Pacific Journal of Mathematics

Volume 274 No. 1 March 2015

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

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

The subscription price for 2015 is US $420/year for the electronic version, and $570/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.1

UNIMODAL SEQUENCES AND "STRANGE" FUNCTIONS: A FAMILY OF QUANTUM MODULAR FORMS

KATHRIN BRINGMANN, AMANDA FOLSOM AND ROBERT C. RHOADES

We construct an infinite family of quantum modular forms from combinatorial rank "moment" generating functions for strongly unimodal sequences. The first member of this family is Kontsevich's "strange" function studied by Zagier. These results rely upon the theory of mock Jacobi forms. As a corollary, we exploit the quantum and mock modular properties of these combinatorial functions in order to obtain asymptotic expansions.

1. Introduction and statement of results

A sequence of integers $\{a_j\}_{j=1}^{s}$ is called a *strongly unimodal sequence of size n* if there exists an integer k such that

$$0 < a_1 < a_2 < \cdots < a_k > a_{k+1} > \cdots > a_s > 0 \tag{1-1}$$

and $a_1 + \cdots + a_s = n$. A number of familiar sequences are strongly unimodal, for example, the sequence of binomial coefficients $\left\{\binom{n}{j-1}\right\}_{j=1}^{n+1}$ with n even. Attached to strongly unimodal sequences is a notion of rank, analogous to the well-known notion of the rank of an integer partition. For more on partition ranks, see for example original works in [Ramanujan 1919; Dyson 1944; Atkin and Swinnerton-Dyer 1954], and the more recent joint work of [Bringmann and Ono 2010] related to mock modular forms. The *rank* of a strongly unimodal sequence is equal to $s - 2k + 1$, the number of terms after the maximal term minus the number of terms that precede it. For example, there are six strongly unimodal sequences of size 5: $\{5\}$, $\{1, 4\}$, $\{4, 1\}$, $\{1, 3, 1\}$, $\{2, 3\}$, $\{3, 2\}$. Their respective ranks are $0, -1, 1, 0, -1, 1$. By letting w (resp. w^{-1}) keep track of the terms after (resp. before) a maximal term, we have that $u(m, n)$, the number of size n and rank m sequences, satisfies

$$U(w; q) := \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} u(m, n)(-w)^m q^n = \sum_{n=0}^{\infty} (wq; q)_n \left(w^{-1}q; q\right)_n q^{n+1}, \tag{1-2}$$

MSC2010: primary 11F99; secondary 11F37, 33D15.
Keywords: quantum modular forms, mock modular forms, Jacobi forms, unimodal sequences, partitions, asymptotics, moment generating functions.

where we set $(w;q)_n := \prod_{j=0}^{n-1}(1-wq^j)$, for $n \in \mathbb{N}_0$.

Recently, Bryson, Ono, Pitman, and the third author [Bryson et al. 2012] studied this function in the special case $w = 1$, namely,[1]

$$U(1;q) = \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} (-1)^m u(m,n) q^n = \sum_{n=1}^{\infty} (u_e(n) - u_o(n)) q^n,$$

where $u_e(n)$ (resp. $u_o(n)$) denotes the number of unimodal sequences of size n with even (resp. odd) rank. They showed that for every root of unity ζ,

$$U(1;\zeta) = F\big(\zeta^{-1}\big),$$

where Kontsevich's "strange" function is defined by

$$F(q) := \sum_{n=0}^{\infty} (q;q)_n.$$

Previously, Zagier [2001] proved that this function satisfies the "identity"

$$F(q) = -\frac{1}{2} \sum_{n=1}^{\infty} n \left(\frac{12}{n}\right) q^{(n^2-1)/24}, \tag{1-3}$$

where $\left(\frac{\cdot}{\cdot}\right)$ is the Kronecker symbol. The two sides of (1-3) don't make sense simultaneously. Indeed, the right-hand side of (1-3) converges in the unit disk $|q| < 1$, but nowhere on the unit circle. The identity (1-3) means that at roots of unity ζ, $F(\zeta)$ (which is clearly a finite sum) agrees with the limit as q approaches ζ radially within the unit disk of the function on the right-hand side of (1-3). Moreover, Zagier proved that for $x \in \mathbb{Q} \setminus \{0\}$,

$$\phi(x) + (-ix)^{-3/2} \phi\left(-\frac{1}{x}\right) = \frac{\sqrt{3i}}{2\pi} \int_0^{i\infty} (w+x)^{-3/2} \eta(w)\, dw, \tag{1-4}$$

where

$$\phi(x) := e^{-\pi i x/12} F\big(e^{-2\pi i x}\big)$$

and

$$\eta(w) := e^{\pi i w/12} \prod_{n=1}^{\infty} \big(1 - e^{2\pi i n w}\big)$$

is the Dedekind eta function. Note that the constant $\sqrt{3i}/2\pi$ in (1-4) is given explicitly in [Bryson et al. 2012]. There, the authors also gave a new proof of (1-4), using the fact that $U(1;q)$ is a (weak) mixed mock modular form for $|q| < 1$. Here, we slightly modify the definition of "mixed mock modular form" given in

[1] Note that the function $U(w;q)$, given in (1-2), is equal to the function $U(-w;q)$ as defined in [Bryson et al. 2012].

[Dabholkar et al. 2014] to mean functions that lie in the tensor product of the general spaces of mock modular forms and weakly holomorphic modular forms (up to possible rational multiples of q powers). In particular, we do not require these functions to be holomorphic at the cusps, as in [loc. cit.]. Weak mixed mock modular forms in this sense occur in a variety of areas including combinatorics [Andrews 2005], algebraic geometry [Vafa and Witten 1994], Lie theory [Kac and Wakimoto 2001], Joyce invariants [Mellit and Okada 2009], and quantum black holes [Manschot 2011; Dabholkar et al. 2014].

The similarity between (1-4) and the usual modular transformation formula of a modular form in part motivated Zagier [2010] to introduce the notion of a quantum modular form. A *quantum modular form of weight $k \in \frac{1}{2}\mathbb{Z}$* is a complex-valued function f on $\mathbb{Q}$ such that for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, the complex-valued function h_γ defined on $\mathbb{Q} \setminus \gamma^{-1}(\infty)$ by

$$h_\gamma(x) := f(x) - \varepsilon(\gamma)(cx+d)^{-k} f\left(\frac{ax+b}{cx+d}\right) \tag{1-5}$$

satisfies a "suitable" property of continuity or analyticity. The $\varepsilon(\gamma)$ in (1-5) are suitable complex numbers, such as those in the theory of half-integral weight modular forms when $k \in \frac{1}{2}\mathbb{Z} \setminus \mathbb{Z}$.

This paper gives an infinite family of quantum modular forms from the "moments" of the unimodal rank statistic. In general, such moment functions are of both number theoretic and combinatorial interest. For example, in their celebrated work, Atkin and Garvan [2003] discovered a partial differential equation relating the bivariate generating functions for the partition statistics rank and crank, leading to exact linear relations between rank and crank moments. Andrews [2007] provided a beautiful combinatorial interpretation of partition rank moments in terms of "k-marked Durfee symbols". Andrews [2008] also discovered a relationship between partition rank moments and the "smallest parts" partition statistic, which has led to further work by Garvan [2011], for example. In addition to intrinsic combinatorial interest, moment functions have been shown to satisfy modular properties. For example, works including [Bringmann et al. 2009; 2010; Alfes et al. 2011] exhibit relationships to weak Maass forms and mock theta functions.

To state our results, we define for $r \in \mathbb{N}_0$ the "weighted" moment functions

$$\phi_r(\tau) := (\pi i)^{2r+1} \sum_{n=1}^{\infty} \sum_{m \in \mathbb{Z}} (-1)^m u(m,n) Q_r\left(m^2, n - \tfrac{1}{24}\right) q^{n-\frac{1}{24}}, \tag{1-6}$$

where here and throughout we set $q := e^{2\pi i \tau}$ and

$$Q_r(X,Y) := \sum_{\substack{0 \le \mu \le r \\ 0 \le \ell \le r-\mu}} c_r(\mu,\ell) X^\ell Y^\mu \in \mathbb{Q}[X,Y], \tag{1-7}$$

the rational coefficients $c_r(\mu, \ell)$ being defined in (1-9). For example, the first few polynomials $\big($normalized, with $Y \to Y - \frac{1}{24}\big)$ are given by

$$\begin{aligned}
&Q_0\big(X, Y - \tfrac{1}{24}\big) = -2,\\
&Q_1\big(X, Y - \tfrac{1}{24}\big) = -4(X + 2Y),\\
&Q_2\big(X, Y - \tfrac{1}{24}\big) = -\tfrac{4}{105}\big(10X + 35X^2 + 6Y + 180XY + 108Y^2\big),\\
&Q_3\big(X, Y - \tfrac{1}{24}\big) = -\tfrac{4}{3465}\big(7X + 140X^2 + 154X^3 + 2Y + 420XY\\
&\qquad\qquad\qquad + 1260X^2Y + 120Y^2 + 2520XY^2 + 720Y^3\big).
\end{aligned}$$

Note that in particular the first member of the family $\phi_r(\tau)$ is (up to a constant) the "strange" function studied by Zagier and Kontsevich discussed above. That is, $\phi_0(\tau) = -2\pi i q^{-1/24} U(1; q) = -2\pi i \phi(\tau)$. It is not difficult to see that the functions $\phi_r(\tau)$ may also be written in terms of the "twisted" unimodal moment functions u_r, defined for integers $r \geq 0$ by

$$u_r(q) := \sum_{n=1}^{\infty} \sum_{m \in \mathbb{Z}} (-1)^m u(m, n) m^r q^n.$$

The moments $\sum_m u(m, n)m^r$ of the unimodal rank statistic are analogous with the rank and crank partition moments, functions which have drawn wide combinatorial interest since Atkin and Garvan [2003] famously introduced them. There is a vast literature on such objects, including asymptotic questions and congruence properties. While the unimodal rank moments are exponentially large for even r [Bringmann et al. $\geq$ 2015], it is surprising that the twisted moments $\sum_m (-1)^m u(m, n)m^r$, as a consequence of our results, are only polynomially large in n. We have chosen to handle the more complicated expressions $\sum_m (-1)^m u(m, n) Q_r(m^2, n - \frac{1}{24})$ because the generating functions for these numbers have a fixed weight as modular objects as seen in Theorem 1.1, while the generating function for the twisted moments will have a mixed weight. To relate these generating functions $\phi_r(\tau)$ to the twisted unimodal moments $u_r(\tau)$, by symmetry, we note that $u_{2r+1}(q) = 0$ for integers $r \geq 0$. In particular, using (1-6), we find that

$$\phi_r(\tau) = (\pi i)^{2r+1} \sum_{\substack{0 \leq \mu \leq r\\ 0 \leq \ell \leq r-\mu}} \frac{c_r(\mu, \ell)}{(2\pi i)^\mu} \cdot \frac{\partial^\mu}{\partial \tau^\mu}\big(u_{2\ell}(q) q^{-\frac{1}{24}}\big), \tag{1-8}$$

where we define

$$c_r(\mu, \ell) := \frac{-2^{2\ell+1} 6^\mu \Gamma\big(\frac{1}{2} + 2r - \mu\big)}{\Gamma\big(\frac{1}{2} + 2r\big) \mu! (2\ell)! (2r - 2\mu - 2\ell + 1)!} \in \mathbb{Q}. \tag{1-9}$$

The coefficients $c_r(\mu, \ell)$ are indeed in $\mathbb{Q}$, as it is well known for integers $k \in \mathbb{N}_0$, that $\Gamma(\frac{1}{2} + k) \in \sqrt{\pi} \cdot \mathbb{Q}$. The twisted moment functions also naturally extend the unimodal function $U(1; q)$ discussed above; namely, $u_0(q) = U(1; q) = -q^{1/24} (2\pi i)^{-1} \phi_0(\tau)$.

To state our first result, we define another polynomial

$$P_r(X,Y) := \sum_{\substack{0\le N\le r\\ 0\le M\le 3r}} b_r(N,M)X^{2N+1}Y^M, \tag{1-10}$$

where the coefficients $b_r(N,M)$ are given explicitly in (3-13). Our first theorem establishes that the unimodal moment functions ϕ_r are quantum modular forms on $\mathbb{Q}\backslash\{0\}$, and that their transformation law also extends to $\mathbb{H}$. The function $\mathcal{H}_r$ below is defined in (3-14).

Theorem 1.1. *Let $r \in \mathbb{N}_0$. If $\tau \in \mathbb{H}\cup\mathbb{Q}\backslash\{0\}$, we have*

$$\begin{aligned}\phi_r(\tau) - (-i\tau)^{-3/2-2r}\phi_r\Big(-\frac{1}{\tau}\Big)\\ = \int_{\mathbb{R}} P_r\big(w,(-i\tau)^{-1}\big)e^{\pi i\tau w^2/3}\frac{\sinh\big(\frac{2\pi w}{3}\big)}{\cosh(\pi w)}\,dw + \mathcal{H}_r(\tau),\end{aligned} \tag{1-11}$$

where $\mathcal{H}_r(\tau) = 0$ for $\tau \in \mathbb{Q}\setminus\{0\}$. In particular, the functions ϕ_r are quantum modular forms.

Remarks. (1) The transformation law given in (1-11) in the case $\tau\in\mathbb{H}$ essentially establishes the mock modular properties of the unimodal rank moment functions $\phi_r(\tau)$.

(2) In the course of proving (1-11) in the case $\tau \in \mathbb{Q}\setminus\{0\}$, we show that for each integer $r \ge 0$, the function ϕ_r is defined for $\tau\in\mathbb{Q}$. Moreover, in Theorem 5.1 of Section 5, we pay special attention to the case $r = 1$, and establish an explicit finite value for $\phi_1(h/k)$ $(h,k\in\mathbb{Z})$ as the value of a polynomial in the root of unity $e^{2\pi ih/k}$.

(3) Our functions naturally arise from mock Jacobi forms. It would be interesting to investigate whether a theory of quantum Jacobi forms could be developed that contains functions arising in this paper as special cases.

Our next theorem exploits the automorphic properties given in Theorem 1.1, and establishes the asymptotic behavior of the moment functions u_r. While such properties are of independent interest, we also point out that these functions are related to the quantum moment functions ϕ_r by (1-8). To describe their asymptotic behavior, we use the Bernoulli polynomials $B_k(x)$ and Euler polynomials $E_k(x)$, defined by the generating functions

$$\frac{ze^{xz}}{e^z-1} = \sum_{k=0}^{\infty} B_k(x)\frac{z^k}{k!} \tag{1-12}$$

and

$$\frac{2e^{xz}}{e^z+1} = \sum_{k=0}^{\infty} E_k(x)\frac{z^k}{k!}. \tag{1-13}$$

Theorem 1.2. *For nonnegative integers r, as $t \to 0^+$, we have*

$$e^{\pi t/12} u_{2r}\left(e^{-2\pi t}\right) = \frac{3^{2r+1}}{2r+1} \sum_{k=0}^{\infty} \frac{(3\pi t)^k}{k!} \sum_{0 \le n \le r} \binom{2r+1}{2n} 3^{-2n} B_{2n}\left(\frac{1}{2}\right) E_{2r+1+2k-2n}\left(\frac{5}{6}\right),$$

In particular, we have

$$e^{\pi t/12} u_{2r}\left(e^{-2\pi t}\right) \sim \frac{2 \cdot 6^{2r}}{2r+1}\left(B_{2r+1}\left(\frac{2}{3}\right) + B_{2r+1}\left(\frac{5}{6}\right)\right).$$

The paper is organized as follows. In Section 2 we provide relevant background information on modular forms, Jacobi forms, and mock Jacobi forms, as well as Bernoulli and Euler polynomials. In Section 3 we prove Theorem 1.1, and in Section 4 we establish Theorem 1.2. In Section 5 we pay special consideration to the moment function ϕ_1.

2. Preliminaries

Here, we provide preliminary information on automorphic forms in Section 2A, and Bernoulli and Euler polynomials in Section 2B.

2A. ***Automorphic forms.*** In this section, we recall some fundamental properties of certain modular and (mock) Jacobi forms. We start with the well-known transformation law for the Dedekind η-function.

Lemma 2.1. *For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we have*

$$\eta(\gamma\tau) = \chi(\gamma)(c\tau + d)^{1/2}\eta(\tau), \tag{2-1}$$

where $\chi(\gamma)$ is a 24-th root of unity, which can be given explicitly in terms of Dedekind sums [Rademacher 1973]. *In particular, we have*

$$\eta\left(-\frac{1}{\tau}\right) = \sqrt{-i\tau}\,\eta(\tau).$$

Here and throughout the square root is defined by the principal branch of the logarithm. Moreover, we require the usual Jacobi theta function, defined for $z \in \mathbb{C}$ and $\tau \in \mathbb{H}$ by

$$\vartheta(z;\tau) := \sum_{\nu \in \frac{1}{2}+\mathbb{Z}} e^{\pi i \nu^2 \tau + 2\pi i \nu\left(z+\frac{1}{2}\right)}. \tag{2-2}$$

This function is well known to satisfy the following transformation law [Rademacher 1973, (80.31) and (80.8)]:

Lemma 2.2. *For $\lambda, \mu \in \mathbb{Z}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we have*

$$\vartheta(z+\lambda\tau+\mu;\tau) = (-1)^{\lambda+\mu} q^{-\lambda^2/2} e^{-2\pi i\lambda z}\vartheta(z;\tau),$$
$$\vartheta\left(\frac{z}{c\tau+d};\gamma\tau\right) = \chi^3(\gamma)(c\tau+d)^{1/2} e^{\pi i c z^2/(c\tau+d)}\vartheta(z;\tau).$$

In particular,

$$\vartheta\left(\frac{z}{\tau};-\frac{1}{\tau}\right) = -i\sqrt{-i\tau}\, e^{\pi i z^2/\tau}\vartheta(z;\tau).$$

The Jacobi theta function also satisfies the well-known triple product identity ($w = e^{2\pi i z}$)

$$\vartheta(z;\tau) = -iq^{1/8}w^{-1/2}\prod_{n=1}^{\infty}\left(1-q^n\right)\left(1-wq^{n-1}\right)\left(1-w^{-1}q^n\right).$$

Additionally, we require the following classical Taylor expansion (see for example [Zagier 1991]):

(2-3) $$\vartheta(z;\tau) = -2\pi z\cdot\eta^3(\tau)\exp\left(-2\sum_{k=1}^{\infty} G_{2k}(\tau)\frac{(2\pi i z)^{2k}}{(2k)!}\right).$$

Here for even integers $k \geq 2$, the Eisenstein series are defined by

$$G_k(\tau) := -\frac{B_k}{2k} + \sum_{n=1}^{\infty}\sigma_{k-1}(n)q^n,$$

where $\sigma_\ell(n) := \sum_{d|n} d^\ell$ and B_k denotes the k-th Bernoulli number.

We also make use of Zwegers' functions $A_\ell(z_1, z_2;\tau)$ [2010] (see also [Bringmann 2008; Andrews et al. 2013]), defined for $\ell \in \mathbb{N}$, $\tau \in \mathbb{H}$, $z_2 \in \mathbb{C}$, and $z_1 \in \mathbb{C}\setminus(\mathbb{Z}\tau+\mathbb{Z})$ by

(2-4) $$A_\ell(z_1,z_2;\tau) := e^{\ell\pi i z_1}\sum_{n\in\mathbb{Z}}\frac{(-1)^{\ell n}q^{\ell n(n+1)/2}e^{2\pi i n z_2}}{1-q^n e^{2\pi i z_1}}.$$

These functions may be "completed" into nonholomorphic Jacobi forms by setting

$$\widehat{A}_\ell(z_1,z_2;\tau) := A_\ell(z_1,z_2;\tau) + R_\ell(z_1,z_2;\tau).$$

The nonholomorphic completions of these higher-level Appell functions are defined by

$$R_\ell(z_1,z_2;\tau) := \frac{i}{2}\sum_{k=0}^{\ell-1} e(kz_1)\vartheta\left(z_2+k\tau+\frac{\ell-1}{2};\ell\tau\right)R\left(\ell z_1 - z_2 - k\tau - \frac{\ell-1}{2};\ell\tau\right),$$

where $e(x) := e^{2\pi i x}$ and where (with $\tau = u + iv$)

$$R(z;\tau) := \sum_{n\in\frac{1}{2}+\mathbb{Z}} \left(\operatorname{sgn}(n) - E\left(\left(n+\frac{\operatorname{Im}(z)}{v}\right)\sqrt{2v}\right)\right)(-1)^{n-\frac{1}{2}} q^{-n^2/2} e^{-2\pi i n z},$$

with $E(z) := 2\int_0^z e^{-\pi t^2}\,dt$. Proposition 2.3 below shows that the so-called "error to modularity" of the function $R(z;\tau)$ is the Mordell integral, defined for $z\in\mathbb{C}$ and $\tau\in\mathbb{H}$ by

$$h(z;\tau) := \int_{\mathbb{R}} \frac{e^{\pi i\tau w^2 - 2\pi z w}}{\cosh(\pi w)}\,dw. \tag{2-5}$$

Proposition 2.3 [Zwegers 2002]. *For $z\in\mathbb{C}$ and $\tau\in\mathbb{H}$, we have*

$$\begin{aligned} R(z+1;\tau) &= -R(z;\tau),\\ R\left(\frac{z}{\tau};-\frac{1}{\tau}\right) &= \sqrt{-i\tau}\,e^{-\pi i z^2/\tau}(-R(z;\tau)+h(z;\tau)). \end{aligned}$$

The completed higher-level Appell functions $A_\ell(z_1,z_2;\tau)$ transform as follows.

Proposition 2.4 [Zwegers 2010]. *For $n_1,n_2,m_1,m_2\in\mathbb{Z}$ and $\gamma=\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$, we have*

$$\begin{aligned} &\widehat{A}_\ell(z_1+n_1\tau+m_1, z_2+n_2\tau+m_2;\tau)\\ &\qquad = (-1)^{\ell(n_1+m_1)} e(z_1(\ell n_1-n_2)-n_1z_2) q^{\ell n_1^2/2-n_1n_2}\widehat{A}_\ell(z_1,z_2;\tau),\\ &\widehat{A}_\ell\left(\frac{z_1}{c\tau+d},\frac{z_2}{c\tau+d};\gamma\tau\right) = (c\tau+d)e\left(\frac{c(-\ell z_1^2+2z_1z_2)}{2(c\tau+d)}\right)\widehat{A}_\ell(z_1,z_2;\tau). \end{aligned}$$

We further require "dissection properties" of the functions ϑ and R (see [Shimura 1973; Zwegers 2010; Bringmann and Folsom 2013]).

Lemma 2.5. *With notation as above, we have for $n\in\mathbb{N}$,*

$$\vartheta\left(z;\frac{\tau}{n}\right) = \sum_{\ell=0}^{n-1} q^{\left(\ell-\frac{n-1}{2}\right)^2/(2n)} e^{2\pi i\left(\ell-\frac{n-1}{2}\right)\left(z+\frac{1}{2}\right)}\vartheta\left(nz+\left(\ell-\frac{n-1}{2}\right)\tau+\frac{n-1}{2};n\tau\right),$$

$$R\left(z;\frac{\tau}{n}\right) = \sum_{\ell=0}^{n-1} q^{-\left(\ell-\frac{n-1}{2}\right)^2/(2n)} e^{-2\pi i\left(\ell-\frac{n-1}{2}\right)\left(z+\frac{1}{2}\right)} R\left(nz+\left(\ell-\frac{n-1}{2}\right)\tau+\frac{n-1}{2};n\tau\right).$$

2B. *Bernoulli and Euler polynomials.* In this section, we recall certain properties of the Bernoulli polynomials $B_k(x)$ and Euler polynomials $E_k(x)$, defined in (1-12) and (1-13), respectively, as well as their special values

$$B_k := B_k(0), \qquad E_k := 2^k E_k\left(\tfrac{1}{2}\right).$$

One property we make use of is a "dissection" property of the Bernoulli polynomials (see [Abramowitz and Stegun 1964, Chapter 23])

$$B_k(mx) = m^{k-1}\sum_{a=0}^{m-1} B_k\left(x+\frac{a}{m}\right) \quad \text{for } m \in 2\mathbb{N}_0+1. \tag{2-6}$$

Another "splitting" property that we use is

$$2^k B_k\left(\frac{x+y}{2}\right) = \sum_{j=0}^{k}\binom{k}{j} B_j(x)E_{k-j}(y), \tag{2-7}$$

which follows easily from the definition of the Euler and Bernoulli polynomials, using the fact that

$$\frac{2z\cdot e^{(x+y)z}}{e^{2z}-1} = \frac{ze^{xz}}{e^z-1}\cdot\frac{2e^{yz}}{e^z+1}.$$

Here and throughout, we let $\zeta_N := e^{2\pi i/N}$ for $N \in \mathbb{N}$. The next lemma expresses derivatives of secant in terms of Euler polynomials.

Lemma 2.6. *With notation as above, we have, for $c \in \mathbb{N}_0$,*

$$\sec^{(2c+1)}\left(\tfrac{\pi}{3}\right) = (-1)^c\sqrt{3}\cdot 6^{2c+1}E_{2c+1}\left(\tfrac{5}{6}\right).$$

Proof. This follows quickly from [Cvijović 2009, Theorem 2]. Namely, using the facts that $E_{2c-1}\left(\frac{1}{6}\right) = -E_{2c-1}\left(\frac{5}{6}\right)$ and $E_{2c-1}\left(\frac{1}{2}\right) = 0$ gives the claim. □

A fourth property that we use expresses the Euler numbers as integrals. Namely, it is known (see [Erdélyi et al. 1981, p. 42, Equation (18)] for example) that for $k \in \mathbb{N}_0$,

$$\int_{\mathbb{R}} \frac{w^{2k}}{\cosh(\pi w)}\,dw = (2i)^{-2k}E_{2k}. \tag{2-8}$$

Note that $E_{2k-1} = 0$ for $k \in \mathbb{N}$.

3. Proof of Theorem 1.1

Here, we ultimately conclude Theorem 1.1 from Propositions 3.6–3.8 below. In Section 3A, we establish properties of mock Jacobi forms related to the unimodal rank generating function; and in Section 3B, we construct mock modular forms from its Taylor coefficients. In Section 3C, we establish quantum modularity and prove Theorem 1.1. Until otherwise indicated, throughout this section, we take $\tau \in \mathbb{H}$.

3A. ***Mock Jacobi forms and unimodal ranks.*** Here we establish properties of mock Jacobi forms associated to the unimodal rank generating function. We begin by writing $U(w;q)$ in terms of the Appell functions $A_\ell(u,v;\tau)$ defined in (2-4). Throughout, for $w_1, w_2 \in \mathbb{C}$, we let

$$\mathcal{U}(w_1;w_2) := U(e(w_1);e(w_2)).$$

Lemma 3.1. *Let $w = e(z)$. With notation as above, we have*

$$\mathcal{U}(z;\tau) = \frac{1}{\left(w^{1/2}-w^{-1/2}\right)(q;q)_\infty}\left(A_1(z,-z;\tau) - w^{-1}A_3(z,-\tau;\tau)\right).$$

Proof. Entry 3.4.7 of "Ramanujan's lost notebook" (see [Andrews and Berndt 2009, p. 67]) gives with $a=-w$, $b=-w^{-1}$ that $\mathcal{U}(z;\tau)$ equals

$$\text{(3-1)} \qquad \frac{-1}{(1-w)(1-w^{-1})}\sum_{n=0}^{\infty}\frac{q^{n^2}}{(wq;q)_n(w^{-1}q;q)_n} + \frac{1}{(1-w^{-1})(q;q)_\infty}\sum_{n\in\mathbb{Z}}\frac{(-1)^n q^{n(n+1)/2}w^{-n}}{1-wq^n}.$$

We note that the second sum on the right-hand side of (3-1) is easily seen to equal

$$\frac{1}{(w^{1/2}-w^{-1/2})(q;q)_\infty}A_1(z,-z;\tau).$$

Using these facts, the result follows after applying the identity (see [Atkin and Swinnerton-Dyer 1954])

$$\frac{-1}{(1-w^{-1})(1-w)}\sum_{n=0}^{\infty}\frac{q^{n^2}}{(wq;q)_n(w^{-1}q;q)_n} = \frac{-1}{(w^{1/2}-w^{-1/2})}\frac{1}{(q;q)_\infty}A_3(z,-\tau;\tau). \qquad \square$$

Next we define a normalization of the function $\mathcal{U}(z;\tau)$

$$\text{(3-2)} \qquad \begin{aligned} Y^+(z;\tau) &:= -\left(w^{1/2}-w^{-1/2}\right)q^{-1/24}\cdot\mathcal{U}(z;\tau) \\ &= \eta^{-1}(\tau)\left(w^{-1}A_3(z,-\tau;\tau) - A_1(z,-z;\tau)\right), \end{aligned}$$

where the second equality follows from Lemma 3.1. Using Proposition 3.3, we now establish a transformation law for Y^+, which is a key step in showing quantum modularity of the functions ϕ_r. To state this, we define

$$H(z;\tau) := \frac{i}{2}\frac{\vartheta(z;\tau)}{\eta(\tau)}h(2z;\tau) - g(z;\tau),$$

where $h(z;\tau)$ is given in (2-5), and

$$g(z;\tau) := \frac{i}{\sqrt{3}} \int_{\mathbb{R}} e^{\pi i \tau w^2/3 - 2\pi wz} \frac{\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\, dw.$$

Proposition 3.2. *With notation as above, we have*

$$-ie^{3\pi i z^2/\tau} Y^+\left(\frac{z}{\tau}; -\frac{1}{\tau}\right) \frac{1}{\sqrt{-i\tau}} - Y^+(z;\tau) = H(z;\tau).$$

To prove Proposition 3.2 we rather work with a second normalization of the function $\mathcal{U}(z;\tau)$, namely,

$$\begin{aligned} X^+(z;\tau) &:= -e^{-3\pi z^2/(2v)} \left(w^{1/2} - w^{-1/2}\right)(q;q)_\infty \mathcal{U}(z;\tau) \\ &= \left(w^{-1} A_3(z, -\tau; \tau) - A_1(z, -z; \tau)\right) e^{-3\pi z^2/(2v)}. \end{aligned}$$

Moreover we need the completed function

$$\begin{aligned} \widehat{X}(z;\tau) &:= \left(w^{-1} \widehat{A}_3(z, -\tau; \tau) - \widehat{A}_1(z, -z; \tau)\right) e^{-3\pi z^2/(2v)} \\ &= \left(\widehat{A}_3(z, 0; \tau) - \widehat{A}_1(z, -z; \tau)\right) e^{-3\pi z^2/(2v)}, \end{aligned} \tag{3-3}$$

where the second equality follows from the first transformation in Proposition 2.4.

Using Proposition 2.4, it is not difficult to establish a modularity result for $\widehat{X}(z;\tau)$:

Proposition 3.3. *With notation as above, for* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, *we have*

$$\widehat{X}\left(\frac{z}{c\tau + d}; \gamma\tau\right) = (c\tau + d)\widehat{X}(z;\tau).$$

From Proposition 3.3, we can establish a transformation property of $X^+(z;\tau)$:

Proposition 3.4. *With notation as above, we have that*

$$\begin{aligned} &X^+\left(\frac{z}{\tau}; -\frac{1}{\tau}\right)\tau^{-1} - X^+(z;\tau) \\ &\qquad = \left(\frac{i}{2}\vartheta(z;\tau)h(2z;\tau) + \frac{i}{2\sqrt{3}}\eta(\tau) \sum_{\pm} \pm h\left(z \pm \frac{1}{3}; \frac{\tau}{3}\right)\right) e^{-3\pi z^2/(2v)}. \end{aligned}$$

Proof. Using Proposition 3.3 we obtain that

$$\left(X^+\left(\frac{z}{\tau}; -\frac{1}{\tau}\right)\tau^{-1} - X^+(z;\tau)\right) 2i = f_1(z;\tau) + f_2(z;\tau),$$

with

$$f_1(z;\tau) := \vartheta\Big(-\frac{1}{\tau};-\frac{3}{\tau}\Big)e^{-3\pi z^2\bar{\tau}/(2v\tau)}\tau^{-1}\sum_{\pm}\pm e^{\pm 2\pi i z/\tau}R\Big(\frac{3z}{\tau}\pm\frac{1}{\tau};-\frac{3}{\tau}\Big) \\ -\vartheta(\tau;3\tau)e^{-3\pi z^2/(2v)}\sum_{\pm}\pm e^{\pm 2\pi i z}R(3z\mp\tau;3\tau),$$

$$f_2(z;\tau) := \vartheta\Big(\frac{z}{\tau};-\frac{1}{\tau}\Big)R\Big(\frac{2z}{\tau};-\frac{1}{\tau}\Big)e^{-3\pi z^2\bar{\tau}/(2v\tau)}\tau^{-1}-\vartheta(z;\tau)R(2z;\tau)e^{-3\pi z^2/(2v)}.$$

We next simplify f_1 and f_2. Firstly, using Lemma 2.2 and Proposition 2.3, we obtain that

$$f_2(z;\tau) = -\vartheta(z;\tau)h(2z;\tau)e^{-3\pi z^2/(2v)}. \tag{3-4}$$

Next Lemma 2.2 and Proposition 2.3 yield that

$$\vartheta\Big(-\frac{1}{\tau};-\frac{3}{\tau}\Big)e^{-3\pi z^2\bar{\tau}/(2v\tau)}\tau^{-1}\sum_{\pm}\pm e^{\pm 2\pi i z/\tau}R\Big(\frac{3z}{\tau}\pm\frac{1}{\tau};-\frac{3}{\tau}\Big) \\ = -\frac{1}{3}e^{-3\pi z^2/(2v)}\vartheta\Big(-\frac{1}{3};\frac{\tau}{3}\Big)\sum_{\pm}\pm\Big(-R\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big)+h\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big)\Big).$$

Now Lemma 2.5, the fact that $\vartheta(0;\tau)=0$, and Proposition 2.3, give that

$$\vartheta\Big(-\frac{1}{3};\frac{\tau}{3}\Big) = 2i\sin\Big(\frac{\pi}{3}\Big)q^{1/6}\vartheta(\tau;3\tau),$$
$$R\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big) = -q^{-\frac{1}{6}}e^{2\pi i(z\pm\frac{1}{3})}R(3z-\tau;3\tau)+R(3z;3\tau) \\ -q^{-1/6}e^{-2\pi i(z\pm\frac{1}{3})}R(3z+\tau;3\tau).$$

Thus

$$\sum_{\pm}\mp R\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big) = 2i\sin\Big(\frac{2\pi}{3}\Big)q^{-1/6}\sum_{\pm}\pm e^{\pm 2\pi i z}R(3z\mp\tau;3\tau),$$

and hence

$$f_1(z;\tau) = -\frac{i}{\sqrt{3}}q^{1/6}\vartheta(\tau;3\tau)\sum_{\pm}\pm h\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big)e^{-3\pi z^2/(2v)}. \tag{3-5}$$

Combining (3-4), (3-5), and the fact that $\vartheta(\tau;3\tau)=-iq^{-1/6}\eta(\tau)$ gives the claim. □

Proof of Proposition 3.2. First note that

$$\sum_{\pm}\pm h\Big(z\pm\frac{1}{3};\frac{\tau}{3}\Big) = 2i\sqrt{3}\cdot g(z;\tau).$$

The result now follows immediately from Proposition 3.4 and Lemma 2.1, using the fact that

$$Y^+(z;\tau)=\frac{e^{3\pi z^2/(2v)}}{\eta(\tau)}X^+(z;\tau). \qquad \square$$

3B. *Taylor coefficients and unimodal ranks.* Using the results from Section 3A, we next construct mock modular forms from the Taylor coefficients of the unimodal rank generating function. The functions $H(z;\tau)$ and $Y^+(z;\tau)$ are holomorphic in z, and it is not difficult to see that they are both odd functions in z. So we may write

$$Y^+(z;\tau)=\sum_{r=0}^{\infty}a_{2r}(\tau)z^{2r+1}, \tag{3-6}$$

$$H(z;\tau)=\sum_{r=0}^{\infty}h_{2r}(\tau)z^{2r+1}. \tag{3-7}$$

The next lemma describes the modularity properties of the Taylor coefficients $a_{2r}(\tau)$ of $Y^+(z;\tau)$.

Lemma 3.5. *With notation as above, we have*

$$a_{2r}\Big(-\frac{1}{\tau}\Big)(-i\tau)^{-3/2-2r}=\sum_{0\le j\le r}\frac{(3\pi)^{r-j}}{(r-j)!}(-1)^{j+1}(-i\tau)^{j-r}(a_{2j}(\tau)+h_{2j}(\tau)).$$

Proof. Proposition 3.2 directly yields

$$Y^+\Big(\frac{z}{\tau};-\frac{1}{\tau}\Big)=ie^{-3\pi iz^2/\tau}\sqrt{-i\tau}(Y^+(z;\tau)+H(z;\tau)).$$

Inserting (3-6), (3-7), and the Taylor expansion of the exponential function, we obtain

$$\begin{aligned}\sum_{r=0}^{\infty}a_{2r}\Big(-\frac{1}{\tau}\Big)\Big(\frac{z}{\tau}\Big)^{2r+1}&\\ &=i\sqrt{-i\tau}\sum_{\ell=0}^{\infty}\frac{(-3\pi iz^2/\tau)^{\ell}}{\ell!}\sum_{j=0}^{\infty}(a_{2j}(\tau)+h_{2j}(\tau))z^{2j+1}\\ &=i\sqrt{-i\tau}\sum_{r=0}^{\infty}z^{2r+1}\sum_{0\le j\le r}\frac{(3\pi)^{r-j}}{(r-j)!}(-1)^{r+j}(-i\tau)^{j-r}(a_{2j}(\tau)+h_{2j}(\tau)).\end{aligned}$$

Equating the coefficients of z^{2r+1} gives the claim. $\square$

To prove the transformation law for the functions ϕ_r, we define for $r\in\mathbb{N}_0$,

$$b_{2r}(\tau):=\sum_{0\le\mu\le r}\frac{(3\pi i)^{\mu}\Gamma\big(\frac12+2r-\mu\big)}{\Gamma\big(\frac12+2r\big)\mu!}a_{2r-2\mu}^{(\mu)}(\tau). \tag{3-8}$$

We will later show that $\phi_r(\tau) = b_{2r}(\tau)$. The functions $b_{2r}(\tau)$ transform as described in the following proposition, a fact which follows as in [Eichler and Zagier 1985], using Lemma 3.5.

Proposition 3.6. *With notation as above, for $r \in \mathbb{N}_0$, we have*

$$b_{2r}\left(-\frac{1}{\tau}\right)(-i\tau)^{-3/2-2r} - b_{2r}(\tau)$$
$$= -(-i\tau)^{-3/2-2r} \sum_{0\le\mu\le r} \frac{(3\pi i)^{\mu}\Gamma\left(\frac{1}{2}+2r-\mu\right)}{\Gamma\left(\frac{1}{2}+2r\right)\mu!}$$
$$\times \sum_{0\le j\le r-\mu} \frac{(3\pi)^{r-\mu-j}(-1)^j}{(r-\mu-j)!} \frac{\partial^{\mu}}{\partial\tau^{\mu}}\left((-i\tau)^{j+r-\mu+\frac{3}{2}} h_{2j}(\tau)\right).$$

Our next proposition shows that the "errors to modularity" h_{2r} are C^{∞}, a fact we use in the course of establishing the quantum modularity of the unimodal rank functions ϕ_r in Theorem 1.1. In doing so, we split the Taylor expansion of $H(z;\tau)$ into two pieces

$$H(z;\tau) = H_1(z;\tau) + H_2(z;\tau), \tag{3-9}$$

with

$$H_1(z;\tau) = \sum_{r=0}^{\infty} h_{1,2r}(\tau) z^{2r+1} := \frac{i}{2}\frac{\vartheta(z;\tau)}{\eta(\tau)} h(2z;\tau),$$
$$H_2(z;\tau) = \sum_{r=0}^{\infty} h_{2,2r}(\tau) z^{2r+1} := -g(z;\tau).$$

Proposition 3.7. *The functions h_{2r} are C^{∞} on $\mathbb{R}$. To be more precise, $h_{1,2r}(\tau)$ vanishes to infinite order for $\tau \in \mathbb{Q}$, and we extend this function to equal 0 on all of $\mathbb{R}$. Moreover, for $\tau \in \mathbb{H} \cup \mathbb{Q}$, the function $h_{2,2r}$ satisfies*

$$h_{2,2r}(\tau) = \frac{i}{\sqrt{3}} \frac{(2\pi)^{2r+1}}{(2r+1)!} \int_{\mathbb{R}} e^{\pi i\tau w^2/3} w^{2r+1} \frac{\sinh\left(\frac{2\pi w}{3}\right)}{\sinh(\pi w)}\, dw.$$

Proof. Firstly, we have

$$H_1(z;\tau)$$
$$= \frac{i}{2\eta(\tau)} \sum_{r=0}^{\infty} \frac{\partial^r}{\partial z^r}[\vartheta(z;\tau)h(2z;\tau)]_{z=0} \frac{z^r}{r!}$$
$$= \frac{i}{2\eta(\tau)} \sum_{r=0}^{\infty} \frac{z^{2r+1}}{(2r+1)!} \sum_{\ell=0}^{r} \binom{2r+1}{2\ell+1} \frac{\partial^{2\ell+1}}{\partial z^{2\ell+1}}[\vartheta\,(z;\tau)]_{z=0} \frac{\partial^{2r-2\ell}}{\partial z^{2r-2\ell}}[h(2z;\tau)]_{z=0},$$

so that

$$h_{1,2r}(\tau) = \frac{i}{2\eta(\tau)} \sum_{\ell=0}^{r} \frac{1}{(2\ell+1)!\,(2r-2\ell)!} \frac{\partial^{2\ell+1}}{\partial z^{2\ell+1}} [\vartheta(z;\tau)]_{z=0} \frac{\partial^{2r-2\ell}}{\partial z^{2r-2\ell}} [h(2z;\tau)]_{z=0}.$$

It is not hard to see that $h(2z;\tau)$ is C^∞ as a function of τ near $z=0$. Moreover by (2-3), we see that

$$\frac{i}{2\eta(\tau)} \frac{\partial^{2\ell+1}}{\partial z^{2\ell+1}} [\vartheta(z;\tau)]_{z=0}$$

gives a linear combination of Eisenstein series multiplied by $\eta^2(\tau)$. It is well known that the Eisenstein series satisfy

$$G_k\Big(-\frac{1}{\tau}\Big) = \tau^k G_k(\tau) \quad (k>2,\text{ even})$$

and

$$G_2\Big(-\frac{1}{\tau}\Big) = \tau^2 G_2(\tau) + \frac{i\tau}{4\pi}.$$

This implies that the function $h_{2r}(\tau)$ and its derivatives vanish exponentially for $\tau \in \mathbb{Q}$. The second claim follows directly by inserting the Taylor expansion of $e^{-2\pi z x}$. $\square$

3C. *Quantum unimodal ranks.* Building from the results in Sections 3A and 3B, here we prove Theorem 1.1.

Proof of Theorem 1.1. We first relate the Taylor coefficients of $Y^+(z;\tau)$ to the unimodal moments u_{2r}. Using the definition of u_{2r}, it is not difficult to verify that

(3-10) $$\mathcal{U}(z;\tau) = \sum_{r=0}^{\infty} u_{2r}(q) \frac{(2\pi i z)^{2r}}{(2r)!}.$$

Using the Taylor expansion of $\sin(\pi z)$ we find that

$$\begin{aligned} Y^+(z;\tau) &= -2iq^{-1/24}\sin(\pi z)\mathcal{U}(z;\tau) \\ &= -(2\pi i z)\sum_{r=0}^{\infty}(2\pi i z)^{2r} \sum_{0\le \ell\le r} \frac{u_{2\ell}(q)q^{-1/24}2^{2\ell-2r}}{(2\ell)!\,(2r-2\ell+1)!}, \end{aligned}$$

yielding

(3-11) $$\frac{a_{2r}(\tau)}{(2\pi i)^{2r+1}} = -\sum_{0\le\ell\le r} \frac{u_{2\ell}(q)q^{-1/24}2^{2\ell-2r}}{(2\ell)!\,(2r-2\ell+1)!}.$$

Using (3-11), the definition of $\phi_r(\tau)$ in (1-6), or its equivalent formulation given in (1-8), as well as the definition of $b_{2r}(\tau)$ in (3-8), it is not difficult to see that for each $r \in \mathbb{N}_0$, $b_{2r}(\tau) = \phi_r(\tau)$. Combining this with the fact that

$$h_{2j}(\tau) = h_{1,2j}(\tau) + h_{2,2j}(\tau),$$

Proposition 3.6 yields

$$\begin{aligned}\phi_r\Big(-\frac{1}{\tau}\Big)(-i\tau)^{-3/2-2r} - \phi_r(\tau) &\\ = -(-i\tau)^{-3/2-2r} \sum_{\substack{0\le\mu\le r\\ 0\le j\le r-\mu}} &\frac{(3\pi)^{r-j}(-1)^j i^\mu \Gamma\big(\frac{1}{2}+2r-\mu\big)}{\Gamma(\frac{1}{2}+2r)\mu!(r-\mu-j)!}\\ &\times \frac{\partial^\mu}{\partial\tau^\mu}\Big((-i\tau)^{j+r-\mu+\frac{3}{2}}(h_{1,2j}(\tau)+h_{2,2j}(\tau))\Big).\end{aligned} \tag{3-12}$$

By continuation, (3-12) and what follows hold on $\mathbb{H} \cup \mathbb{Q} \setminus \{0\}$.

We first consider the first summand. We have by Proposition 3.7

$$\begin{aligned}&\frac{\partial^\mu}{\partial\tau^\mu}\Big((-i\tau)^{j+r-\mu+\frac{3}{2}}h_{2,2j}(\tau)\Big)\\ &= \frac{i}{\sqrt{3}}\frac{(2\pi)^{2j+1}}{(2j+1)!}\\ &\quad\times \int_{\mathbb{R}}\sum_{\ell=0}^{\mu}\binom{\mu}{\ell}\frac{\partial^\ell}{\partial\tau^\ell}\Big((-i\tau)^{j+r-\mu+\frac{3}{2}}\Big)\frac{\partial^{\mu-\ell}}{\partial\tau^{\mu-\ell}}\Big(e^{\pi i w^2\tau/3}\Big)w^{2j+1}\frac{\sinh\big(\frac{2\pi w}{3}\big)}{\cosh(\pi w)}\,dw\\ &= \sum_{\ell=0}^{\mu}(-1)^\ell i^{\mu+1}\pi^{2j+1+\mu-\ell}2^{2j+1}3^{\ell-\mu-\frac{1}{2}}\binom{\mu}{\ell}\frac{\Gamma\big(j+r-\mu+\frac{5}{2}\big)}{(2j+1)!\Gamma\big(j+r-\mu+\frac{5}{2}-\ell\big)}\\ &\qquad\qquad\times(-i\tau)^{j+r+\frac{3}{2}-\mu-\ell}\int_{\mathbb{R}}w^{2j+2\mu-2\ell+1}e^{\pi i w^2\tau/3}\frac{\sinh\big(\frac{2\pi w}{3}\big)}{\cosh(\pi w)}\,dw.\end{aligned}$$

We now define the numbers

$$b_r(\mu, j, \ell) := \frac{i(-1)^{j+\ell+\mu}2^{2j+1}\pi^{r+j+\mu+1-\ell}3^{r+\ell-\mu-j-\frac{1}{2}}\Gamma\big(\frac{1}{2}+2r-\mu\big)\Gamma\big(j+r-\mu+\frac{5}{2}\big)}{(2j+1)!\ell!(\mu-\ell)!(r-\mu-j)!\Gamma\big(\frac{1}{2}+2r\big)\Gamma\big(j+r-\mu+\frac{5}{2}-\ell\big)},$$

and let

$$b_r(N, M) := \sum_{0\le\mu\le r}\ \sum_{\substack{0\le j\le r-\mu\\ 0\le\ell\le\mu\\ N=j+\mu-\ell\\ M=\mu+\ell+r-j}} b_r(\mu, j, \ell). \tag{3-13}$$

Moreover, we define $\mathcal{H}_r(\tau)$ to be

$$(3\text{-}14)\quad (-i\tau)^{-\frac{3}{2}-2r} \times \sum_{\substack{0\le\mu\le r\\ 0\le j\le r-\mu}} \frac{(3\pi)^{r-j}(-1)^j i^\mu \Gamma\left(\frac{1}{2}+2r-\mu\right)}{\Gamma\left(\frac{1}{2}+2r\right)\mu!(r-\mu-j)!} \frac{\partial^\mu}{\partial\tau^\mu}\Big((-i\tau)^{j+r-\mu+\frac{3}{2}} h_{1,2j}(\tau)\Big).$$

Note that $\mathcal{H}_r(\tau)=0$ for $\tau\in\mathbb{Q}\setminus\{0\}$. We have thus shown for $\tau\in\mathbb{H}\cup\mathbb{Q}\setminus\{0\}$,

$$(-i\tau)^{-\frac{3}{2}-2r}\phi_r\Big(-\frac{1}{\tau}\Big)-\phi_r(\tau) = -\int_{\mathbb{R}} P_r\big(w,(-i\tau)^{-1}\big)e^{\pi i\tau w^2/3}\frac{\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\,dw-\mathcal{H}_r(\tau),$$

as claimed in (1-11).

Finally, under the translation $\tau\to\tau+1$, it is clear using the definition of $\phi_r(\tau)$ in (1-6) that $\phi_r(\tau+1)=e^{-\pi i/12}\phi_r(\tau)$. With the proof of Proposition 3.8 below, using (1-8), Theorem 1.1 now follows. □

We are left to show the existence of the moment functions and their derivatives.

Proposition 3.8. *For $r,n\in\mathbb{N}_0$, the moment functions*

$$\frac{\partial^n}{\partial\tau^n}\big[q^{-1/24}u_{2r}(q)\big]$$

are defined for every root of unity $q=\zeta$ and lie in $\mathbb{Z}[\zeta]$.

Proof. For ease of notation, we let

$$D_\alpha := \alpha\frac{\partial}{\partial\alpha},$$
$$J_m(w;q) := (wq;q)_m\big(w^{-1}q;q\big)_m.$$

To finish the proof it is enough to show that for m sufficiently large, and every $n,r\in\mathbb{N}_0$, the function

$$(3\text{-}15)\qquad D_q^n\big(D_w^r[J_m(w;q)]_{w=1}\big)$$

vanishes for $q=\zeta$.

It is not difficult to see that for $m\in\mathbb{N}$,

$$(3\text{-}16)\qquad \frac{D_w(J_m(w;q))}{J_m(w;q)} = -\sum_{k=1}^m\frac{wq^k}{1-wq^k}+\sum_{k=1}^m\frac{w^{-1}q^k}{1-w^{-1}q^k} =: R_m(w;q).$$

We further relax notation and let $J := J_m(w;q)$, $R := R_m(w;q)$, and $R^{(r)} := D_w^r R$ for $r \in \mathbb{N}_0$. Using (3-16), we find that

$$\begin{aligned}
D_w J &= JR,\\
D_w^2 J &= J\big(R^2 + R^{(1)}\big),\\
D_w^3 J &= J\big(R^3 + 3RR^{(1)} + R^{(2)}\big),\\
D_w^4 J &= J\big(R^4 + 4RR^{(2)} + 3\big(R^{(1)}\big)^2 + 6R^2R^{(1)} + R^{(3)}\big),\\
&\ \ \vdots
\end{aligned}$$

Note that each $D_w^r J$ can be expressed as J multiplied by a sum over the partitions of r. That is, given a partition $\pi = \ell_1(\pi)\cdot 1 + \ell_2(\pi)\cdot 2 + \cdots + \ell_{r-1}(\pi)\cdot(r-1) + \ell_r(\pi)\cdot r$ of r (where each $\ell_j(\pi) \in \mathbb{N}_0$), we may assign the product

$$\prod_{1\le j\le r} \big(D_w^{j-1}R\big)^{\ell_j(\pi)}.$$

Conversely, every such product appearing as a summand as above for $D_w^r J$ corresponds to a partition of r. In general, we have

$$D_w^r[J_m(w;q)]_{w=1} = (q;q)_m^2 \sum_{\pi\vdash r} c(\pi) \prod_{1\le j\le r} \big(D_w^{j-1}[R_m(w;q)]_{w=1}\big)^{\ell_j(\pi)},$$

where we sum over all partitions π of r. The exponents $\ell_j(\pi)$ correspond to the number of parts of the partition π of r, and the constants $c(\pi) = c_r(\pi)$ also depend on the partition π of r. Now using the definition of $R_m(w;q)$ in (3-16), we may write

$$\sum_{\pi\vdash r} c(\pi) \prod_{1\le j\le r} \big(D_w^{j-1}[R_m(w;q)]_{w=1}\big)^{\ell_j(\pi)} = \sum_{\vec{k}=(k_1,\ldots,k_c)} \frac{P_{\vec{k},r}(q)}{\prod_{j=1}^{c}(1-q^{k_j})^r} =: R_{m,r}(q), \tag{3-17}$$

where $c = c_r \in \mathbb{N}$ depends only on r, and $P_{\vec{k},r} \in \mathbb{Z}[q]$. Next we apply the operator D_q^n to $(q;q)_m^2$ multiplied by $R_{m,r}(q)$ in (3-17) above. Using the product rule, we have (3-15) equals

$$\sum_{0\le j\le n} \binom{n}{j} D_q^j\big((q;q)_m^2\big) D_q^{n-j}(R_{m,r}(q)).$$

It is not difficult to see that

$$\frac{D_q\big((q;q)_m^2\big)}{(q;q)_m^2} = -2\sum_{k=1}^{m} \frac{kq^k}{1-q^k} =: T_m(q),$$

and for $l \in \mathbb{N}$, that

$$D_q^{\ell-1}(T_m(q)) = \sum_{k=1}^{m} \frac{Q_{k,l}(q)}{(1-q^k)^\ell},$$

with $Q_{k,l}(q) \in \mathbb{Z}[q]$. Therefore, we may conclude that (3-15) has the shape

$$(q;q)_m^2 \sum_{\vec{k}=(k_1,\ldots,k_d)} \frac{P_{\vec{k},r,n}(q)}{\prod_{j=1}^{d}\left(1-q^{k_j}\right)^{r+n}},$$

where $d = d_{r,n} \in \mathbb{N}$ depends only on r and n, and $P_{\vec{k},r,n} \in \mathbb{Z}[q]$. Now if $\zeta = \zeta_m$ then $(q;q)_M^2$ ($M \in \mathbb{N}$) vanishes at $q = \zeta$ of order $\geq 2\lfloor m/M \rfloor$. On the other hand, each term

$$\frac{P_{\vec{k},r,n}(q)}{\prod_{j=1}^{d}\left(1-q^{k_j}\right)^{r+n}}$$

vanishes at $q = \zeta$ of order at most $d(r+n)$, which is a constant independent of m. Thus, the claim follows. □

4. Proof of Theorem 1.2

To prove Theorem 1.2, we recall (3-2). It is not difficult to see from Proposition 3.2 that

$$Y^+(z;it) = -H(z;it) + \sum_{r\geq 0} \beta_r(t) z^r$$

with

$$\beta_r(t) \ll_r e^{-N/t}$$

for some $N > 0$. To find the asymptotic expansion of $H(z;it)$, we split as in (3-9) and bound using (2-3)

$$h_{1,2r}(it) \ll e^{-M/t}$$

for some $M > 0$. Thus we are left to determine the asymptotic expansion of $H_2(z;it)$. For this, we write

$$\begin{aligned} H_2(z;it) &= -\frac{i}{\sqrt{3}} \int_{\mathbb{R}} e^{-\pi t w^2/3 - 2\pi w z} \frac{\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\, dw \\ &= -\frac{i}{\sqrt{3}} \sum_{r=0}^{\infty} \frac{(-2\pi z)^{2r+1}}{(2r+1)!} \sum_{k=0}^{\infty} \frac{(-\pi t/3)^k}{k!} \int_{\mathbb{R}} \frac{w^{2r+2k+1}\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\, dw, \end{aligned} \tag{4-1}$$

where the identity in (4-1) refers to an asymptotic expansion. Thus, to determine the asymptotic expansion of $H_2(z; it)$, we are left to evaluate explicitly for $a \in \mathbb{N}_0$,

$$\begin{aligned}\mathcal{C}_a &:= \int_{\mathbb{R}} \frac{w^{2a+1}\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\, dw \\ &= \frac{1}{2}\int_{\mathbb{R}} \frac{w^{2a+1}\left(e^{2\pi w/3}-e^{-2\pi w/3}\right)}{\cosh(\pi w)}\, dw = \sum_{r=1}^{\infty} \frac{\left(\frac{2\pi}{3}\right)^{2r-1}}{(2r-1)!}\int_{\mathbb{R}} \frac{w^{2a+2r}}{\cosh(\pi w)}\, dw.\end{aligned}$$

From (2-8), we have that the integral above equals $(2i)^{-2a-2r}E_{2a+2r}$, yielding

$$\mathcal{C}_a = (-2i)^{-2a-1}\sum_{r=1}^{\infty}\frac{\left(\frac{\pi i}{3}\right)^{2r-1}}{(2r-1)!}E_{2a+2r} = (-2i)^{-2a-1}\sum_{r=0}^{\infty}\frac{\left(\frac{\pi i}{3}\right)^{r}}{r!}E_{2a+r+1}.$$

The second equality above holds because $E_j = 0$ for j odd.

We are thus left to understand $\sum_{r=0}^{\infty}(v^r/r!)E_{r+b}$ for positive integers b and $v = \pi i/3$. Set

$$f(v) := \sum_{r=0}^{\infty}\frac{E_r}{r!}v^r = \operatorname{sech}(v),$$

where the second equality above is simply the definition of the Euler numbers. Then

$$f^{(b)}(v) = \sum_{r=0}^{\infty}\frac{E_{r+b}}{r!}v^r.$$

Thus

$$\mathcal{C}_a = (-2i)^{-2a-1}\operatorname{sech}^{(2a+1)}\left(\frac{\pi i}{3}\right) = 2^{-2a-1}\sec^{(2a+1)}\left(\frac{\pi}{3}\right). \tag{4-2}$$

Next we deduce from (1-12) that

$$\frac{i}{2}\frac{1}{\sin(\pi z)} = -\sum_{n=0}^{\infty}\frac{B_{2n}\left(\frac{1}{2}\right)}{(2n)!}(2\pi i z)^{2n-1}.$$

Combining the above, we have established that the asymptotic expansion of $\mathcal{U}(z; it)e^{\pi t/12}$ as $t \to 0^+$ is given by

$$\frac{1}{\sqrt{3}}\sum_{r=0}^{\infty}(2\pi i z)^{2r}(-1)^r\sum_{0\le n\le r}\frac{B_{2n}\left(\frac{1}{2}\right)}{(2n)!}\frac{(-1)^n}{(2r-2n+1)!}\sum_{k=0}^{\infty}\frac{\left(-\pi t/3\right)^k}{k!}\mathcal{C}_{r-n+k}.$$

Thus, using (4-2), we have the asymptotic expansion as $t \to 0^+$,

$$\begin{aligned}e^{\pi t/12}u_{2r}\left(e^{-2\pi t}\right) &= \frac{(2r)!(-1)^r 2^{-2r-1}}{\sqrt{3}}\sum_{k=0}^{\infty}\frac{t^k}{k!}\left(-\frac{\pi}{3}\right)^k 2^{-2k} \\ &\quad\times\sum_{0\le n\le r}\frac{(-1)^n B_{2n}\left(\frac{1}{2}\right)2^{2n}}{(2n)!(2r-2n+1)!}\sec^{(2r-2n+2k+1)}\left(\frac{\pi}{3}\right).\end{aligned} \tag{4-3}$$

Using Lemma 2.6 together with (4-3), we have

$$
\begin{aligned}
&e^{\pi t/12} u_{2r}\left(e^{-2\pi t}\right) \\
&\quad = \frac{3^{2r+1}}{2r+1} \sum_{k=0}^{\infty} \frac{(3\pi t)^k}{k!} \sum_{0\le n\le r} \binom{2r+1}{2n} 3^{-2n} B_{2n}\left(\frac{1}{2}\right) E_{2r+2k+1-2n}\left(\frac{5}{6}\right),
\end{aligned}
\tag{4-4}
$$

which concludes the proof of the first statement of Theorem 1.2.

Next we prove the claimed asymptotic for the main term. Since $B_{2n+1}\left(\frac{1}{2}\right) = 0$, we may rewrite the $k = 0$ summand of (4-4) as

$$
\frac{3^{2r+1}}{2r+1} \sum_{0\le n\le 2r+1} \binom{2r+1}{n} 3^{-n} B_n\left(\frac{1}{2}\right) E_{2r+1-n}\left(\frac{5}{6}\right). \tag{4-5}
$$

Now we use (2-6), which yields that

$$
B_n\left(\frac{1}{2}\right) = 3^{n-1} \sum_{a=0}^{2} B_n\left(\frac{1}{6} + \frac{a}{3}\right).
$$

Thus, (4-5) equals

$$
\frac{3^{2r}}{2r+1} \sum_{a=0}^{2} \sum_{0\le n\le 2r+1} \binom{2r+1}{n} B_n\left(\frac{1}{6} + \frac{a}{3}\right) E_{2r+1-n}\left(\frac{5}{6}\right). \tag{4-6}
$$

Using (2-7), (4-6) reduces to

$$
\frac{2\cdot 6^{2r}}{2r+1} \sum_{a=0}^{2} B_{2r+1}\left(\frac{1}{2} + \frac{a}{6}\right).
$$

Noting again that $B_{2r+1}\left(\frac{1}{2}\right) = 0$, we find that as claimed, as $t \to 0^+$,

$$
e^{\pi t/12} u_{2r}\left(e^{-2\pi t}\right) \sim \frac{2\cdot 6^{2r}}{2r+1} \left(B_{2r+1}\left(\frac{2}{3}\right) + B_{2r+1}\left(\frac{5}{6}\right)\right).
$$

5. An example: the moment function $\phi_1(\tau)$

In this section, we give an exact value for the quantum moment function

$$
\begin{aligned}
\phi_1(\tau) &= 4\pi^3 i q^{-1/24} \sum_{n=1}^{\infty} \sum_{m\in\mathbb{Z}} (-1)^m u(m,n)\left(m^2 + 2n\right) q^n \\
&= 4\pi^3 i q^{-1/24} \left(u_2(q) - i\pi^{-1} \frac{\partial}{\partial\tau} u_0(q)\right).
\end{aligned}
\tag{5-1}
$$

To describe this, we define for positive integers n the polynomials

$$d_n(q) := n(q;q)_{n-1}^2 q^n - 2q^{n+2}(q;q)_n \sum_{j=1}^{n} j q^{j-1} \prod_{\substack{k=1\\k\neq j}}^{n} \left(1-q^k\right) \in \mathbb{Z}[q], \tag{5-2}$$

$$b_n(q) := q^{n+1} \sum_{j=1}^{n} q^j \prod_{\substack{k=1\\k\neq j}}^{n} \left(1-q^k\right)^2 \in \mathbb{Z}[q]. \tag{5-3}$$

Theorem 5.1. *If* $h, k \in \mathbb{N}$, *with* $\gcd(h,k) = 1$, *we have*

$$\phi_1\left(\frac{h}{k}\right) = 8\pi^3 i \zeta_{24k}^{-h} \left(\sum_{n=1}^{k} d_n\left(\zeta_k^h\right) - \sum_{n=1}^{2k-1} b_n\left(\zeta_k^h\right) \right).$$

Remark. Theorem 5.1, together with (1-11) in the case $\tau \in \mathbb{Q}\setminus\{0\}$ of Theorem 1.1, gives an exact value for the integral

$$\int_{\mathbb{R}} P_1\left(w, (-i\tau)^{-1}\right) e^{\pi i \tau w^2/3} \frac{\sinh\left(\frac{2\pi w}{3}\right)}{\cosh(\pi w)}\, dw.$$

To prove Theorem 5.1, we first establish Propositions 5.2 and 5.3 below. These propositions give alternate expressions for the functions defining $\phi_1(\tau)$ (see (5-1)), which we subsequently evaluate for $q = \zeta$, where ζ is any root of unity.

Proposition 5.2. *With notation as above, we have*

$$\frac{\partial}{\partial\tau} u_0(q) = 2\pi i \sum_{n\geq 1} d_n(q).$$

Moreover, if $\gcd(h,k) = 1$, *we have*

$$\frac{\partial}{\partial\tau} [u_0(q)]_{q=\zeta_k^h} = 2\pi i \sum_{n=1}^{k} d_n\left(\zeta_k^h\right).$$

Proof. The first statement follows by straightforward differentiation, using that $u_0(q) = \mathcal{U}(0;\tau)$, definition (1-2), and the fact that $1/(2\pi i)(\partial/\partial\tau) = q(d/dq)$. To prove the second statement, we observe that $d_n(q)$ is of the form $d_n(q) = (q;q)_{n-1}\tilde{d}_n(q)$, where $\tilde{d}_n(\zeta_k^h) < \infty$. The statement now follows, observing that for $n \geq k+1$, the factor $(q;q)_{n-1}$ of $d_n(q)$ vanishes when $q = \zeta_k^h$. □

Proposition 5.3. *With notation as above, we have*

$$(2\pi i)^2 u_2(q) = \frac{\partial^2}{\partial z^2} [\mathcal{U}(z;\tau)]_{z=0} = -2(2\pi i)^2 \sum_{n\geq 1} b_n(q).$$

Moreover, if $h, k \in \mathbb{N}$, *with* $\gcd(h,k) = 1$, *we have*

$$(2\pi i)^2 u_2\big(\zeta_k^h\big) = -2(2\pi i)^2 \sum_{n=1}^{2k-1} b_n\big(\zeta_k^h\big).$$

Proof. The first statement follows by straightforward differentiation, using definition (1-2), and the fact that $1/(2\pi i)(\partial/\partial z) = w(d/dw)$ for $w = e^{2\pi i z}$. To prove the second statement, using the first statement, we see for $n \geq 2k$, the j-th summand defining $b_n(q)$ (for any $j \geq 1$) contains either the factor $(1-q^k)$ or $(1-q^{2k})$ (or both), both of which vanish when $q = \zeta_k^h$. □

Proof of Theorem 5.1. Theorem 5.1 now follows from the definition of $\phi_1(\tau)$ (see (5-1)), Propositions 5.2 and 5.3. □

Acknowledgements

The authors thank the referee for helpful comments and suggestions. Bringmann's research was supported by the Alfried Krupp Prize for Young University Teachers of the Krupp foundation and the research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant agreement n. 335220-AQSER. Folsom is grateful for the support of National Science Foundation CAREER grant DMS-1252815, and the Max Planck Institute for Mathematics in Bonn, Germany.

References

[Abramowitz and Stegun 1964] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series **55**, U.S. Government Printing Office, Washington, DC, 1964. Reprinted by Dover, New York, 1974. MR 29 #4914 Zbl 0171.38503

[Alfes et al. 2011] C. Alfes, K. Bringmann, and J. Lovejoy, "Automorphic properties of generating functions for generalized odd rank moments and odd Durfee symbols", *Math. Proc. Cambridge Philos. Soc.* **151**:3 (2011), 385–406. MR 2838344 Zbl 1287.11062

[Andrews 2005] G. E. Andrews, "Partitions with short sequences and mock theta functions", *Proc. Natl. Acad. Sci. USA* **102**:13 (2005), 4666–4671. MR 2006a:11131 Zbl 1207.11106

[Andrews 2007] G. E. Andrews, "Partitions, Durfee symbols, and the Atkin–Garvan moments of ranks", *Invent. Math.* **169**:1 (2007), 37–73. MR 2008d:05013 Zbl 1214.11116

[Andrews 2008] G. E. Andrews, "The number of smallest parts in the partitions of *n*", *J. Reine Angew. Math.* **624** (2008), 133–142. MR 2009m:11172 Zbl 1153.11053

[Andrews and Berndt 2009] G. E. Andrews and B. C. Berndt, *Ramanujan's lost notebook, Part II*, Springer, New York, 2009. MR 2010f:11002 Zbl 1180.11001

[Andrews et al. 2013] G. E. Andrews, R. C. Rhoades, and S. P. Zwegers, "Modularity of the concave composition generating function", *Algebra Number Theory* **7**:9 (2013), 2103–2139. MR 3152010 Zbl 1282.05016

[Atkin and Garvan 2003] A. O. L. Atkin and F. G. Garvan, "Relations between the ranks and cranks of partitions", *Ramanujan J.* **7**:1-3 (2003), 343–366. MR 2005e:11131 Zbl 1039.11069

[Atkin and Swinnerton-Dyer 1954] A. O. L. Atkin and H. P. F. Swinnerton-Dyer, "Some properties of partitions", *Proc. London Math. Soc.* (3) **4** (1954), 84–106. MR 15,685d Zbl 0055.03805

[Bringmann 2008] K. Bringmann, "On the explicit construction of higher deformations of partition statistics", *Duke Math. J.* **144**:2 (2008), 195–233. MR 2009e:11203 Zbl 1154.11034

[Bringmann and Folsom 2013] K. Bringmann and A. Folsom, "On the asymptotic behavior of Kac–Wakimoto characters", *Proc. Amer. Math. Soc.* **141**:5 (2013), 1567–1576. MR 3020844 Zbl 1277.11035

[Bringmann and Ono 2010] K. Bringmann and K. Ono, "Dyson's ranks and Maass forms", *Ann. of Math.* (2) **171**:1 (2010), 419–449. MR 2011e:11165 Zbl 1277.11096

[Bringmann et al. 2009] K. Bringmann, F. G. Garvan, and K. Mahlburg, "Partition statistics and quasiharmonic Maass forms", *Int. Math. Res. Not.* **2009**:1 (2009), 63–97. MR 2009j:11073 Zbl 1156.11021

[Bringmann et al. 2010] K. Bringmann, J. Lovejoy, and R. Osburn, "Automorphic properties of generating functions for generalized rank moments and Durfee symbols", *Int. Math. Res. Not.* **2010**:2 (2010), 238–260. MR 2011c:11159 Zbl 1230.05034

[Bringmann et al. $\geq$ 2015] K. Bringmann, K. Mahlburg, and R. C. Rhoades, "Peak positions of strongly unimodal sequences", In preparation.

[Bryson et al. 2012] J. Bryson, K. Ono, S. Pitman, and R. C. Rhoades, "Unimodal sequences and quantum and mock modular forms", *Proc. Natl. Acad. Sci. USA* **109**:40 (2012), 16063–16067. MR 2994899

[Cvijović 2009] D. Cvijović, "Closed-form formulae for the derivatives of trigonometric functions at rational multiples of π", *Appl. Math. Lett.* **22**:6 (2009), 906–909. MR 2010e:33001 Zbl 1231.33001

[Dabholkar et al. 2014] A. Dabholkar, S. Murthy, and D. Zagier, "Quantum black holes, wall crossing, and mock modular forms", preprint, 2014. Submitted for publication. arXiv 1208.4074

[Dyson 1944] F. J. Dyson, "Some guesses in the theory of partitions", *Eureka* **8** (1944), 10–15. MR 3077150

[Eichler and Zagier 1985] M. Eichler and D. Zagier, *The theory of Jacobi forms*, Progress in Mathematics **55**, Birkhäuser, Boston, 1985. MR 86j:11043 Zbl 0554.10018

[Erdélyi et al. 1981] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions, I*, Krieger, Melbourne, FL, 1981. Based on notes left by Harry Bateman. MR 84h:33001a Zbl 0051.30303

[Garvan 2011] F. G. Garvan, "Higher order spt-functions", *Adv. Math.* **228**:1 (2011), 241–265. MR 2012g:11185 Zbl 1268.11143

[Kac and Wakimoto 2001] V. G. Kac and M. Wakimoto, "Integrable highest weight modules over affine superalgebras and Appell's function", *Comm. Math. Phys.* **215**:3 (2001), 631–682. MR 2001j:17017 Zbl 0980.17002

[Manschot 2011] J. Manschot, "Wall-crossing of D4-branes using flow trees", *Adv. Theor. Math. Phys.* **15**:1 (2011), 1–42. MR 2888006 Zbl 06074658

[Mellit and Okada 2009] A. Mellit and S. Okada, "Joyce invariants for $K3$ surfaces and mock theta functions", *Commun. Number Theory Phys.* **3**:4 (2009), 655–676. MR 2011f:14019 Zbl 1219.14025

[Rademacher 1973] H. Rademacher, *Topics in analytic number theory*, edited by E. Grosswald et al., Die Grundlehren der mathematischen Wissenschaften **169**, Springer, New York, 1973. MR 51 #358 Zbl 0253.10002

[Ramanujan 1919] S. Ramanujan, "Some properties of $p(n)$, the number of partitions of n", *Proc. Cambridge Philos. Soc.* **19** (1919), 207–210. MR 2280868 JFM 47.0885.01

[Shimura 1973] G. Shimura, "On modular forms of half integral weight", *Ann. of Math.* (2) **97** (1973), 440–481. MR 48 #10989 Zbl 0266.10022

[Vafa and Witten 1994] C. Vafa and E. Witten, "A strong coupling test of S-duality", *Nuclear Phys. B* **431**:1-2 (1994), 3–77. MR 95k:81138 Zbl 0964.81522

[Zagier 1991] D. Zagier, "Periods of modular forms and Jacobi theta functions", *Invent. Math.* **104**:3 (1991), 449–465. MR 92e:11052 Zbl 0742.11029

[Zagier 2001] D. Zagier, "Vassiliev invariants and a strange identity related to the Dedekind eta-function", *Topology* **40**:5 (2001), 945–960. MR 2002g:11055 Zbl 0989.57009

[Zagier 2010] D. Zagier, "Quantum modular forms", pp. 659–675 in *Quanta of maths* (Paris, 2007), edited by E. Blanchard et al., Clay Math. Proc. **11**, Amer. Math. Soc., Providence, RI, 2010. MR 2012a:11066 Zbl 05902011

[Zwegers 2002] S. P. Zwegers, *Mock theta functions*, thesis, Utrecht University, 2002. Zbl 1194.11058 arXiv 0807.4834

[Zwegers 2010] S. P. Zwegers, "Multivariable Appell functions", preprint, University College Dublin, 2010, Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.6121.

Received November 13, 2013. Revised April 25, 2014.

KATHRIN BRINGMANN
MATHEMATICAL INSTITUTE
UNIVERSITY OF COLOGNE
WEYERTAL, 86-90
D-50931 KÖLN
GERMANY

kbringma@math.uni-koeln.de

AMANDA FOLSOM
MATHEMATICS
YALE UNIVERSITY
P.O. BOX 208283
NEW HAVEN, CT 06520-8283
UNITED STATES

amanda.folsom@yale.edu

ROBERT C. RHOADES
CENTER FOR COMMUNICATIONS RESEARCH
805 BUNN DRIVE
PRINCETON, NJ 08540
UNITED STATES

rob.rhoades@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.27

CONGRUENCE PRIMES FOR IKEDA LIFTS AND THE IKEDA IDEAL

JIM BROWN AND RODNEY KEATON

Let f be a newform of level 1 and weight $(2\kappa - n)$ for positive even integers κ and n. We study congruence primes for the Ikeda lift of f. In particular, we consider a conjecture of Katsurada stating that primes dividing certain L-values of f are congruence primes for the Ikeda lift. Instead of focusing on a congruence to a single eigenform, we deduce a lower bound on the number of all congruences between the Ikeda lift of f and forms not lying in the space spanned by Ikeda lifts.

1. Introduction

Let κ be an integer and let χ be a Dirichlet character of conductor N satisfying $\chi(-1) = (-1)^{\kappa}$. One has an associated Eisenstein series $E_{\kappa,\chi}$. It is a well-known fact that for a prime $\ell \nmid N$ and a prime $\mathfrak{l}$ dividing ℓ in a suitably large extension of $\mathbb{Z}$ so that $\mathfrak{l} \mid L(1-\kappa, \chi)$ there exists a cuspidal eigenform f of level M with $N \mid M$ such that $f \equiv E_{\kappa,\chi} \pmod{\mathfrak{l}}$. Such congruences between cusp forms and Eisenstein series have been studied by many authors. For instance, one can use such congruences to make deductions on the structure of the residual Galois representation of the cusp form, which can then be used to study Selmer groups associated to the cusp form (see [Ribet 1976; Wiles 1990; Skinner and Urban 2014] for some prominent examples of this type of argument).

If we view the Eisenstein series as a "lift" of the Dirichlet character χ from GL(1) to GL(2), then we can fit the congruences mentioned above into a more general framework. Namely, one can consider more general automorphic forms and lift them to automorphic forms on other algebraic groups. This approach has also received considerable attention as it can also be used to study Selmer groups of higher-degree Galois representations; see [Skinner and Urban 2006; Klosin 2009;

Keaton was partially supported by the National Security Agency under grant H98230-11-1-0137.

The authors would like to thank the referee for numerous suggestions that improved the exposition of this paper.
MSC2010: primary 11F33; secondary 11F46, 11F30, 11F55.
Keywords: Ikeda lifts, Siegel modular forms, congruences between modular forms.

Agarwal and Brown 2014] for specific examples and [Mazur 2011] for a survey of this method. This makes classifying primes for which one will have a congruence between a lifted form and a nonlifted form a natural question to study. In this paper we investigate this problem for Ikeda lifts.

Let κ and n be positive even integers, $f \in S_{2\kappa-n}(\mathrm{SL}_2(\mathbb{Z}))$ a newform, and $I_n(f) \in S_\kappa(\mathrm{Sp}_{2n}(\mathbb{Z}))$ the Ikeda lift of f. Katsurada [2011] states a conjecture on when a prime $\mathfrak{l}$ will satisfy that there is an eigenform $F \in S_\kappa(\mathrm{Sp}_{2n}(\mathbb{Z}))$ that is not an Ikeda lift and is congruent to $I_n(f)$ modulo $\mathfrak{l}$. The conjecture is in terms of divisibilities of special values of L-functions of f by $\mathfrak{l}$. One can see Conjecture 9 for the precise statement. To provide evidence for his conjecture he proves that if a prime divides the required L-values and does not divide other L-values then one indeed does have such a congruence (see Theorem 10). In this paper we also consider Ikeda lifts, but instead of focusing on producing one congruence we introduce the Ikeda ideal. This ideal is an analogue of the Eisenstein ideal in the GL(2) case and measures congruences between $I_n(f)$ and all other eigenforms. We then show that under similar hypotheses as given in [Katsurada 2011], we can do better and bound from below the congruences between $I_n(f)$ and all other eigenforms that are not lifts. One can see Theorem 14 for the precise result.

One thing to note here is that while the Saito–Kurokawa lift is useful for studying the p-adic Bloch–Kato conjecture for the L-value $L_{\mathrm{alg}}(\kappa, f)$ due to the fact that the value $L_{\mathrm{alg}}(\kappa, f)$ "controls" the congruence between the Saito–Kurokawa lift and a nonlifted form (see [Brown 2011; Agarwal and Brown 2014] for example), the L-values that control the congruence for an Ikeda lift are given by $L_{\mathrm{alg}}(\kappa, f)\prod_{j=1}^{n/2-1} L_{\mathrm{alg}}(2j+1, \mathrm{ad}^0 f)$. This indicates that if one knew the existence of Galois representations for automorphic forms on GSp_{2n}, as well as expected properties of these representations, one could use the congruence results produced in this paper to study the ℓ-adic Bloch–Kato conjecture not only for $L_{\mathrm{alg}}(\kappa, f)$, but also for the values $L_{\mathrm{alg}}(2j+1, \mathrm{ad}^0 f)$ when $j = 1, \dots, \frac{n}{2}-1$. This makes such congruences particularly interesting.

The structure of the paper is as follows. Section 2 recalls the basic definitions we will need throughout the paper. We recall the Ikeda lift and some necessary properties in Section 3. In Section 4 we state Katsurada's conjecture and result, introduce the Ikeda ideal, and show how Katsurada's congruence can be recovered by studying the Ikeda ideal. We then state our main result and discuss the major hypotheses in Section 5. Section 6 gives a somewhat detailed description of an Eisenstein series originally introduced by Shimura and some results needed to prove the main theorem. Finally, we conclude by proving the main theorem in Section 7.

Throughout the paper ℓ denotes an odd prime. We fix once and for all an algebraic closure $\overline{\mathbb{Q}}$ of the rationals and $\overline{\mathbb{Q}}_\ell$ of $\mathbb{Q}_\ell$. We also fix compatible embeddings $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_\ell \hookrightarrow \mathbb{C}$.

2. Modular forms

In this section we recall the basics on modular forms and Siegel modular forms that will be needed throughout the rest of the paper.

2.1. ***Basic definitions.*** Given a ring R with identity, we write $\mathrm{Mat}_n(R)$ for the ring of $n \times n$ matrices with entries in R.

Set

$$J_n = \begin{pmatrix} 0_n & -1_n \\ 1_n & 0_n \end{pmatrix}$$

and recall that the degree-n symplectic group is defined by

$$G_n = \mathrm{GSp}_{2n} = \{g \in \mathrm{GL}_{2n} : {}^t g J_n g = \mu_n(g) J_n,\ \mu_n(g) \in \mathrm{GL}_1\}.$$

We set $\mathrm{Sp}_{2n} = \ker \mu_n$. We denote $\mathrm{Sp}_{2n}(\mathbb{Z})$ by Γ_n to ease notation. We say that $\Gamma \subset \Gamma_n$ is a congruence subgroup if it contains $\Gamma^{(n)}(N)$ as a subgroup of finite index for some integer $N \geq 1$, where

$$\Gamma^{(n)}(N) = \{\gamma \in \Gamma_n : \gamma \equiv 1_{2n} \pmod N\}.$$

Given a matrix $z \in \mathrm{Mat}_n(\mathbb{C})$, we can write $z = x + \sqrt{-1}y$ for $x, y \in \mathrm{Mat}_n(\mathbb{R})$. When we write $z = x + \sqrt{-1}y$, we will always mean $x, y \in \mathrm{Mat}_n(\mathbb{R})$. The Siegel upper half-space is given by

$$\mathfrak{h}_n = \{z = x + \sqrt{-1}y \in \mathrm{Mat}_n(\mathbb{C}) : {}^t z = z,\ y > 0\}.$$

We have an action of $G_n^+(\mathbb{R}) = \{g \in G_n(\mathbb{R}) : \mu_n(g) > 0\}$ on $\mathfrak{h}_n$ given by

$$gz = (a_g z + b_g)(c_g z + d_g)^{-1} \quad \text{for } g = \begin{pmatrix} a_g & b_g \\ c_g & d_g \end{pmatrix},$$

where $a_g, b_g, c_g, d_g \in \mathrm{Mat}_n(\mathbb{R})$.

For $g \in G_n^+(\mathbb{R})$ and $z \in \mathfrak{h}_n$, we set

$$j(g, z) = \det(c_g z + d_g).$$

Let κ be a positive integer. Given $f : \mathfrak{h}_n \to \mathbb{C}$, we define the slash operator on f by

$$(f|_\kappa g)(z) = \mu_n(g)^{n\kappa/2} j(g, z)^{-\kappa} f(gz).$$

Let $\Gamma \subset \Gamma_n$ be a congruence subgroup. We say that such an f is a genus-n Siegel modular form of weight κ and level Γ if f is holomorphic and satisfies

$$(f|_\kappa \gamma)(z) = f(z)$$

for all $\gamma \in \Gamma$. If $n = 1$ we also require that f is holomorphic at the cusps so that we recover the theory of elliptic modular forms. We denote the space of genus-n, level-Γ, and weight-κ modular forms by $M_\kappa(\Gamma)$.

Let $f \in M_\kappa(\Gamma)$ and let $\gamma \in G_n^+(\mathbb{Q})$. Then $f|_\kappa\gamma$ has a Fourier expansion of the form

$$(f|_\kappa\gamma)(z) = \sum_{T \in \Lambda_n} a_{f|_\kappa\gamma}(T) e(\mathrm{Tr}(Tz)),$$

where Λ_n is defined to be the set of $n \times n$ half-integral positive semidefinite symmetric matrices and $e(w) := e^{2\pi i w}$. We say that f is a cusp form if for all $T \in \Lambda_n$ with $\det(T) = 0$ we have $a_{f|_\kappa\gamma}(T) = 0$ for all $\gamma \in G_n^+(\mathbb{Q})$. We write $S_\kappa(\Gamma)$ for the cusp forms in $M_\kappa(\Gamma)$. Given a ring $R \subset \mathbb{C}$, we write $M_\kappa(\Gamma; R)$ for those modular forms whose Fourier coefficients all lie in R and likewise for the cusp forms.

Let $f_1, f_2 \in M_\kappa(\Gamma)$ with at least one of them a cusp form. The Petersson inner product of f_1 and f_2 is defined by

$$\langle f_1, f_2 \rangle_\Gamma = \int_{\Gamma\backslash\mathfrak{h}_n} f_1(z)\overline{f_2(z)}(\det y)^\kappa \, d\mu z,$$

where $z = x + iy$ with $x = (x_{\alpha,\beta})$, $y = (y_{\alpha,\beta}) \in \mathrm{Mat}_n(\mathbb{R})$, and

$$d\mu z = (\det y)^{-(n+1)} \prod_{\alpha \leq \beta} dx_{\alpha,\beta} \prod_{\alpha \leq \beta} dy_{\alpha,\beta},$$

with $dx_{\alpha,\beta}$ and $dy_{\alpha,\beta}$ the usual Lebesgue measure on $\mathbb{R}$. We will use the following scaled definition that is independent of the congruence subgroup considered:

$$\langle f_1, f_2 \rangle = \frac{1}{[\overline{\Gamma}_n : \overline{\Gamma}]} \langle f_1, f_2 \rangle_\Gamma,$$

where $\overline{\Gamma}_n = \Gamma_n / \{\pm 1_{2n}\}$ and $\overline{\Gamma}$ is the image of Γ in $\overline{\Gamma}_n$.

2.2. *Hecke algebras.* Let $\Gamma \subset \Gamma_n$ be a congruence subgroup. Given $g \in G_n^+(\mathbb{Q})$, we write $T(g)$ to denote the double coset $\Gamma g \Gamma$. We define the usual action of $T(g)$ on Siegel modular forms by setting

$$T(g)f = \sum_i f|_\kappa g_i,$$

where $\Gamma g \Gamma = \coprod_i \Gamma g_i$ and $f \in M_\kappa(\Gamma)$. Let p be prime and define

$$T^{(n)}(p) = T(\mathrm{diag}(1_n, p1_n)),$$

and for $i = 1, \dots, n$, set

$$T_i^{(n)}(p^2) = T(\mathrm{diag}(1_{n-i}, p1_i, p^2 1_{n-i}, p1_i)).$$

These Hecke operators generate the local Siegel Hecke algebra at p [van der Geer 2008, Theorem 9].

Let $\mathcal{H}_{\mathbb{Z}}^{(n)}$ denote the $\mathbb{Z}$-subalgebra of $\mathrm{End}_{\mathbb{C}}(S_\kappa(\Gamma))$ generated by $T^{(n)}(p)$ and $T_i^{(n)}(p^2)$ for $i = 1, \ldots, n$. Given any $\mathbb{Z}$-algebra A, we write $\mathcal{H}_A^{(n)}$ for $\mathcal{H}_{\mathbb{Z}}^{(n)} \otimes_{\mathbb{Z}} A$.

Let E be a finite extension of $\mathbb{Q}_\ell$ and $\mathcal{O}_E$ the ring of integers of E. Then $\mathcal{H}_{\mathcal{O}_E}^{(n)}$ is a semilocal complete finite $\mathcal{O}_E$-algebra. One has

$$\mathcal{H}_{\mathcal{O}_E}^{(n)} = \prod_{\mathfrak{m}} \mathcal{H}_{\mathfrak{m}}^{(n)}, \tag{1}$$

where the product runs over all maximal ideals of $\mathcal{H}_{\mathcal{O}_E}^{(n)}$ and $\mathcal{H}_{\mathfrak{m}}^{(n)}$ denotes the localization of $\mathcal{H}_{\mathcal{O}_E}^{(n)}$ at $\mathfrak{m}$.

2.3. *Congruences.* Let $f, g \in M_\kappa(\Gamma_n; K)$, with $K \subseteq \overline{\mathbb{Q}}_\ell$. Let $\mathcal{O}$ denote the ring of integers of K with $\mathfrak{l}$ the prime of $\mathcal{O}$. We write

$$f \equiv g \pmod{\mathfrak{l}^b}$$

to denote

$$\mathrm{val}_{\mathfrak{l}}(a_f(T) - a_g(T)) \geq b$$

for all $T \in \Lambda_n$. We refer to this as a congruence of Fourier coefficients.

We will also use the notion of a congruence of eigenvalues. Let $f, g \in M_\kappa(\Gamma_n)$ be eigenforms and now suppose that $K/\mathbb{Q}_\ell$ is the minimal extension containing all Hecke eigenvalues of f and g. Note that this is a finite extension by [Kurokawa 1981, Theorem 1]. Furthermore, by the remark in [Mizumoto 1991, Section 2] we may assume that f, g are normalized so that the Fourier coefficients are also contained in K. We shall assume throughout the remainder of the paper that all eigenforms are normalized in this way.

In this case, if f and g are eigenforms for all $t \in \mathcal{H}_{\mathcal{O}}^{(n)}$ with eigenvalues $\lambda_f(t)$ and $\lambda_g(t)$, respectively, we write

$$f \equiv_{\mathrm{ev}} g \pmod{\mathfrak{l}^b}$$

to denote

$$\mathrm{val}_{\mathfrak{l}}(\lambda_f(t) - \lambda_g(t)) \geq b$$

for all $t \in \mathcal{H}_{\mathcal{O}}^{(n)}$.

2.4. *L-functions.* In this section we introduce the L-functions that will be needed in this paper. In the case of the relevant L-functions attached to elliptic modular forms, we also introduce the appropriate canonical periods.

Given local Euler factors $L_p(s)$ and a finite set of primes Σ, we define

$$L^\Sigma(s) = \prod_{p \notin \Sigma} L_p(s).$$

If $\Sigma = \{p \mid N\}$ we write $L^N(s)$ for $L^\Sigma(s)$. We set $L(s) = L^\varnothing(s)$.

We begin with the case of an elliptic modular form $f \in S_\kappa(\Gamma_1)$. We assume that f is a normalized eigenform with Fourier expansion

$$f(z) = \sum_{n \geq 1} a_f(n) e(nz).$$

Let $\pi_f = \bigotimes'_p \pi_{f,p}$ be the automorphic representation associated to f. For each prime p there exists a character σ_p such that $\pi_{f,p} = \pi(\sigma_p, \sigma_p^{-1})$, where $\pi(\sigma_p, \sigma_p^{-1})$ is the principal series representation of $\mathrm{GL}_2(\mathbb{Q}_p)$. The p-Satake parameter of f is given by $\alpha_0(p; f) = \sigma_p(p)$. We will drop the f from the notation when it is clear from context. The L-function of f is given by

$$\begin{aligned} L(s, f) &= \prod_p (1 - \alpha_0(p) p^{-s+(\kappa-1)/2})^{-1} (1 - \alpha_0(p)^{-1} p^{-s+(\kappa-1)/2})^{-1} \\ &= \prod_p (1 - a_f(p) p^{-s} + p^{\kappa-1-2s})^{-1} = \sum_{n \geq 1} a_f(n) n^{-s}. \end{aligned}$$

Given a Dirichlet character χ, we will also make use of the twisted L-function

$$L(s, f, \chi) = \sum_{n \geq 1} \chi(n) a_f(n) n^{-s}.$$

Let $\ell \geq \kappa$ be a prime and let K be a suitably large finite extension of $\mathbb{Q}_\ell$ with ring of integers $\mathcal{O}$. Let $f \in S_\kappa(\Gamma_1; \mathcal{O})$ be a normalized eigenform. Let $\rho_{f,\ell}$ be the ℓ-adic Galois representation associated to f and assume that the residual representation $\bar{\rho}_{f,\ell}$ is irreducible. Then we have canonical complex periods $\Omega_f^\pm$ (determined up to ℓ-units) by [Vatsal 1999]. Vatsal showed that such periods exist for level greater than 3, but using arguments in [Hida 1987] we can define $\Omega_f^\pm$ for arbitrary level. One can see [Brown 2007] for more details. Using these periods we have:

Theorem 1 [Shimura 1977; Vatsal 1999]. *Let $f \in S_\kappa(\Gamma_1; \mathcal{O})$ be as in the above discussion. There exist complex periods $\Omega_f^\pm$ such that for each integer m with $0 < m < \kappa$ and every Dirichlet character χ one has*

$$\frac{L(m, f, \chi)}{\tau(\chi)(2\pi\sqrt{-1})^m} \in \begin{cases} \Omega_f^+ \mathcal{O}_\chi & \text{if } \chi(-1) = (-1)^m, \\ \Omega_f^- \mathcal{O}_\chi & \text{if } \chi(-1) = (-1)^{m-1}, \end{cases}$$

where $\tau(\chi)$ is the Gauss sum of χ and $\mathcal{O}_\chi$ is the extension of $\mathcal{O}$ generated by the values of χ.

With this theorem in mind we set the following notation for the algebraic part of $L(m, f, \chi)$ with $0 < m < \kappa$:

$$L_{\mathrm{alg}}(m, f, \chi) := \frac{L(m, f, \chi)}{\tau(\chi)(2\pi\sqrt{-1})^m \Omega_f^\pm},$$

where the choice of period is from the theorem.

For Siegel modular forms of genus greater than 1 there are two relevant L-functions: the standard and spinor L-functions. Let $f \in S_\kappa(\Gamma_n)$ be an eigenform. Associated to f is a cuspidal automorphic representation π_f of $\mathrm{PGSp}_{2n}(\mathbb{A})$. We can decompose π_f into local components as $\pi_f = \bigotimes' \pi_{f,p}$, with $\pi_{f,p}$ an Iwahori spherical representation of $\mathrm{PGSp}_{2n}(\mathbb{Q}_p)$. We refer the reader to [Asgari and Schmidt 2001, Section 3] for the details concerning the construction of cuspidal automorphic representations associated to Siegel cusp forms. The representation $\pi_{f,p}$ is given as $\pi(\chi_0, \chi_1, \dots, \chi_n)$ for χ_i unramified characters of $\mathbb{Q}_p^\times$. One can see [Asgari and Schmidt 2001, Section 3.2] for the definition of this spherical representation. Let $\alpha_0(p; f) = \chi_0(p), \dots, \alpha_n(p; f) = \chi_n(p)$ denote the p-Satake parameters of f. Note these are normalized so that

$$\alpha_0(p; f)^2 \alpha_1(p; f) \cdots \alpha_n(p; f) = 1.$$

We drop f and/or p in the notation for the Satake parameters when they are clear from context. Set $\tilde{\alpha}_0 = p^{(2n\kappa - n(n+1))/4}\alpha_0$ and

$$L_p(X, f; \mathrm{spin}) = (1 - \tilde{\alpha}_0 X) \prod_{j=1}^{n} \prod_{1 \le i_1 \le \cdots \le i_j \le n} (1 - \tilde{\alpha}_0 \alpha_{i_1} \cdots \alpha_{i_j} X).$$

The spinor L-function associated to f is given by

$$L(s, f; \mathrm{spin}) = \prod_p L_p(p^{-s}, f; \mathrm{spin})^{-1}.$$

One should note that in the case that f is an elliptic modular form the spinor L-function is exactly $L(s, f)$ defined above.

Set

$$L_p(X, f; \mathrm{st}) = (1 - X) \prod_{i=1}^{n} (1 - \alpha_i(p) X)(1 - \alpha_i(p)^{-1} X).$$

Then, we define the standard L-function associated to f by

$$L(s, f; \mathrm{st}) = \prod_p L_p(p^{-s}, f; \mathrm{st})^{-1}.$$

Given a Hecke character χ, the twisted standard L-function is given by

$$L(s, f, \chi; \mathrm{st}) = \prod_p L_p(\chi(p) p^{-s}, f; \mathrm{st})^{-1}.$$

In the case that $f \in S_\kappa(\Gamma_1; \mathcal{O})$ is an elliptic modular form the standard L-function is usually denoted by $L(s, \mathrm{ad}^0 f)$, i.e., it is the adjoint L-function. Then the corollary to [Zagier 1977, Theorem 2] gives that

$$\frac{L(m, \mathrm{ad}^0 f)}{\pi^{2m+\kappa-1} \Omega_f^+ \Omega_f^-} \in \overline{\mathbb{Q}}$$

for $m = 1, 3, \dots, \kappa - 1$ and

$$\frac{L(m, \mathrm{ad}^0 f)}{\pi^{m+\kappa-1}\Omega_f^+\Omega_f^-} \in \overline{\mathbb{Q}}$$

for $m = 2-\kappa, 4-\kappa, \dots, 0$. We will only be interested in the first case; we denote this algebraic value by $L_{\mathrm{alg}}(m, \mathrm{ad}^0 f)$.

3. The Ikeda lift

In this section we will present an introduction to the Ikeda lift. For the details the reader is referred to [Kohnen 2002] or Ikeda's original paper [2001]. The Ikeda lift can be viewed as a composition of the Shintani map from the space of elliptic modular forms to the space of half-integral weight modular forms and a map from the space of half-integral weight forms to the correct space of Siegel modular forms.

Throughout we assume κ, n to be positive even integers with $2\kappa - n > 1$. We note here that we begin with weight $2\kappa - n$ instead of 2κ as used in [Ikeda 2001; Kohnen 2002]. This normalization is more convenient for our purposes.

Recall the algebraic version of Shintani's lift that we require. One has:

Theorem 2 [Shintani 1975]. *There is a linear function*

$$\theta_{\kappa,n} : S_{2\kappa-n}(\Gamma_1) \to S^+_{\kappa-\frac{n}{2}+\frac{1}{2}}(\Gamma_0(4))$$

that is Hecke equivariant, i.e., one has $\theta_{\kappa,n}(f \mid T(p)) = \theta_{\kappa,n}(f) \mid T(p^2)$ *for any prime* p.

The next result will be pivotal for the algebraic construction.

Proposition 3 [Stevens 1994, Proposition 2.3.1]. *Let* $f \in S_{2\kappa-n}(\Gamma_1; \mathcal{O})$ *be a Hecke eigenform, where* $\mathcal{O}$ *is the ring of integers of a field that can be embedded in* $\mathbb{C}$. *Then there is a nonzero complex number* $\Omega(f) \in \mathbb{C}^\times$ *so that*

$$\frac{1}{\Omega(f)}\theta_{\kappa,n}(f) \in S^+_{\kappa-\frac{n}{2}+\frac{1}{2}}(\Gamma_0(4); \mathcal{O}).$$

Moreover, if $\mathcal{O}$ *is a discrete valuation ring then* $\Omega(f)$ *can be chosen so that at least one of the Fourier coefficients of* $(1/\Omega(f))\theta_{\kappa,n}(f)$ *is a unit in* $\mathcal{O}$.

From now on we write $\theta^{\mathrm{alg}}_{\kappa,n}(f)$ for $(1/\Omega(f))\theta_{\kappa,n}(f)$ and will always choose the period so that $\theta^{\mathrm{alg}}_{\kappa,n}(f)$ has a unit Fourier coefficient in the case that $\mathcal{O}$ is a discrete valuation ring. We write

$$\theta^{\mathrm{alg}}_{\kappa,n}(f)(z) = \sum_{\substack{m>0 \\ m\equiv 0,1 \pmod 4}} c(m)e(mz).$$

Let $T > 0$ be in Λ_n, i.e., T is an $n \times n$ half-integral positive definite symmetric matrix. Set D_T to be the determinant of $2T$, Δ_T the absolute value of the discriminant of $\mathbb{Q}(\sqrt{D_T})$, χ_T the primitive Dirichlet character associated to $\mathbb{Q}(\sqrt{D_T})/\mathbb{Q}$, and $\mathfrak{f}_T$ the rational number satisfying $D_T = \Delta_T \mathfrak{f}_T^2$.

Let $S_n(R)$ denote the set of symmetric $n \times n$ matrices over a ring R. For a rational prime p, let $\psi_p : \mathbb{Q}_p \to \mathbb{C}^\times$ be the unique additive character given by

$$\psi_p(x) = \exp(-\{x\}_p),$$

where $\{x\}_p \in \mathbb{Z}\left[\frac{1}{p}\right]$ is the p-adic fractional part of x. The Siegel series for T is

$$b_p(T, s) := \sum_{S \in S_n(\mathbb{Q}_p)/S_n(\mathbb{Z}_p)} \psi_p(\mathrm{Tr}(TS))\, p^{-s \operatorname{ord}_p(\nu(S))} \quad \text{for } \mathrm{Re}(s) \gg 0,$$

where $\nu(S) := \det(S_1) \cdot \mathbb{Z}_p$, and S_1 is from the factorization $S = S_1^{-1} S_2$ for a symmetric coprime pair of matrices S_1, S_2. We have a factorization of the Siegel series

$$b_p(T, s) = \gamma_p(T, p^{-s}) F_p(T, p^{-s}),$$

where

$$\gamma_p(T, X) = \frac{1 - X}{1 - p^{\frac{n}{2}} \chi_T(p) X} \prod_{i=1}^{\frac{n}{2}} (1 - p^{2i} X^2),$$

and $F_p(T, X) \in \mathbb{Z}[X]$ has constant term 1 and $\deg(F_p(T, X)) = 2 \operatorname{ord}_p(\mathfrak{f}_T)$. Using this polynomial $F_p(T, X)$ we define

$$\tilde{F}_p(T, X) := X^{-\operatorname{ord}_p(\mathfrak{f}_T)} F_p(T, p^{-\frac{n}{2} - \frac{1}{2}} X).$$

For each $T > 0$ in Λ_n, define

$$a(T) = c(|\Delta_T|) \mathfrak{f}_T^{\kappa - \frac{n+1}{2}} \prod_p \tilde{F}_p(T, \alpha_0(p; f)), \tag{2}$$

and form the series

$$I_n(f)(z) = \sum_{T > 0} a(T) e(\mathrm{Tr}(Tz)),$$

where $\alpha_0(p; f)$ is the p-th Satake parameter of f. Then we have:

Theorem 4 [Ikeda 2001, Theorems 3.2 and 3.3]. *The series $I_n(f)(z)$, referred to as the Ikeda lift of f, is an eigenform in $S_\kappa(\Gamma_n)$ whose standard L-function factors as*

$$L(s, F; \mathrm{st}) = \zeta(s) \prod_{i=1}^{n} L(s + \kappa - i, f).$$

We will also need further information about the integrality of the Fourier coefficients of $I_n(f)$. In particular, the following result is essential to our applications.

Theorem 5 [Kohnen 2002, Theorem 1]. *Let $\theta_{\kappa,n}^{\mathrm{alg}}(f)$ be as above and let $a(T)$ be as in (2). Then*

$$a(T)=\sum_{d\mid\mathfrak{f}_T} d^{\kappa-1}\phi(d;T)c(|\Delta_T|(\mathfrak{f}_T/d)^2),$$

where $\phi(d;T)$ is an integer-valued function.

As an immediate consequence of this theorem and Proposition 3 we have:

Corollary 6. *Let $f\in S_{2\kappa-n}(\Gamma_1;\mathcal{O})$ be a Hecke eigenform, where $\mathcal{O}$ is the ring of integers of a field that can be embedded in $\mathbb{C}$. Then $I_n(f)$ has Fourier coefficients in $\mathcal{O}$.*

We will also make use of the following result.

Proposition 7 [Katsurada 2011, Proposition 4.6]. *Let $f\in S_{2\kappa-n}(\Gamma_1)$ be a normalized eigenform with Ikeda lift $I_n(f)$. Let $\mathcal{O}$ be the ring of integers of a field that can be embedded in $\mathbb{C}$ and let $\mathfrak{l}$ be a prime in $\mathcal{O}$. If there is a fundamental discriminant D such that the D-th Fourier coefficient of $\theta_{\kappa,n}^{\mathrm{alg}}(f)$ is not divisible by $\mathfrak{l}$, then there is a Fourier coefficient of $I_n(f)$ that is not divisible by $\mathfrak{l}$. In particular, if $\mathcal{O}$ is the ring of integers of some $K\subset\overline{\mathbb{Q}}_\ell$ with prime $\mathfrak{l}$, then $I_n(f)$ has a Fourier coefficient that is a unit modulo $\mathfrak{l}$.*

Proof. The only thing to prove is the last statement, but this follows immediately from our normalization of $\theta_{\kappa,n}^{\mathrm{alg}}$. □

Let $f_1,\dots,f_r$ be an orthogonal basis of $S_{2\kappa-n}(\Gamma_1)$ consisting of normalized eigenforms. We denote the span of $I_n(f_1),\dots,I_n(f_r)$ in $S_\kappa(\Gamma_n)$ by $S_\kappa^{\mathrm{Ik}}(\Gamma_n)$. We denote the orthogonal complement of $S_\kappa^{\mathrm{Ik}}(\Gamma_n)$ in $S_\kappa(\Gamma_n)$ with respect to the Petersson product by $S_\kappa^{\text{N-Ik}}(\Gamma_n)$.

4. A conjecture of Katsurada and the Ikeda ideal

In this section we present a conjecture of Katsurada on the congruence primes of Ikeda lifts. We then introduce the Ikeda ideal and show how one can use the Ikeda ideal to study all the congruences between $I_n(f)$ and forms in $S_\kappa^{\text{N-Ik}}(\Gamma_n)$ at once. This allows us to prove a stronger congruence result under roughly the same conditions as given in [Katsurada 2011].

We fix some notation used throughout this section. Let K denote a number field, $\mathcal{O}_K$ the ring of integers of K, and $\mathfrak{l}$ a prime of $\mathcal{O}_K$ of residue characteristic ℓ. We let $\mathcal{O}$ be the completion of $\mathcal{O}_K$ at $\mathfrak{l}$ and let λ denote a uniformizer of $\mathfrak{l}$ in $\mathcal{O}$.

4.1. *A conjecture of Katsurada.*

Definition 8. Let $F\in S_\kappa(\Gamma_n;\mathcal{O})$ be an eigenform. We say $\mathfrak{l}$ is a congruence prime of F with respect to $V\subset(\mathbb{C}F)^\perp$ if there exists an eigenform $G\in V$ such that $F\equiv_{\mathrm{ev}} G \pmod{\mathfrak{l}}$. (Note that in order for this congruence to make sense we may need to extend K so that $G\in S_\kappa(\Gamma_n;\mathcal{O})$ as well.)

One should note this definition can be extended to levels other than Γ_n, but we will have no need of such a definition in this paper.

Let $f \in S_{2\kappa-n}(\Gamma_1)$ be a normalized eigenform. Katsurada's conjecture states that all of the primes dividing certain special values of L-functions of f are congruence primes for the Ikeda lift $I_n(f)$ with respect to the space $S_\kappa^{\mathrm{Ik}}(\Gamma_n)^\perp$.

Conjecture 9 [Katsurada 2011, Conjecture A]. *Let $\kappa > n$ be integers and let $f = f_1, f_2, \dots, f_r \in S_{2\kappa-n}(\Gamma_1; \mathcal{O})$ be a basis of normalized eigenforms. Assume $\ell \nmid (2\kappa-1)!$. Then $\mathfrak{l}$ is a congruence prime of $I_n(f)$ with respect to $S_\kappa^{\mathrm{Ik}}(\Gamma_n)^\perp$ if*

$$\mathfrak{l} \,\Big|\, L_{\mathrm{alg}}(\kappa, f) \prod_{i=1}^{\frac{n}{2}-1} L_{\mathrm{alg}}(2i+1, \mathrm{ad}^0 f).$$

As evidence for this conjecture, Katsurada proves the following theorem.

Theorem 10 [ibid., Theorem 4.7]. *Let $\mathcal{O}$, f, and $\mathfrak{l}$ be as in the conjecture with $\kappa > 2n+4$. Then $\mathfrak{l}$ is a congruence prime for $I_n(f)$ with respect to $S_\kappa^{\mathrm{Ik}}(\Gamma_n)^\perp$ if the following are satisfied*:

$$\mathfrak{l} \,\Big|\, L_{\mathrm{alg}}(\kappa, f) \prod_{i=1}^{\frac{n}{2}-1} L_{\mathrm{alg}}(2i+1, \mathrm{ad}^0 f). \tag{1}$$

(2) *For some integer m satisfying $\frac{n}{2} < m < \frac{\kappa}{2} - \frac{n}{2}$ and some fundamental discriminant D satisfying $(-1)^{\frac{n}{2}} D > 0$,*

$$\mathfrak{l} \nmid D(m-1)!\,\zeta_{\mathrm{alg}}(2m) L_{\mathrm{alg}}\Big(\kappa - \frac{n}{2}, \chi_D\Big) \prod_{i=1}^{n} L_{\mathrm{alg}}(2m+\kappa-i, f),$$

where $\zeta_{\mathrm{alg}}(2m) = \zeta(2m)/\pi^{2m}$.

(3) *For a constant $C_{\kappa,n} := \prod_{j \le (2\kappa-n)/12} (1 + j + \cdots + j^{n-1})$ if $n > 2$ and $C_{\kappa,2} = 1$,*

$$\mathfrak{l} \nmid \frac{C_{\kappa,n} \langle f, f\rangle}{\Omega_f^+ \Omega_f^-}.$$

As noted by Katsurada, the second condition allows freedom to choose m, so it is reasonable to expect that one can find an m with $\mathfrak{l} \nmid \zeta_{\mathrm{alg}}(2m) \prod_{i=1}^n L_{\mathrm{alg}}(2m+\kappa-i, f)$ in many cases.

4.2. ***The Ikeda ideal: definition and bounds.*** The conjecture in the previous subsection gives conditions when one will have a congruence between an Ikeda lift $I_n(f)$ and a form in $S_\kappa^{\mathrm{N\text{-}Ik}}(\Gamma_n)$. In this section we will introduce the Ikeda ideal associated to $I_n(f)$ that will capture this information as well. In fact, the ideal captures more information as it measures all congruences between $I_n(f)$ and forms in $S_\kappa^{\mathrm{N\text{-}Ik}}(\Gamma_n)$.

Let f be a normalized eigenform in $S_{2\kappa-n}(\Gamma_1; \mathcal{O})$ and $I_n(f)$ the Ikeda lift of f. Recall that the Hecke algebra over $\mathcal{O}$ acting on $S_\kappa(\Gamma_n)$ is denoted by $\mathcal{H}_{\mathcal{O}}^{(n)}$.

Let $X \subseteq S_\kappa^{\mathrm{Ik}}(\Gamma_n)$ be a Hecke-stable subspace and let Y be the orthogonal complement in $S_\kappa(\Gamma_n)$ to X under the Petersson product. In particular, the examples we will be interested in are when $X = \mathbb{C}I_n(f)$ or $X = S_\kappa^{\mathrm{Ik}}(\Gamma_n)$. Let $\mathcal{H}_{\mathcal{O}}^{(n),Y}$ denote the image of $\mathcal{H}_{\mathcal{O}}^{(n)}$ in $\mathrm{End}_{\mathbb{C}}(Y)$ and let $\phi : \mathcal{H}_{\mathcal{O}}^{(n)} \to \mathcal{H}_{\mathcal{O}}^{(n),Y}$ denote the natural surjection.

We let $\mathrm{Ann}(I_n(f))$ denote the annihilator of $I_n(f)$ in $\mathcal{H}_{\mathcal{O}}^{(n)}$. We have that $I_n(f)$ induces an $\mathcal{O}$-algebra homomorphism $\mathcal{H}_{\mathcal{O}}^{(n)} \to \mathcal{O}$ by sending a Hecke operator to its eigenvalue. Since this is an $\mathcal{O}$-algebra homomorphism it is surjective and it clearly has kernel $\mathrm{Ann}(I_n(f))$. Thus, there is an isomorphism

$$\mathcal{H}_{\mathcal{O}}^{(n)} / \mathrm{Ann}(I_n(f)) \cong \mathcal{O}.$$

Using that ϕ is surjective we have that $\phi(\mathrm{Ann}(I_n(f)))$ is an ideal in $\mathcal{H}_{\mathcal{O}}^{(n),Y}$. We refer to this ideal as the *Ikeda ideal associated to $I_n(f)$ with respect to Y* and denote it by $\mathcal{I}_n^Y(f)$. We will be interested in the index of this ideal. In particular, one has that there exists an integer m such that

$$\mathcal{H}_{\mathcal{O}}^{(n),Y} / \mathcal{I}_n^Y(f) \cong \mathcal{O}/\mathfrak{l}^m\mathcal{O}.$$

We give here two elementary propositions to relate this index to Katsurada's conjecture.

Proposition 11. *With the notation as above, if there exists $G \in Y$, not necessarily an eigenform, such that*

$$I_n(f) \equiv G \pmod{\mathfrak{l}^b},$$

then $m \geq b$.

Proof. Assume that $b > m$, and consider the diagram

$$\begin{tikzcd}
\mathcal{H}_{\mathcal{O}}^{(n)} \arrow[r, "\phi"] \arrow[d] & \mathcal{H}_{\mathcal{O}}^{(n),Y} \arrow[d] \\
\mathcal{H}_{\mathcal{O}}^{(n)} / \mathrm{Ann}(I_n(f)) \arrow[r, "\phi"] \arrow[d, "\simeq"] & \mathcal{H}_{\mathcal{O}}^{(n),Y} / \mathcal{I}_n^Y(f) \arrow[d, "\simeq"] \\
\mathcal{O} \arrow[r] & \mathcal{O}/\mathfrak{l}^m\mathcal{O}
\end{tikzcd}$$

Each map here is an $\mathcal{O}$-algebra surjection. Let $t \in \phi^{-1}(\lambda^m) \subset \mathcal{H}_{\mathcal{O}}^{(n)}$. Then by definition we have $tG = \lambda^m G$. Moreover, by the commutativity of the diagram we see that $t \in \mathrm{Ann}(I_n(f))$, so the assumed congruence gives

$$\lambda^m G \equiv 0 \pmod{\mathfrak{l}^b},$$

i.e.,

$$G \equiv 0 \pmod{\mathfrak{l}^{b-m}}.$$

However, since we are assuming $b > m$, this gives

$$I_n(f) \equiv G \equiv 0 \pmod{\mathfrak{l}}.$$

This contradicts Proposition 7, and so it must be that $b \leq m$. □

Proposition 12. *With the notation as above, suppose $m \geq 1$. Then there exists an eigenform $G \in Y$ such that*

$$I_n(f) \equiv_{\text{ev}} G \pmod{\mathfrak{l}}.$$

Proof. Extend K if necessary so that $I_n(f) \in S_\kappa(\Gamma_n; \mathcal{O})$ and we have an orthogonal basis $F_1, \ldots, F_r$ of Y with each $F_i \in S_\kappa(\Gamma_n; \mathcal{O})$. Suppose that there are no eigenforms $G \in Y$ eigenvalue-congruent to $I_n(f)$.

Let $\mathcal{S}$ denote the $\mathbb{C}$-vector space spanned by $I_n(f), F_1, \ldots, F_r$. Let $\mathcal{H}_{\mathcal{O}}^{(n),S}$ denote the image of the Hecke algebra $\mathcal{H}_{\mathcal{O}}^{(n)}$ in $\mathrm{End}_{\mathbb{C}}(\mathcal{S})$. For each eigenform $F \in \mathcal{S}$ with eigenvalues in $\mathcal{O}$ we obtain a maximal ideal $\mathfrak{m}_F$ of $\mathcal{H}_{\mathcal{O}}^{(n),S}$ given as the kernel of the map $\mathcal{H}_{\mathcal{O}}^{(n),S} \to \mathcal{O}/\mathfrak{l}\mathcal{O} : t \mapsto \lambda_F(t) \pmod{\mathfrak{l}}$. We have that eigenforms F and G are eigenvalue-congruent modulo $\mathfrak{l}$ if and only if $\mathfrak{m}_F = \mathfrak{m}_G$.

We now use the fact that $I_n(f)$ is not congruent to any of $F_1, \ldots, F_r$ to conclude that

$$\mathcal{H}_{\mathcal{O}}^{(n),S} = \mathcal{H}_{\mathfrak{m}_{I_n(f)}}^{(n),S} \times \prod_{\mathfrak{m}} \mathcal{H}_{\mathfrak{m}}^{(n),S},$$

where the product is over the maximal ideals corresponding to $F_1, \ldots, F_r$. However, this gives that $\mathcal{I}_n^Y(f) = \prod_{\mathfrak{m}} \mathcal{H}_{\mathfrak{m}}^{(n),S}$, and this is exactly $\mathcal{H}_{\mathcal{O}}^{(n),Y}$. This contradicts the assumption that $m \geq 1$. Thus, it must be that there is a congruence as desired. □

To match the previous results with Katsurada's, simply take $X = S_\kappa^{\mathrm{Ik}}(\Gamma_n)$ and $Y = S_\kappa^{\text{N-Ik}}(\Gamma_n)$. In fact, one has that the index of the Ikeda ideal measures all congruences between forms in Y and $I_n(f)$. This follows from Proposition 13. One should note that we use the fact that the space of Ikeda lifts satisfies multiplicity one [Ikeda 2013, Theorem 7.1] in order to apply this result.

Proposition 13 [Berger et al. $\geq$ 2015, Proposition 4.3]. *Let X and Y be as above and let $F_1, \ldots, F_r$ be a basis of Y. For each $1 \leq i \leq r$, let m_i be the largest integer so that*

$$I_n(f) \equiv_{\text{ev}} F_i \pmod{\mathfrak{l}^{m_i}}.$$

Then one has

$$\frac{1}{e}\sum_{i=1}^{r} m_i \geq \mathrm{val}_\lambda\big(\#\mathcal{H}_{\mathcal{O}}^{(n),Y}/\mathcal{I}_n^Y(f)\big),$$

where e is the ramification index of $\mathcal{O}$ over $\mathbb{Z}_\ell$.

Thus, one can view results giving a lower bound on the Ikeda ideal as a strengthening of the results of [Katsurada 2011], where one is only concerned with a congruence modulo $\mathfrak{l}$ to a single eigenform.

5. Main results

We now state the main result of this paper. The proof will be given in Section 7. After stating the theorem, we discuss the main hypotheses.

Theorem 14. *Let κ and n be positive even integers with $\kappa > n+1$ and let ℓ be a prime so that $\ell > 2\kappa - n$. Assume $\ell \nmid \prod_{p \leq (2\kappa-n)/12}(1+p+\cdots+p^{n-1})$. Let $f \in S_{2\kappa-n}(\Gamma_1)$ be a newform and let $\mathcal{O}$ be a suitably large finite extension of $\mathbb{Z}_\ell$ that contains all the eigenvalues of f. Let $\mathfrak{l}$ denote the prime of $\mathcal{O}$. Furthermore, assume that $\bar{\rho}_{f,\ell}$ is irreducible and $\mathrm{val}_{\mathfrak{l}}(\langle f, f\rangle/(\Omega_f^+\Omega_f^-)) = 0$. We make these assumptions:*

(1) *There exists an integer $N > 1$ prime to ℓ and a Dirichlet character χ of conductor N with $\chi(-1) = (-1)^\kappa$ such that*
$$\mathrm{val}_{\mathfrak{l}}\biggl(L^N(n-\kappa+1, \chi)\prod_{j=1}^{n} L_{\mathrm{alg}}^N(n+1-j, f, \chi)\biggr) = 0.$$

(2) *There exists a fundamental discriminant D prime to ℓ such that $(-1)^{n/2}D > 0$, $\chi_D(-1) = -1$, and*
$$\mathrm{val}_{\mathfrak{l}}\Bigl(L_{\mathrm{alg}}\Bigl(\kappa - \frac{n}{2}, f, \chi_D\Bigr)\Bigr) = 0.$$

(3) *We have*
$$\mathrm{val}_{\mathfrak{l}}\biggl(L_{\mathrm{alg}}(\kappa, f)\prod_{j=1}^{\frac{n}{2}-1} L_{\mathrm{alg}}(2j+1, \mathrm{ad}^0 f)\biggr) = b > 0.$$

Then we have
$$\mathrm{val}_{\mathfrak{l}}\bigl(\#\mathcal{H}_{\mathcal{O}}^{(n),Y}/\mathcal{I}_n^Y(f)\bigr) \geq b,$$
where Y is the orthogonal complement of $X = \mathbb{C}I_n(f)$ in $S_\kappa(\Gamma_n)$.

Corollary 15. *With the same setup and assumptions as in Theorem 14, if $F_1, \ldots, F_r$ is a basis of eigenforms of $S_\kappa^{\text{N-Ik}}(\Gamma_n; \mathcal{O})$ (where we enlarge $\mathcal{O}$ if necessary) and if we let m_i be the largest integer so that*
$$I_n(f) \equiv_{\mathrm{ev}} F_i \pmod{\mathfrak{l}^{m_i}},$$
then we have
$$\frac{1}{e}\sum_{i=1}^{r} m_i \geq b,$$
where e is the ramification index of $\mathcal{O}$ over $\mathbb{Z}_\ell$.

Proof. Let $F_1, \dots, F_r$ be a basis of eigenforms of $S_\kappa^{\text{N-Ik}}(\Gamma_n; \mathcal{O})$ and $F_{r+1}, \dots, F_s$, $I_n(f)$ a basis of $S_\kappa^{\text{Ik}}(\Gamma_n)$. For $i = 1, \dots, s$ let m_i be the largest integer so that

$$I_n(f) \equiv_{\text{ev}} F_i \pmod{\mathfrak{l}^{m_i}}.$$

Theorem 14 and Proposition 13 give

$$\frac{1}{e}\sum_{i=1}^{s} m_i \geq b.$$

However, we have $m_{r+1} = \dots = m_s = 0$ as the assumption $\operatorname{val}_\mathfrak{l}(\langle f, f\rangle/(\Omega_f^+\Omega_f^-)) = 0$ guarantees that there are no eigenvalue congruences between $I_n(f)$ and other Ikeda lifts by the proof of [Katsurada 2011, Theorem 4.7]. Thus, we obtain the result. □

We first discuss the hypotheses that $\operatorname{val}_\mathfrak{l}(\langle f, f\rangle/(\Omega_f^+\Omega_f^-)) = 0$. This condition is equivalent to assuming that there are no other normalized eigenforms $g \in S_{2\kappa-n}(\Gamma_1; \mathcal{O})$ that are eigenvalue-equivalent to f modulo $\mathfrak{l}$. One can see [Hida 1981; Ribet 1983] for further discussion. For a particular f this condition can be easily checked [Bosma et al. 1997; Stein et al. 2013].

The two hypotheses we focus on are the ones concerning the $\mathfrak{l}$-indivisibility of L-values. We begin with the assumption that $\operatorname{val}_\mathfrak{l}\big(L_{\text{alg}}\big(\kappa - \frac{n}{2}, f, \chi_D\big)\big) = 0$. Note this is a central critical value since the weight of f is $2\kappa - n$. There have been several results on the $\mathfrak{l}$-divisibility of this particular special value due to its relation with the Fourier coefficients of the half-integral weight modular form $\theta_{\kappa,n}^{\text{alg}}(f)$. For example, Corollary 3 of [Bruinier and Ono 2003] shows that for nonexceptional primes ℓ there is a period Ω of f with the property that for infinitely many fundamental discriminants D prime to ℓ with $(-1)^{n/2}D > 0$ one has

$$\operatorname{ord}_\mathfrak{l}\left(\frac{D^{\kappa-\frac{n}{2}-\frac{1}{2}}L_{\text{alg}}\big(\kappa - \frac{n}{2}, f, \chi_D\big)}{\Omega}\right) = 0.$$

Since we assume $\bar{\rho}_{f,\ell}$ is irreducible, ℓ is automatically a nonexceptional prime for f [Swinnerton-Dyer 1973, Corollary 2]. However, we are unable to apply this result in our situation as the period Ω used is not the canonical period Ω_f^+ that we are using to normalize the L-value. We are unaware of any period relation between Ω and Ω_f^+. However, this does reduce the consideration to another period ratio; and since we have already assumed that $\mathfrak{l}$ does not divide a period ratio, this assumption is a reasonable one as well.

We next consider $L(n - \kappa + 1, \chi)$. Let p be a prime with $p \neq \ell$, $m \geq 1$ and φ be a Dirichlet character. In this setting Washington [1978] proves that for all but finitely many Dirichlet characters ψ of p-power conductor with $\varphi\psi(-1) = (-1)^m$,

$$\operatorname{val}_\mathfrak{l}(L(1 - m, \varphi\psi)/2) = 0.$$

In our setup we can take $m=\kappa-n$, $\chi=\varphi\psi$, and observe that $\chi(-1)=(-1)^{\kappa}=(-1)^{\kappa-n}$ to see there are infinitely many χ so that

$$\mathrm{val}_{\mathfrak{l}}(L(n-\kappa+1,\chi))=0.$$

If this were the only L-value controlled by χ we would be able to remove the hypothesis regarding this L-value. However, we also require that

$$\mathrm{val}_{\mathfrak{l}}\left(\prod_{j=1}^{n} L_{\mathrm{alg}}^{N}(n+1-j,f,\chi)\right)=0.$$

This means that we must choose a χ so that all of these L-values are simultaneously $\mathfrak{l}$-adic units. This is a much more delicate issue. We note here that we have a great deal of freedom in choosing such a χ, namely, the only restrictions concern the parity of χ and that its conductor be prime to ℓ. Thus, we have infinitely many characters to choose from so it is reasonable to expect that one can often find such a χ. In the case $n=2$, i.e., when one considers Saito–Kurokawa lifts, one can find computational evidence supporting the existence of such a χ in [Agarwal and Brown 2013]. One can use the same methods to produce computational evidence for $n>2$.

6. Siegel Eisenstein series

In this section we recall the definition of a Siegel Eisenstein series associated to a character. Following Shimura we then make a suitable choice of a section so that the Fourier coefficients of the Eisenstein series can be computed. Finally, we consider the pullback of our Siegel Eisenstein series and recall an inner product formula of Shimura. Throughout this section we assume that κ and n are even integers with $\kappa>n+1$.

6.1. ***Siegel Eisenstein series — general setup.*** Let P_n be the Siegel parabolic subgroup of G_n given by $P_n=\{g\in G_n : c_g=0\}$. We have that P_n factors as $P_n=N_{P_n}M_{P_n}$, where N_{P_n} is the unipotent radical

$$N_{P_n}=\left\{n(x)=\begin{pmatrix}1_n & x\\ 0_n & 1_n\end{pmatrix} : {}^{t}x=x,\, x\in \mathrm{Mat}_n\right\}$$

and M_{P_n} is the Levi subgroup

$$M_{P_n}=\left\{\begin{pmatrix}A & 0_n\\ 0_n & \alpha({}^{t}A)^{-1}\end{pmatrix} : A\in \mathrm{GL}_n,\, \alpha\in \mathrm{GL}_1\right\}.$$

Let $\mathbb{A}$ denote the rational adeles. Fix an idele class character χ and consider the induced representation

$$I(\chi)=\mathrm{Ind}_{P_n(\mathbb{A})}^{G_n(\mathbb{A})}(\chi)=\bigotimes_{v} I_v(\chi_v)$$

consisting of smooth functions $\mathfrak{f}$ on $G_n(\mathbb{A})$ that satisfy

$$\mathfrak{f}(pg) = \chi(\det(A_p))\mathfrak{f}(g) \quad \text{for } p = \begin{pmatrix} A_p & B_p \\ 0 & D_p \end{pmatrix} \in P_n(\mathbb{A}),\ g \in G_n(\mathbb{A}).$$

For $s \in \mathbb{C}$ and $\mathfrak{f} \in I(\chi)$ define

$$\mathfrak{f}(pg, s) = \chi(\det(A_p))|\det(A_p D_p^{-1})|^s \mathfrak{f}(g).$$

For v a place of $\mathbb{Q}$ we define $I_v(\chi_v)$ and $\mathfrak{f}_v(pg, s)$ analogously. We associate to such a section the Siegel Eisenstein series

$$E_{\mathbb{A}}(g, s; \mathfrak{f}) = \sum_{\gamma \in P_n(\mathbb{Q}) \backslash G_n(\mathbb{Q})} \mathfrak{f}(\gamma g, s). \tag{3}$$

Observe that $E_{\mathbb{A}}(g, s; \mathfrak{f})$ converges absolutely and uniformly for (g, s) on compact subsets of $G_n(\mathbb{A}) \times \{s \in \mathbb{C} : \mathrm{Re}(s) > (n+1)/2\}$. One can see [Shimura 1997, Appendix A.3] for this fact. Moreover, (3) defines an automorphic form on $G_n(\mathbb{A})$ and a holomorphic function on $\{s \in \mathbb{C} : \mathrm{Re}(s) > 0\}$ with meromorphic continuation to $\mathbb{C}$ with at most finitely many poles. Furthermore, Langlands [1976] gives a functional equation for $E_{\mathbb{A}}(g, s; \mathfrak{f})$ relating the value at $(n+1)/2 - s$ to the value at s.

6.2. ***A choice of section.*** For our applications we need to restrict the possible χ and pick a particular section $\mathfrak{f}$. Let $N > 1$ be an integer.

Let $\chi = \bigotimes_v \chi_v$ be an idele class character of $\mathbb{Q}$ that satisfies

$$\chi_\infty(x) = \left(\frac{x}{|x|}\right)^\kappa,$$
$$\chi_p(x) = 1 \quad \text{if } p \nmid \infty,\ x \in \mathbb{Z}_p^\times,\ \text{and } x \equiv 1 \pmod{N}.$$

For each finite prime p, we set

$$K_{0,p}^{(n)}(N) = \{g \in G_n(\mathbb{Q}_p) : a_g, b_g, d_g \in \mathrm{Mat}_n(\mathbb{Z}_p), c_g \in \mathrm{Mat}_n(N\mathbb{Z}_p)\}.$$

From this definition it is immediate that if $p \nmid N$ we have

$$K_{0,p}^{(n)}(N) = G_n(\mathbb{Q}_p) \cap \mathrm{Mat}_{2n}(\mathbb{Z}_p).$$

At the infinite place we put

$$K_\infty^{(n)} = \{g \in \mathrm{Sp}_{2n}(\mathbb{R}) : g(i_n) = i_n\}.$$

Set

$$K_0^{(n)}(N) = \prod_p K_{0,p}^{(n)}(N).$$

We choose our section $\mathfrak{f} = \bigotimes_v \mathfrak{f}_v$ as follows.

(1) We set $\mathfrak{f}_\infty$ to be the unique vector in $I_\infty(\chi_\infty, s)$ so that

$$\mathfrak{f}_\infty(k, \kappa) = j(k, i)^{-\kappa}$$

for all $k \in K_\infty^{(n)}$.

(2) If $p \nmid N$ we set $\mathfrak{f}_p$ to be the unique $K_{0,p}^{(n)}(N)$-fixed vector so that

$$\mathfrak{f}_p(1) = 1.$$

(3) If $p \mid N$ we set $\mathfrak{f}_p$ to be the vector given by

$$\mathfrak{f}_p(k) = \chi_p(\det(a_k)) \quad \text{for all } k \in K_{0,p}^{(n)}(N),\ k = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix}$$

and

$$\mathfrak{f}_p(g) = 0 \quad \text{for all } g \notin P_n(\mathbb{Q}_p) K_{0,p}^{(n)}(N).$$

The Eisenstein series $E_{\mathbb{A}}$ is the same as in [Shimura 1995; 1997].

Define

$$\Lambda_n^N(s, \chi) = L^N(2s, \chi) \prod_{i=1}^{[\frac{n}{2}]} L^N(4s - 2i, \chi^2)$$

and normalize $E_{\mathbb{A}}$ by setting

$$\mathbf{E}_{\mathbb{A}}(g, s; \mathfrak{f}) = \pi^{-n(n+2)/4} \Lambda_n^N(s, \chi) E_{\mathbb{A}}(g J_n^{-1}, s; \mathfrak{f}).$$

Set

$$G_\kappa^n(z; \mathfrak{f}) = \mathbf{E}_{\mathbb{A}}\left(\begin{pmatrix} y^{1/2} & x y^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix}, \frac{n+1}{2} - \frac{\kappa}{2}; \mathfrak{f} \right). \tag{4}$$

We have that $G_\kappa^n(z; \mathfrak{f})$ is a Siegel modular form of weight κ and level $\Gamma_0^{(n)}(N)$ [Shimura 1983], where

$$\Gamma_0^{(n)}(N) = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_n : C \equiv 0 \pmod{N} \right\}.$$

Write

$$G_\kappa^n(z; \mathfrak{f}) = \sum_{T \in \Lambda_n} a(T; \mathfrak{f}) e(\operatorname{Tr}(Tz)).$$

The Fourier coefficients $a(T; \mathfrak{f})$ are well known for this particular choice of section and normalization [Shimura 1997, Chapters 18 and 19]. In particular:

Theorem 16 [Brown 2007, Theorem 4.4]. *Let $\ell \geq n+1$ be an odd prime with $\ell \nmid N$. Then*

$$G_\kappa^n(z; \mathfrak{f}) \in M_\kappa\big(\Gamma_0^{(n)}(N); \mathbb{Z}_\ell\big[\chi, \sqrt{-1}^{\,n\kappa}\big]\big).$$

6.3. ***Pullbacks of Siegel Eisenstein series.*** Let $N > 1$ be an integer and $\ell > n+1$ a prime with $\ell \nmid N$.

Consider the diagonal embedding of $\mathfrak{h}^n \times \mathfrak{h}^n$ into $\mathfrak{h}^{2n}$ via the map

$$(z, w) \mapsto \mathrm{diag}[z, w] = \begin{pmatrix} z & 0 \\ 0 & w \end{pmatrix}.$$

We also have an embedding of $\Gamma_n \times \Gamma_n$ into Γ_{2n} given by

$$\left(\begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \right) \mapsto \begin{pmatrix} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{pmatrix}.$$

This allows us to view the natural action of $\Gamma_n \times \Gamma_n$ on $\mathfrak{h}^n \times \mathfrak{h}^n$ as a restriction of the action of Γ_{2n} on $\mathfrak{h}^{2n}$.

We will be interested in the restriction of the Eisenstein series $G_\kappa^{2n}(Z; \mathfrak{f})$ to $\mathfrak{h}^n \times \mathfrak{h}^n$. We refer to such a restriction as a pullback. These pullbacks have been considered in [Garrett 1984; Böcherer 1985; Garrett 1992; Shimura 1995; 1997]. In general, if F is a modular form of degree $2n$, level $\Gamma_0^{(2n)}(N)$, and weight κ, then the restriction of F to $\mathfrak{h}^n \times \mathfrak{h}^n$ is a modular form of degree n, level $\Gamma_0^{(n)}(N)$, and weight κ when considered as a function of z or w.

Shimura calculates the following set of representatives for $P_{2n} \backslash G_{2n}/(G_n \times G_n)$.

Lemma 17 [Shimura 1995, Lemma 4.2]. *For $0 \le r \le n$ let τ_r denote the element of G_{2n} given by*

$$\tau_r = \begin{pmatrix} 1_{2n} & 0 \\ \rho_r & 1_{2n} \end{pmatrix}, \quad \rho_r = \begin{pmatrix} 0_n & e_r \\ {}^t e_r & 0_n \end{pmatrix}, \quad e_r = \begin{pmatrix} 1_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Then the τ_r form a complete set of representatives for $P_{2n} \backslash G_{2n}/(G_n \times G_n)$.

We will make use of τ_n. Let $F \in S_\kappa(\Gamma_n)$ be an eigenform. We can specialize [ibid., Equation (6.17)] to obtain

$$\text{(5)} \quad \langle (G_\kappa^{2n} \mid \tau_n)(\mathrm{diag}[z, w]; \mathfrak{f}), F^c(w) \rangle = \mathcal{A}_{\kappa,n,N} \pi^{-n(n+1)/2} L(n+1-\kappa, F, \chi; \mathrm{st}) F(z),$$

where we have used $F \mid J_n = F$ since F has level Γ_n, and

$$\mathcal{A}_{\kappa,n,N} = \frac{2^{n(2\kappa-3n+2)/2}}{[\Gamma_n : \Gamma_0^{(n)}(N)]} \prod_{j=0}^{n-1} \frac{\Gamma((n-j)/2)}{\Gamma((2n+1-j)/2)}.$$

Since it will be important in the next section, we note that since $G_\kappa^{2n}(z; \mathfrak{f}) \in M_\kappa(\Gamma_0^{(2n)}(N); \mathbb{Z}_\ell[\chi])$, we have $(G_\kappa^{2n} \mid \tau_n)(z; \mathfrak{f}) \in M_\kappa(\tau_n^{-1} \Gamma_0^{(2n)}(N) \tau_n; \mathbb{Z}_\ell[\chi])$ by the q-expansion principle for Siegel modular forms [Chai and Faltings 1990, Proposition 1.5]. The Fourier expansion of $(G_\kappa^{2n} \mid \tau_n)(\operatorname{diag}[z, w]; \mathfrak{f})$ can be written as

$$(G_\kappa^{2n} \mid \tau_n)(\operatorname{diag}[z, w]; \mathfrak{f}) = \sum_{T_1, T_2 \in \Lambda_n} \Bigg(\sum_{T \in \Lambda_{2n}(T_1, T_2)} a(T; G_\kappa^{2n} \mid \tau_n) \Bigg) e(\operatorname{Tr}(T_1 z)) e(\operatorname{Tr}(T_2 w)),$$

where $a(T; G_\kappa^{2n} \mid \tau_n)$ is the T-th Fourier coefficient of $G_\kappa^{2n} \mid \tau_n$, and for $T_1, T_2 \in \Lambda_n$ we define

$$\Lambda_{2n}(T_1, T_2) = \left\{ T \in \Lambda_{2n} : T = \begin{pmatrix} T_1 & b \\ b & T_2 \end{pmatrix} \right\}.$$

This immediately gives that the Fourier coefficients of $(G_\kappa^{2n} \mid \tau_n)(\operatorname{diag}[z, w]; \mathfrak{f})$ lie in $\mathbb{Z}_\ell[\chi]$ as well.

7. Constructing a congruence

In this section we prove Theorem 14. We work under the hypotheses listed after the theorem. We again let $\mathcal{O}$ be a suitably large finite extension of $\mathbb{Z}_\ell$ with prime $\mathfrak{l}$ and uniformizer λ.

Our first step in constructing the congruence is to replace the Eisenstein series $(G_\kappa^{2n} \mid \tau_n)(\operatorname{diag}[z, w]; \mathfrak{f})$ with a form of level $\Gamma_n \times \Gamma_n$. To do this, we take the trace

$$\widetilde{G}_\kappa^{2n}(\operatorname{diag}[z, w]; \mathfrak{f}) = \sum_{\gamma_1, \gamma_2} (G_\kappa^{2n} \mid \tau_n)(\operatorname{diag}[z, w]; \mathfrak{f}) \mid (\gamma_1 \times \gamma_2)$$

where the sum is over $(\Gamma_n \times \Gamma_n)/(\tau_n^{-1} \Gamma_0^{(n)}(N) \tau_n \times \tau_n^{-1} \Gamma_0^{(n)}(N) \tau_n)$. We note again that this has Fourier coefficients in $\mathbb{Z}_\ell[\chi]$ by the q-expansion principle. Moreover, we know that $\widetilde{G}_\kappa^{2n}$ is a cusp form in each variable via [Brown 2011, Section 3.2].

Let $F_0 = I_n(f), F_1, \ldots, F_r$ be an orthogonal basis of eigenforms for $S_\kappa(\Gamma_n)$. Note that $F_0^c, \ldots, F_r^c$ is also an orthogonal basis of eigenforms for $S_\kappa(\Gamma_n)$. Applying [Shimura 1995, Equation (7.7)] we may write

$$\widetilde{G}_\kappa^{2n}(\operatorname{diag}[z, w]; \mathfrak{f}) = \sum_{\substack{0 \le i \le r \\ 0 \le j \le r}} c_{i,j} F_i(z) F_j^c(w)$$

for some $c_{i,j} \in \mathbb{C}$. Furthermore, from [Brown 2011, Lemma 5.1] we can rewrite

$$\widetilde{G}_\kappa^{2n}(\operatorname{diag}[z, w]; \mathfrak{f}) = c_0 I_n(f)(z) I_n(f)(w) + \sum_{0 < j \le r} c_j F_j(z) F_j^c(w), \tag{6}$$

where we write $c_j = c_{j,j}$ and we have used that since $f^c = f$, Corollary 6 gives $I_n(f)^c = I_n(f)$.

We now turn our attention to the constant c_0. Our goal is to show that we can write c_0 as a product of an element of $\mathcal{O}^\times$ and λ^{-m} for some $m > 0$.

Consider the inner product $\langle \widetilde{G}_{\kappa}^{2n}(\mathrm{diag}[z,w];\mathfrak{f}), I_n(f)(w)\rangle$. Note that

$$\langle \widetilde{G}_{\kappa}^{2n}(\mathrm{diag}[z,w];\mathfrak{f}), I_n(f)(w)\rangle = \langle (G_{\kappa}^{2n} \mid \tau_n)(\mathrm{diag}[z,w];\mathfrak{f}), I_n(f)(w)\rangle,$$

where we view the forms on the left-hand side as being level Γ_n and on the right-hand side as being level $\tau_n^{-1}\Gamma_0^{(n)}(N)\tau_n$. Taking the inner product of both sides of (6) with $I_n(f)(w)$, applying (5), and solving for c_0 we obtain

$$c_0 = \frac{\mathcal{A}_{k,n,N} L^N(n-\kappa+1, I_n(f), \chi; \mathrm{st})}{\pi^{n(n+1)/2}\langle I_n(f), I_n(f)\rangle}.$$

Ikeda [2006] made a conjecture relating $\langle I_n(f), I_n(f)\rangle$ to $\langle f, f\rangle$. We have the following theorem, which proves Ikeda's conjecture assuming n is even. We rephrase their result to suit our purposes.

Theorem 18 [Katsurada and Kawamura 2013, Theorem 2.3]. *Let κ be a positive even integer and let $\ell > n+1$ be a prime. Let $f \in S_{2\kappa-n}(\Gamma_1;\mathcal{O})$ be a newform with $\mathcal{O}$ a suitably large finite extension of $\mathbb{Z}_\ell$. Assume* $\mathrm{val}_{\mathfrak{l}}(\langle f,f\rangle/(\Omega_f^+\Omega_f^-)) = 0$. *Let D be a fundamental discriminant such that $(-1)^{n/2}D > 0$, $\chi_D(-1) = -1$, and assume $\ell \nmid D$. Then if $I_n(f)$ is the Ikeda lift of f as given above, we have*

$$\frac{\langle I_n(f), I_n(f)\rangle}{\langle f,f\rangle^{n/2}} = u_1 \cdot \frac{\Gamma(\kappa)\prod_{j=1}^{\frac{n}{2}-1}\Gamma(2j+2\kappa-n)|c(|D|)|^2\prod_{j=1}^{\frac{n}{2}}\zeta_{\mathrm{alg}}(2j)}{\Gamma\left(\kappa-\frac{n}{2}\right)} \times \frac{L_{\mathrm{alg}}(\kappa,f)\prod_{j=1}^{\frac{n}{2}-1}L_{\mathrm{alg}}(2j+1,\mathrm{ad}^0 f)}{L_{\mathrm{alg}}\left(\kappa-\frac{n}{2},f,\chi_D\right)},$$

where $\mathrm{val}_{\mathfrak{l}}(u_1) = 0$, *$c(|D|)$ is the $|D|$-th Fourier coefficient of $\theta_{\kappa,n}^{\mathrm{alg}}(f)$ from above and we have used the assumption on $\langle f,f\rangle/(\Omega_f^+\Omega_f^-)$ to normalize the adjoint L-function to our conventions.*

We now apply this result to remove the period $\langle I_n(f), I_n(f)\rangle$ in our expression for c_0 to obtain

$$c_0 = \frac{\mathcal{B}_{\kappa,n}}{|c(|D|)|^2}\cdot\frac{L^N(n-\kappa+1, I_n(f),\chi;\mathrm{st})L_{\mathrm{alg}}\left(\kappa-\frac{n}{2},f,\chi_D\right)}{\pi^{\frac{n(n+1)}{2}}\langle f,f\rangle^{\frac{n}{2}}\zeta_{\mathrm{alg}}(n)\prod_{i=1}^{\frac{n}{2}-1}\zeta_{\mathrm{alg}}(2i)L_{\mathrm{alg}}(2i+1,\mathrm{ad}^0 f)L_{\mathrm{alg}}(\kappa,f)},$$

where

$$\mathcal{B}_{\kappa,n} = u_2\cdot\frac{\Gamma\left(\kappa-\frac{n}{2}\right)\prod_{j=1}^{n-1}\Gamma\left(\frac{n-j}{2}\right)}{[\Gamma_n:\Gamma_0^{(n)}(N)]\Gamma(k)\prod_{j=1}^{n-1}\Gamma\left(\frac{2n+1-j}{2}\right)\prod_{j=1}^{\frac{n}{2}-1}\Gamma(2i+2k-n)},$$

where u_2 satisfies $\mathrm{val}_{\mathfrak{l}}(u_2) = 0$.

The following factorization is a direct consequence of Theorem 4:

$$L^N(n-k+1, I_n(f),\chi;\mathrm{st}) = L^N(n-k+1,\chi)\prod_{i=1}^{n}L^N(n+1-i,f,\chi). \tag{7}$$

Applying the assumption that $\mathrm{val}_{\mathfrak{l}}(\langle f, f\rangle/(\Omega_f^+\Omega_f^-)) = 0$, we can replace $\langle f, f\rangle^{n/2}$ by $u_3(\Omega_f^+\Omega_f^-)^{n/2}$ for u_3 an $\mathfrak{l}$-adic unit. Furthermore, note that if $\Omega_f^{\pm}$ is the period associated to $L(n+1-i, f, \chi)$ as in Theorem 1, then $\Omega_f^{\mp}$ is the period associated to $L(n+1-(i+1), f, \chi)$. Using this, we can rewrite our expression for c_0 as

$$c_0 = u_4 \cdot \mathcal{B}_{\kappa,n} \cdot \mathcal{C}_{D,n,\chi} \cdot \mathcal{L}_{f,\chi,D},$$

where u_4 is a $\mathfrak{l}$-adic unit, $\mathcal{B}_{\kappa,n}$ is defined as above,

$$\mathcal{C}_{D,n,\chi} = \frac{1}{|c_h(|D|)|^2 \prod_{i=1}^{\frac{n}{2}} \zeta_{\mathrm{alg}}(2i)},$$

and

$$\mathcal{L}_{f,\chi,\chi_D} = \frac{L^N(n-\kappa+1, \chi) L_{\mathrm{alg}}\big(\kappa - \frac{n}{2}, f, \chi_D\big) \prod_{j=1}^{n} L_{\mathrm{alg}}^N(n+1-j, f, \chi)}{L_{\mathrm{alg}}(\kappa, f) \prod_{j=1}^{\frac{n}{2}-1} L_{\mathrm{alg}}(2j+2, \mathrm{ad}^0 f)}.$$

Note that it is shown in [Brown 2007, Section 4.2] that $L^N(n-k+1, \chi) \in \mathbb{Z}_\ell[\chi]$. As $\mathcal{B}_{\kappa,n}$, $\mathcal{C}_{D,n,\chi}$, and $\mathcal{L}_{f,\chi,D}$ are algebraic, we may consider the $\mathfrak{l}$-divisibility of c_0. First, using that n is even and $\ell > n+1$ we have $\mathrm{val}_{\mathfrak{l}}(\mathcal{B}_{k,n}) \le 0$.

Next we turn our attention to $\mathcal{C}_{D,n,\chi}$. Our choice of $\theta_{\kappa,n}^{\mathrm{alg}}(f)$ given in Section 3 gives that $|c(|D|)| \in \mathcal{O}$, and so $\mathrm{val}_{\mathfrak{l}}(|c(|D|)|^2) \ge 0$. Consider $\zeta_{\mathrm{alg}}(2j)$ for some $1 \le j \le \frac{n}{2}$. It is an immediate consequence of the Von Staudt–Clausen Theorem (see for example [Ireland and Rosen 1990, p. 233]) that $\zeta_{\mathrm{alg}}(2j)$ is in $\mathcal{O}$, and hence $\mathrm{val}_{\mathfrak{l}}(\zeta_{\mathrm{alg}}(2j)) \ge 0$. Thus, we have $\mathrm{val}_{\mathfrak{l}}(\mathcal{C}_{D,n,\chi}) \le 0$.

By assumption we have $\mathrm{val}_{\mathfrak{l}}(\mathcal{L}_{f,\chi,\chi_D}) < 0$, so under our assumptions we have $\mathrm{val}_{\mathfrak{l}}(c_0) < 0$. We now show how this gives the desired congruence. Write $c_0 = \alpha\lambda^{-b'}$ for some $b' > 0$ and α an $\mathfrak{l}$-adic unit. Using this, we may rewrite (6) as

$$\widetilde{G}_\kappa^{2n}(\mathrm{diag}[z, w]; \mathfrak{f}) = \alpha\lambda^{-b'} I_n(f)(z) I_n(f)(w) + \sum_{0<j\le r} c_j F_j(z) F_j^c(w). \tag{8}$$

Note that by Proposition 7 there is a T_0 so that $\mathrm{val}_{\mathfrak{l}}(a_{I_n(f)}(T_0)) = 0$. We expand (8) in terms of z and equate the T_0-th Fourier coefficients to obtain

$$\sum_{T_2\in\Lambda_n} \Bigg(\sum_{T\in\Lambda_{2n}(T_0,T_2)} a(T, G_\kappa^{2n} \mid \tau_n) \Bigg) e(\mathrm{Tr}(T_2 w)) \\ = \alpha\lambda^{-b'} a_{I_n(f)}(T_0) I_n(f)(w) + \sum_{0<j\le r} c_j a_{F_j}(T_0) F_j^c(w).$$

Multiply the equation by $\lambda^{b'}$ and recall that $a(T, G_\kappa^{2n}|\tau_n) \in \mathcal{O}$ for all T to see that

$$I_n(f)(w) \equiv -\frac{\lambda^{b'}}{\alpha a_{I_n(f)}(T_0)} \sum_{0<j\le r} c_j a_{F_j}(T_0) F_j^c(w) \pmod{\mathfrak{l}^{b'}}.$$

Note that since $a_{I_n(f)}(T_0)$ is a $\mathfrak{l}$-adic unit, the form on the right-hand side of the congruence cannot be zero modulo $\mathfrak{l}^{b'}$, i.e., we have constructed a nontrivial congruence. Set

$$G(w) = -\frac{\lambda^{b'}}{\alpha a_{I_n(f)}(T_0)} \sum_{0<j\leq r} c_j a_{I_n(f)}(T_0) F_j(w).$$

We now return to the setting of Ikeda ideals. Let $X = \mathbb{C}I_n(f)$ and $Y = (\mathbb{C}I_n(f))^\perp$, where the notation follows that given in Section 4.2. We have constructed a congruence

$$I_n(f) \equiv G \pmod{\mathfrak{l}^{b'}}$$

for some $b' \geq 1$ and $G \in Y$. Note that it is clear from above that

$$b' \geq \mathrm{val}_{\mathfrak{l}}\biggl(L_{\mathrm{alg}}(\kappa, f) \prod_{j=1}^{\frac{n}{2}-1} L_{\mathrm{alg}}(2j+1, \mathrm{ad}^0 f)\biggr),$$

which is what we labeled b in the statement of Theorem 14. Thus, applying Proposition 11 concludes the proof of Theorem 14.

One thing to note here is that we do not obtain a lower bound of b' for the index in the Hecke algebra of the Ikeda ideal with respect to $X = S_\kappa^{\mathrm{Ik}}(\Gamma_n)$ and $Y = S_\kappa^{\text{N-Ik}}(\Gamma_n)$. The reason for this is that while we know $I_n(f)$ cannot be eigenvalue-congruent to any other Ikeda lifts, that does not imply that $G \in S_\kappa^{\text{N-Ik}}(\Gamma_n)$. One can use the fact that $I_n(f)$ is not congruent to any other Ikeda lifts along with (1) to conclude there is an idempotent t in the Hecke algebra $\mathcal{H}_{\mathcal{O}}^{(n)}$ that satisfies

$$tF = \begin{cases} 0 & \text{if } F \not\equiv_{\mathrm{ev}} I_n(f) \pmod{\mathfrak{l}}, \\ F & \text{if } F \equiv_{\mathrm{ev}} I_n(f) \pmod{\mathfrak{l}}. \end{cases}$$

If one acts on G by this Hecke operator one obtains a form tG in $S_\kappa^{\text{N-Ik}}(\Gamma_n)$ with $tG \equiv_{\mathrm{ev}} I_n(f) \pmod{\mathfrak{l}}$. Thus, one only obtains a lower bound of 1 for the Ikeda ideal with respect to $X = S_\kappa^{\mathrm{Ik}}(\Gamma_n)$ and $Y = S_\kappa^{\text{N-Ik}}(\Gamma_n)$. While one would like to have a stronger bound for this Ikeda ideal, Corollary 15 shows that it is not necessary for our results.

References

[Agarwal and Brown 2013] M. Agarwal and J. Brown, "Computational evidence for the Bloch–Kato conjecture for elliptic modular forms of square-free level", preprint, 2013, available at http://www.ces.clemson.edu/~jimlb/ResearchPapers/BlochKatoCompEvid.pdf.

[Agarwal and Brown 2014] M. Agarwal and J. Brown, "On the Bloch–Kato conjecture for elliptic modular forms of square-free level", *Math. Z.* **276**:3-4 (2014), 889–924. MR 3175164

[Asgari and Schmidt 2001] M. Asgari and R. Schmidt, "Siegel modular forms and representations", *Manuscripta Math.* **104**:2 (2001), 173–200. MR 2002a:11044 Zbl 0987.11037

[Berger et al. ≥ 2015] T. Berger, K. Klosin, and K. Kramer, "On higher congruences between automorphic forms", *Math. Res. Lett*. To appear. arXiv 1302.2381

[Böcherer 1985] S. Böcherer, "Über die Fourier–Jacobi-Entwicklung Siegelscher Eisensteinreihen, II", *Math. Z.* **189**:1 (1985), 81–110. MR 86f:11037 Zbl 0558.10022

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR 1484478 Zbl 0898.68039

[Brown 2007] J. Brown, "Saito–Kurokawa lifts and applications to the Bloch–Kato conjecture", *Compos. Math.* **143**:2 (2007), 290–322. MR 2008i:11064 Zbl 1172.11015

[Brown 2011] J. Brown, "On the cuspidality of pullbacks of Siegel Eisenstein series and applications to the Bloch–Kato conjecture", *Int. Math. Res. Not.* **2011**:7 (2011), 1706–1756. MR 2012j:11105 Zbl 05898719

[Bruinier and Ono 2003] J. H. Bruinier and K. Ono, "Coefficients of half-integral weight modular forms", *J. Number Theory* **99**:1 (2003), 164–179. MR 2004b:11056 Zbl 1090.11028

[Chai and Faltings 1990] C.-L. Chai and G. Faltings, *Degeneration of abelian varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **22**, Springer, Berlin, 1990. MR 92d:14036 Zbl 0744.14031

[Garrett 1984] P. B. Garrett, "Pullbacks of Eisenstein series; applications", pp. 114–137 in *Automorphic forms of several variables* (Katata, 1983), edited by I. Satake and Y. Morita, Progr. Math. **46**, Birkhäuser, Boston, 1984. MR 86f:11039 Zbl 0544.10023

[Garrett 1992] P. B. Garrett, "On the arithmetic of Siegel–Hilbert cuspforms: Petersson inner products and Fourier coefficients", *Invent. Math.* **107**:3 (1992), 453–481. MR 93e:11060 Zbl 0769.11024

[van der Geer 2008] G. van der Geer, "Siegel modular forms and their applications", pp. 181–245 in *The 1-2-3 of modular forms* (Nordfjordeid, 2004), edited by J. H. Bruinier et al., Springer, Berlin, 2008. MR 2010c:11059 Zbl 1259.11051

[Hida 1981] H. Hida, "Congruences of cusp forms and special values of their zeta functions", *Invent. Math.* **63**:2 (1981), 225–261. MR 82g:10044 Zbl 0459.10018

[Hida 1987] H. Hida, "Theory of p-adic Hecke algebras and Galois representation", *Sūgaku* **39**:2 (1987), 124–139. In Japanese; translated in *Sūgaku Expositions* **2**:1 (1989), 75–102. MR 89a:11061 Zbl 0641.10025

[Ikeda 2001] T. Ikeda, "On the lifting of elliptic cusp forms to Siegel cusp forms of degree $2n$", *Ann. of Math.* (2) **154**:3 (2001), 641–681. MR 2002m:11037 Zbl 0998.11023

[Ikeda 2006] T. Ikeda, "Pullback of the lifting of elliptic cusp forms and Miyawaki's conjecture", *Duke Math. J.* **131**:3 (2006), 469–497. MR 2007e:11052 Zbl 1112.11022

[Ikeda 2013] T. Ikeda, "On the lifting of automorphic representations of $\mathrm{PGL}_2(\mathbb{A})$ to $\mathrm{Sp}_{2n}(\mathbb{A})$ of $\widehat{\mathrm{Sp}_{2n+1}}(\mathbb{A})$ over a totally real field", preprint, Kyoto University, 2013, available at http://www.math.kyoto-u.ac.jp/~ikeda/adelic1.pdf.

[Ireland and Rosen 1990] K. Ireland and M. Rosen, *A classical introduction to modern number theory*, 2nd ed., Graduate Texts in Mathematics **84**, Springer, New York, 1990. MR 92e:11001 Zbl 0712.11001

[Katsurada 2011] H. Katsurada, "Congruence between Duke–Imamoḡlu–Ikeda lifts and non-Duke–Imamoḡlu–Ikeda lifts", preprint, 2011. arXiv 1101.3377

[Katsurada and Kawamura 2013] H. Katsurada and H. Kawamura, "Ikeda's conjecture on the period of Duke–Imamoḡlu–Ikeda lift", preprint, 2013. arXiv 1008.4195

[Klosin 2009] K. Klosin, "Congruences among modular forms on $\mathrm{U}(2, 2)$ and the Bloch–Kato conjecture", *Ann. Inst. Fourier* (*Grenoble*) **59**:1 (2009), 81–166. MR 2010g:11077 Zbl 1214.11055

[Kohnen 2002] W. Kohnen, "Lifting modular forms of half-integral weight to Siegel modular forms of even genus", *Math. Ann.* **322**:4 (2002), 787–809. MR 2003d:11067 Zbl 1004.11020

[Kurokawa 1981] N. Kurokawa, "On Siegel eigenforms", *Proc. Japan Acad. Ser. A Math. Sci.* **57**:1 (1981), 47–50. MR 83d:10033 Zbl 0482.10025

[Langlands 1976] R. P. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. **544**, Springer, Berlin, 1976. MR 58 #28319 Zbl 0332.10018

[Mazur 2011] B. Mazur, "How can we construct abelian Galois extensions of basic number fields?", *Bull. Amer. Math. Soc.* (*N.S.*) **48**:2 (2011), 155–209. MR 2012b:11173 Zbl 1228.11163

[Mizumoto 1991] S.-i. Mizumoto, "Integrality of Eisenstein liftings", *Proc. Japan Acad. Ser. A Math. Sci.* **67**:1 (1991), 11–13. MR 92f:11065 Zbl 0745.11028

[Ribet 1976] K. A. Ribet, "A modular construction of unramified p-extensions of $Q(\mu_p)$", *Invent. Math.* **34**:3 (1976), 151–162. MR 54 #7424 Zbl 0338.12003

[Ribet 1983] K. A. Ribet, "Mod p Hecke operators and congruences between modular forms", *Invent. Math.* **71**:1 (1983), 193–205. MR 84j:10040 Zbl 0508.10018

[Shimura 1977] G. Shimura, "On the periods of modular forms", *Math. Ann.* **229**:3 (1977), 211–221. MR 57 #3080 Zbl 0363.10019

[Shimura 1983] G. Shimura, "On Eisenstein series", *Duke Math. J.* **50**:2 (1983), 417–476. MR 84k:10019 Zbl 0519.10019

[Shimura 1995] G. Shimura, "Eisenstein series and zeta functions on symplectic groups", *Invent. Math.* **119**:3 (1995), 539–584. MR 96e:11065 Zbl 0845.11020

[Shimura 1997] G. Shimura, *Euler products and Eisenstein series*, CBMS Regional Conference Series in Mathematics **93**, Amer. Math. Soc., Providence, RI, 1997. MR 98h:11057 Zbl 0906.11020

[Shintani 1975] T. Shintani, "On construction of holomorphic cusp forms of half integral weight", *Nagoya Math. J.* **58** (1975), 83–126. MR 52 #10603 Zbl 0316.10016

[Skinner and Urban 2006] C. Skinner and E. Urban, "Sur les déformations p-adiques de certaines représentations automorphes", *J. Inst. Math. Jussieu* **5**:4 (2006), 629–698. MR 2008a:11072 Zbl 1169.11314

[Skinner and Urban 2014] C. Skinner and E. Urban, "The Iwasawa main conjectures for GL_2", *Invent. Math.* **195**:1 (2014), 1–277. MR 3148103

[Stein et al. 2013] W. A. Stein et al., *Sage mathematics software*, Version 5.10, Sage Development Team, 2013, available at http://www.sagemath.org.

[Stevens 1994] G. Stevens, "Λ-adic modular forms of half-integral weight and a Λ-adic Shintani lifting", pp. 129–151 in *Arithmetic geometry: conference on arithmetic geometry with an emphasis on Iwasawa theory* (Tempe, AZ, 1993), edited by N. Childress and J. W. Jones, Contemp. Math. **174**, Amer. Math. Soc., Providence, RI, 1994. MR 95h:11051 Zbl 0869.11042

[Swinnerton-Dyer 1973] H. P. F. Swinnerton-Dyer, "On ℓ-adic representations and congruences for coefficients of modular forms", pp. 1–55 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, Berlin, 1973. MR 53 #10717a Zbl 0267.10032

[Vatsal 1999] V. Vatsal, "Canonical periods and congruence formulae", *Duke Math. J.* **98**:2 (1999), 397–419. MR 2000g:11032 Zbl 0979.11027

[Washington 1978] L. C. Washington, "The non-p-part of the class number in a cyclotomic $\mathbb{Z}_p$-extension", *Invent. Math.* **49**:1 (1978), 87–97. MR 80c:12005 Zbl 0403.12007

[Wiles 1990] A. Wiles, "The Iwasawa conjecture for totally real fields", *Ann. of Math.* (2) **131**:3 (1990), 493–540. MR 91i:11163 Zbl 0719.11071

[Zagier 1977] D. B. Zagier, "Modular forms whose Fourier coefficients involve zeta-functions of quadratic fields", pp. 105–169 in *Modular functions of one variable, VI* (Bonn, 1976), edited by

J.-P. Serre and D. B. Zagier, Lecture Notes in Math. **627**, Springer, Berlin, 1977. MR 58 #5525 Zbl 0372.10017

Received October 16, 2013. Revised January 20, 2014.

JIM BROWN
DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
O-110 MARTIN HALL
BOX 340975
CLEMSON, SC 29634-0975
UNITED STATES
jimlb@clemson.edu

RODNEY KEATON
DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
CLEMSON, SC 29634
UNITED STATES
rkeaton@clemson.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.53

CONSTANT MEAN CURVATURE, FLUX CONSERVATION, AND SYMMETRY

NICK EDELEN AND BRUCE SOLOMON

As first noted by Korevaar, Kusner, and Solomon, constant mean curvature implies a homological conservation law for hypersurfaces in ambient spaces with Killing fields. We generalize that law by relaxing the topological restrictions assumed by Korevaar et al., and by allowing a weighted mean curvature functional. We also prove a partial converse, which roughly says that when flux is conserved along a Killing field, a hypersurface splits into two regions: one with constant (weighted) mean curvature, and one preserved by the Killing field. We demonstrate our theory by using it to derive a first integral for helicoidal surfaces of constant mean curvature in $\mathbb{R}^3$, i.e., "twizzlers".

1. Introduction

Constant mean curvature ("CMC") imposes a homological flux conservation law on hypersurfaces in ambient spaces with nontrivial Killing fields. This was first observed and exploited by Korevaar, Kusner, and Solomon [1989] in their paper on the structure of embedded CMC surfaces in $\mathbb{R}^3$ (see [Kusner 1991] for an alternative exposition). In Theorem 3.5, we generalize that law by relaxing the topological restrictions assumed by Korevaar et al. [1989], and by allowing a weighted version of the mean curvature functional. We further extend the theory via Theorem 4.1, which gives a partial converse to the conservation law. Roughly, it states that when the appropriate flux is conserved along Killing fields, the hypersurface splits into two regions (though either may be empty): a region with constant (weighted) mean curvature, and a region preserved by the Killing fields.

We apply our results in Case study 4.5 by using them to quickly derive the seemingly *ad hoc* first integral that Perdomo [2012], do Carmo and Dajczer [1982], and others have used to analyze the moduli space of CMC surfaces with helicoidal symmetry, also known as *twizzlers*[1]. In general, constancy of weighted mean

MSC2010: 53A10.
Keywords: constant mean curvature, conservation law.

[1]Twizzlers have also been studied by Wunderlich [1952] and, more recently, Halldorsson [2013].

curvature is characterized by a nonlinear second-order PDE, and its Noetherian reduction to a first-order condition makes it easier to analyze.

When a CMC hypersurface Σ in a manifold N is preserved by the action of a continuous isometry group $\mathcal{G}$, one can project it into the orbit space $N/\mathcal{G}$. The projected hypersurface $\Sigma/\mathcal{G}$ will then be stationary for the *weighted* functional introduced in Section 3.2. We analyze the weighted functional and the resulting weighted mean-curvature invariant with an eye toward this fact. We suspect that virtually all we do here could be developed in a more general, stratified context encompassing both Riemannian manifolds and their quotients under smooth group actions.

We stick with smooth ambient manifolds here, but the orbit space viewpoint can be helpful, and Case study 4.5 could easily have been carried out in that setting. The approach we demonstrate there can also be adapted to spherical and hyperbolic space forms. The first author's report [Edelen 2011] sketches out one way to do that, but we describe the orbit-space approach to those examples in our final Remark 4.6.2.

2. Preliminaries

Let N denote an n-dimensional oriented Riemannian manifold, and consider a smooth, connected, oriented, properly immersed hypersurface $f : \Sigma^{n-1} \to N$. We will feel free to write Σ when we mean $f(\Sigma)$ or even $f : \Sigma \to N$, leaving context to clarify our intentions.

Let ν denote the unit normal that completes the orientation of Σ to that of N. The *mean curvature function* $h : \Sigma \to \mathbb{R}$ is the trace of the shape operator $\nabla\nu$. Notationally,

$$h = \operatorname{div}_\Sigma(\nu). \tag{2-1}$$

Here $\operatorname{div}_\Sigma(Y)$ denotes the *intrinsic divergence* of a vector field Y along Σ, that is, the trace of the endomorphism $T\Sigma \to T\Sigma$ gotten at each $p \in \Sigma$ by projecting the ambient covariant derivative ∇Y onto $T_p\Sigma$. One may compute $\operatorname{div}_\Sigma$ locally using any orthonormal basis $\{e_i\}$ for $T_p\Sigma$ via

$$\operatorname{div}_\Sigma(\mathrm{Y}) := \sum_{i=1}^{n-1} \nabla_{e_i} Y \cdot e_i .$$

2.1. *Chains and k-area.*

The homology of the sequence

$$N \to (N, \Sigma) \to \Sigma$$

will play a role below in a way that makes it problematic to work solely with smooth submanifolds. We therefore work with a class of piecewise smooth objects:

Definition 2.1.1. A *smooth r-chain* (or simply *chain*) in a smooth manifold M is a finite union of smoothly immersed oriented r-dimensional simplices. We regard a chain X as a formal homological sum

$$X = \sum_{i=1}^{m} m_i f_i. \tag{2-2}$$

Here each $f_i : \Delta \to M$ immerses the standard closed oriented r-simplex Δ (along with its boundary) smoothly into M. The m_i are (for us) always integers.

We denote the *support* of a chain X by $\|X\|$.

We write $\mathcal{S}_r(M)$ for the *group of smooth r-chains in M*, and ∂X for the homological *boundary* of a chain X, while $\mathfrak{Z}_i(M)$ and $\mathfrak{B}_i(M)$ denote the spaces of i-dimensional *cycles* and *boundaries* (kernel and image of ∂) in M, respectively. Likewise, $\mathfrak{Z}_i(M, A)$ and $\mathfrak{B}_i(M, A)$ indicate spaces of cycles and boundaries modulo a subset $A \subset M$.

Integration of an r-form ϕ on M over such a chain is trivial:

$$\int_X \phi := \sum_{i=1}^{m} m_i \int_\Delta f_i^* \phi,$$

where f_i^* denotes the usual pullback.

Given a Riemannian metric on M, one can also integrate *functions* over chains, and most importantly for our purposes, compute weighted volumes.

Definition 2.1.2. Let $\mu : M \to \mathbb{R}$ be any continuous function. Define the *μ-weighted r-volume* $|X|_\mu$ of the r-chain X in (2-2) as

$$|X|_\mu := \sup\left\{ \int_X e^\mu \phi : \phi \text{ is an } r\text{-form on } M \text{ with } \|\phi\|_\infty \le 1 \right\}.$$

For a single immersed simplex, the usual Riemannian volume integral gives a simpler definition. To allow coincident, oppositely oriented simplices to cancel, however, we need the definition above.[2]

Finally, note that because Stokes' Theorem holds for immersed r-simplices, it holds for r-chains as well.

2.2. *Symmetry.* Our work here is vacuous unless the ambient space N has nontrivial Killing fields.

Write $\mathcal{I}$ and $L(\mathcal{I})$ respectively for the isometry group of N and its Lie algebra. Identify $L(\mathcal{I})$ with the linear space of Killing fields on N in the usual way, associating each $Y \in L(\mathcal{I})$ with the Killing field (also called Y) we get by differentiating

[2] Definition 2.1.2 amounts to a weighted version of the *mass* of X as a *current*, in the sense of geometric measure theory [Federer 1969, p. 358].

the flow that sends $p \in N$ along the path $t \mapsto \exp(tY)p$. We write Y_p for the value of Y at p.

One often studies CMC hypersurfaces (like surfaces of revolution and twizzlers in $\mathbb{R}^3$) in relation to the action of a closed, connected subgroup $\mathscr{L} \subset \mathscr{I}$. Though it complicates our exposition to some extent, the presence of such a subgroup $\mathscr{L}$ — like the density function e^μ — lets us broaden our theory. Even when $\mu \equiv 0$ and $\mathscr{L}$ is the full isometry group of N, however, our results go beyond those of [Korevaar et al. 1989].

In Theorem 4.1 (a converse to our conservation law) we must consider the possibility that all Killing fields associated with $\mathscr{L}$ lie tangent to an open subset S of our hypersurface $\Sigma \subset N$. The following lemma (and its corollary) then lets us deduce $\mathscr{L}$-invariance of S.

Lemma 2.3. *Suppose $S \subset N$ is a hypersurface, and that for some $Y \in L(\mathscr{L})$, we have $Y_p \in T_pS$ for every $p \in S$. Then for each $p \in S$, there exist a compact neighborhood $\mathbb{O}_p \subset S$ and an $\varepsilon > 0$ such that $e^{tY}q \in S$ whenever $|t| < \varepsilon$ and $q \in \mathbb{O}_p$.*

Proof. Since S is a submanifold, some open set $W \subset N$ contains S, but no point of $\bar{S} \setminus S$ ($\bar{S}$ = closure of S). Let $\Theta : S \times \mathbb{R} \to N$ denote the flow of Y, so that $\Theta(q,t) := \exp(tY)q$. Then $\Theta^{-1}(W)$ is an open neighborhood of $S \times \{0\}$.

Now $\Theta(q,t)$ parametrizes the integral curve of Y with initial velocity Y_q. But $Y_p \in T_pS$ for all $p \in S$, and first-order ODE's have unique solutions, so this curve must stay in S for all $(q,t) \in W$. It follows that $\Theta^{-1}(W) = \Theta^{-1}(S)$.

For any compact neighborhood $\mathbb{O}_p$ of $p \in S$, there now exists an $\varepsilon > 0$ such that

$$\mathbb{O}_p \times (-\varepsilon, \varepsilon) \subset \Theta^{-1}(S),$$

and the lemma consequently holds with this choice of $\mathbb{O}_p$ and ε. □

2.4. *Flux.* Korevaar et al. [1989] showed that when a hypersurface $\Sigma \subset N^n$ has constant mean curvature $h \equiv H$ and the homology groups $H_{n-1}(N)$ and $H_{n-2}(N)$ are both trivial (over $\mathbb{Z}$ — all homology groups in this paper have integer coefficients), there exists a *flux homomorphism*

$$\phi : H_{n-2}(\Sigma) \otimes L(\mathscr{I}) \to \mathbb{R}$$

defined by assigning, to any Killing field Y and any class $\boldsymbol{k} \in H_{n-2}(\Sigma)$, the *flux* $\phi(\boldsymbol{k}, Y)$ *of Y across $\boldsymbol{k}$*, where

$$\phi(\boldsymbol{k}, Y) := \int_\Gamma \eta \cdot Y + H \int_K \nu \cdot Y. \tag{2-3}$$

Here,

- Γ can be an $(n-2)$-cycle representing $\boldsymbol{k}$;

- $K \subset N$ can be any $(n-1)$-chain bounded by Γ;
- η is the orienting unit conormal to Γ in Σ; and
- ν is the orienting unit normal to K in N.

To ensure that $\phi(\boldsymbol{k}, Y)$ is well-defined by (2-3), [Korevaar et al. 1989] makes two topological assumptions: namely, *that* $H_{n-1}(N)$ *and* $H_{n-2}(N)$ *both vanish*. The vanishing of $H_{n-2}(N)$ ensures Γ will bound *some* chain K, while that of $H_{n-1}(N)$ means any competing chain K' with $\partial K' = \partial K$ can be written $K' = K + \partial U$ for some n-chain U. Since Killing fields are divergence-free, the divergence theorem then makes the second integral in (2-3) independent of the choice of K.

Here, we extend this flux theory in [Korevaar et al. 1989] in several ways.

First, in Section 3.2, we broaden the mean curvature functional by allowing μ-weighted area and volume as in Definition 2.1.2. This is a minor tweak of the standard theory, but it does *not* correspond to a mere conformal change of metric, since n- and $(n-1)$-dimensional volumes scale differently under conformal change. We do this with a geometric application in mind: the μ-weighted theory relates the geometry of $\mathscr{L}$-invariant CMC hypersurfaces in N to that of hypersurfaces in the orbit space $N/\mathscr{L}$ (see Remark 3.3.2 below).

Second, and more importantly, we eliminate the homological triviality assumptions mentioned above. Though we follow the same variational strategy as in [Korevaar et al. 1989], we show the flux invariant lives more naturally in a certain *relative* homology group. Instead of focusing the invariant on $(n-2)$-cycles in the surface Σ, we realize the flux as an invariant on certain $(n-1)$-dimensional relative cycles we shall call *caps*.

The homological restriction can be naively avoided by defining the flux on $H_{n-1}(N, \Sigma)$. When $H_{n-2}(\Sigma) \neq 0$, however, one gets a more sensitive invariant by designating a set of “reference cycles”. We call this set a *spine*. It not only gives better invariants, but it tends to make flux calculations more tractable.

The new viewpoint reproduces the invariant in [Korevaar et al. 1989] when $H_{n-1}(N) = H_{n-2}(N) = 0$. In that case, the reference cycle is trivial, and the long exact sequence for the pair (N, Σ), namely

$$0 = H_{n-1}(N) \to H_{n-1}(N, \Sigma) \xrightarrow{\partial} H_{n-2}(\Sigma) \to H_{n-2}(N) = 0,$$

shows that $H_{n-1}(N, \Sigma) \cong H_{n-2}(\Sigma)$.

We derive our generalized conservation law in Section 3, and then, in Section 4, develop a partial converse. Before proceeding to these extensions, however, we present a motivating example that we can review later as an illustration of our theory.

Example 2.4.1. *Twizzlers* are “helicoidal” CMC surfaces invariant under a 1-parameter group of screw motions in $\mathbb{R}^3$. Any such surface can be gotten by applying a screw motion to a curve γ in a plane perpendicular to the screw-axis. The resulting helicoidal surface will then have mean curvature $h \equiv H$ if and only if γ satisfies an easily derived second-order ODE. However, as others have noted [do Carmo and Dajczer 1982; Wunderlich 1952; Perdomo 2012; Halldorsson 2013], the second-order ODE has a useful first integral. We show how to derive it from flux conservation below.

The conservation law formulated in [Korevaar et al. 1989], however, yields nothing for twizzlers, since the typical CMC twizzler is generated by a nonperiodic curve γ in the transverse plane, and thus lacks homology. To remedy that, one can mod out the translational period of the helicoidal motion, realizing the twizzler as an immersion of a *cylinder* in $N := \mathbb{R}^2 \times \boldsymbol{S}^1$. Cylinders *do* have nontrivial loops, but those loops don't bound in N, and hence can't be capped off as required by [Korevaar et al. 1989].

Our approach evades that obstruction; see Example 3.1.1 and Case study 4.5.

3. Conservation

Like Korevaar et al. [1989], we derive flux conservation using a constrained first-variation formula. We make two notable modifications, however.

First, we weight both the areas of hypersurfaces and the volumes of domains by an $\mathcal{L}$-invariant *density function*

$$e^{\mu} : N \to (0, \infty). \tag{3-1}$$

Here μ can be any smooth function fixed by $\mathcal{L}$. The formula in [Korevaar et al. 1989] effectively takes $\mu \equiv 0$, as will become clear in Section 3.2 below.

Secondly, we encode the homology of our immersion $f : \Sigma \to N$ into a set of reference cycles B. Let f_* denote the induced homomorphism

$$f_* : H_{n-2}(\Sigma) \to H_{n-2}(N).$$

Definition 3.0.1 (Spine). We call a subgroup $B \subset \mathfrak{X}_{n-2}(N)$ a *spine* for the pair (N, Σ) if:

(a) $B \cap \mathfrak{X}_{n-2}(\Sigma) = 0$;

(b) B generates $f_* H_{n-2}(\Sigma)$;

(c) the composition $B \to \mathfrak{X}_{n-2}(N) \to H_{n-2}(N)$ is injective.

We won't always draw an explicit distinction between the subgroup B and a set of generating cycles for B.

A nontrivial spine lets us assign fluxes to classes in $H_{n-2}(\Sigma)$ that don't bound in N.

Note that a spine for (N, Σ) always exists. Indeed, any independent set of cycles that generate $f_* H_{n-2}(\Sigma)$ in $H_{n-2}(N)$ will satisfy conditions (b) and (c) of Definition 3.0.1, and one can always perturb slightly, if needed, to realize (a). That condition is really an artifact of the language we use to define the flux invariants; we want the sum $\mathcal{Z}_{n-2}(\Sigma) + B$ to be direct. The assumption could be omitted in favor of more precision in distinguishing "caps with nontrivial spines" and "caps without spines".

Definition 3.0.2 (Cap). A *cap* K is any chain in $\mathcal{S}_{n-1}(N)$ such that

$$\partial K \in \mathcal{Z}_{n-2}(\Sigma) \oplus B.$$

As the kernel of the composition

$$\mathcal{S}_{n-1}(N) \xrightarrow{\partial} \mathcal{S}_{n-2}(N) \to \mathcal{S}_{n-2}(N)/(\mathcal{S}_{n-2}(\Sigma) \oplus B),$$

the set of all caps forms a group, which we denote by $\mathcal{Z}(N, \Sigma, B)$.

A *reduced cap* is a class belonging to the quotient

$$\mathcal{K}(N, \Sigma, B) = \mathcal{Z}(N, \Sigma, B)/\mathcal{B}_{n-1}(N, \Sigma).$$

We call two caps K, K' *homologous*, written $K \sim K'$, if they represent the same reduced cap in $\mathcal{K}(N, \Sigma, B)$.

In spirit, a reduced cap is a class in $H_{n-1}(N, \Sigma \cup \|B\|)$, where $\|B\|$ denotes the support of B. Indeed, when $\|B\|$ is disjoint from Σ, we have $\mathcal{K}(N, \Sigma, B) = H_{n-1}(N, \Sigma \cup \|B\|)$. When $\|B\|$ does meet Σ, however, ambiguity can arise as to which part of ∂K to take as β in Observation 3.1 below. The need to remove that ambiguity motivated Definition 3.0.2. The direct sum decomposition of $\mathcal{Z}(N, \Sigma, B)$ there immediately yields the fact we need:

Observation 3.1. *For any cap K, there exists a unique $\beta \in B$ with $\|\partial K - \beta\| \subset \Sigma$.*

To make the notions of *spine* and *cap* more concrete, we illustrate using twizzlers:

Example 3.1.1. As explained in Example 2.4.1, we may regard a twizzler as a cylinder $\Sigma = \mathbb{R} \times \boldsymbol{S}^1$ immersed in $N = \mathbb{C} \times \boldsymbol{S}^1$ and preserved by the helical $\boldsymbol{S}^1$-action

$$[e^{i\theta}](z, e^{it}) = (e^{i\theta} z, e^{i(t+\theta)}). \tag{3-2}$$

The length of the $\boldsymbol{S}^1$-factor is geometrically significant, but we can take it to be the usual 2π for purposes of this example.

We call the orbits of the screw-action *helices*. By construction, both N and the twizzler $f(\Sigma)$ are foliated by such helices, any one of which generates $f_* H_1(\Sigma) = H_1(N)$. It follows that any helix, viewed as a 1-cycle in N, qualifies as a spine for (N, Σ). We take the shortest one, namely $\mathbf{0} \times \boldsymbol{S}^1 \subset N$, as our spine B.

Suppose a twizzler is generated by a particular curve $\gamma : \mathbb{R} \to \mathbb{C}$, so that we can immerse it in N via

$$f(t, e^{i\theta}) = (e^{i\theta}\gamma(t), e^{i\theta}).$$

For each fixed $t \in \mathbb{R}$, the helix $\Gamma_t := f(t, \boldsymbol{S}^1)$ forms a nontrivial cycle in $H_1(\Sigma)$. Any oriented surface that realizes the homology between Γ_t and the compatibly oriented cycle $\beta \in B$ is then a *cap* for Γ_t.

For instance, the line segment (or any arc) joining $\mathbf{0}$ to $\gamma(t)$ in $\mathbb{C}$ will, under the $\boldsymbol{S}^1$-action (3-2), sweep out a cap, and all arcs give rise to the same reduced cap in this way. Such caps are also preserved by the $\boldsymbol{S}^1$-action, a useful property that many other caps lack.

3.2. ***First variation.*** To prepare for our first variation formula, fix a spine B for (N, Σ), and suppose we have homologous caps K, K' in $\mathscr{Z}(N, \Sigma, B)$. There then exists an n-chain U satisfying

$$\partial U = S + K - K' \tag{3-3}$$

for some $S \in \mathscr{S}_{n-1}(\Sigma)$. Applying the boundary operator to (3-3), we then get

$$\partial K - \partial K' = -\partial S. \tag{3-4}$$

In particular, $\partial K - \partial K'$ is a cycle in Σ, and by the definition of a *cap*, there now exist unique β, $\beta' \in B$ such that

$$\partial K - \beta,\ \partial K' - \beta' \in \mathscr{Z}_{n-2}(\Sigma)$$

and (3-4) forces $\beta = \beta'$. This proves:

Proposition 3.3. *If two caps K, $K' \in \mathscr{Z}(N, \Sigma, B)$ are homologous, there exists a unique $\beta \in B$ such that both $\partial K - \beta$ and $\partial K' - \beta$ are supported in Σ.*

Definition 3.3.1. The proposition above lets us define the *spine of a reduced cap* $\boldsymbol{k} \in \mathscr{K}(N, \Sigma, B)$ as the unique $\beta \in B$ with $\partial K - \beta \in \mathscr{Z}_{n-2}(\Sigma)$ for any representative K.

In the situation just described, and in the presence of a density function e^{μ}, we now consider the n- and $(n-1)$-dimensional μ-weighted volumes $|U|_\mu$ and $|S|_\mu$ of the chains U and S respectively (Definition 2.1.2) as we deform along the flow of a smooth vector field Y. Fix a scalar H, and consider the initial derivative of $|S|_\mu - H\,|U|_\mu$ with respect to this flow, written

$$\delta_Y(|S|_\mu - H\,|U|_\mu). \tag{3-5}$$

Calling this the *(μ-weighted) volume-constrained first variation* of S, we obtain our conservation law for hypersurfaces with constant *μ-mean curvature* $h_\mu \equiv H$, as defined in (3-8) below, by evaluating (3-5) on Killing vector fields of N. To simplify the task, we analyze $\delta_Y\,|U|_\mu$ and $\delta_Y\,|S|_\mu$ separately before combining results.

A familiar derivation shows $\delta_Y |U|_\mu$ to equal the integral of $\mathrm{div_N(Y)}$ over U when $\mu \equiv 0$. A routine modification of that calculation shows that for general μ,

$$\delta_Y |U|_\mu = \int_U \mathrm{div_N(e^\mu Y)} = \int_{\partial U} e^\mu Y \cdot \nu,$$

where ν denotes the orienting unit normal along ∂U. By (3-3), we can rewrite this as

$$\delta_Y |U|_\mu = \int_S e^\mu Y \cdot \nu + \int_{K-K'} e^\mu Y \cdot \nu. \tag{3-6}$$

A similar modification of the $\mu \equiv 0$ case, as analyzed in [Simon 1983, pp. 46–51], computes the μ-weighted first variation of $|S|_\mu$ along Y:

$$\delta_Y |S|_\mu = \int_S e^\mu d\mu(\nu)\nu \cdot Y + \mathrm{div}_\Sigma(\mathrm{e}^\mu \mathrm{Y}^\top) + \mathrm{div}_\Sigma(\mathrm{e}^\mu \mathrm{Y}^\perp). \tag{3-7}$$

Here $Y^\top$ and $Y^\perp$ signify the tangential and normal components, respectively, of Y along S.

Recall that for vector fields *tangent* to Σ, the divergence theorem applies in its usual form: given an $(n-1)$-chain S in Σ with oriented unit conormal η along its boundary, we have

$$\int_S \mathrm{div}_\Sigma(\mathrm{X}) \; = \int_{\partial S} X \cdot \eta \quad (X \text{ tangent to } \Sigma).$$

For vector fields *normal* to Σ, on the other hand, the divergence operator invokes the mean curvature of Σ, due to (2-1). When $Z = (Z \cdot \nu)\nu$ is purely normal, then, the Leibniz rule yields

$$\int_S \mathrm{div}_\Sigma(\mathrm{Z}) \; = \int_S (Z \cdot \nu) h \quad (Z \text{ normal to } \Sigma).$$

Accordingly, we define the *μ-mean curvature* h_μ along Σ as

$$h_\mu := h + d\mu(\nu). \tag{3-8}$$

Using this notation, the facts above reduce (3-7) to

$$\delta_Y |S|_\mu = \int_{\partial S} e^\mu Y \cdot \eta + \int_S e^\mu h_\mu Y \cdot \nu. \tag{3-9}$$

Finally, using (3-6), (3-9), and (3-4), we can put our volume-constrained first-variation formula (3-5) into the form we need:

(3-10)

$$\delta_Y(|S|_\mu - H\,|U|_\mu) = -\int_{\partial K - \partial K'} e^\mu \eta \cdot Y \; - H \int_{K-K'} e^\mu \nu \cdot Y + \int_S e^\mu (h_\mu - H)\nu \cdot Y.$$

Remark 3.3.2. The μ-mean curvature h_μ arises naturally in the context of Riemannian submersions, which we encounter here whenever a compact Lie group $\mathscr{G}$ of dimension $k > 0$ acts isometrically on a Riemannian manifold X. In that situation, the principal orbits (roughly speaking, the orbits of highest dimension) foliate a dense open subset $X' \subset X$, and the submersion $X' \to X'/\mathscr{G}$ becomes Riemannian, given the right metric on $X'/\mathscr{G}$ (see [Hsiang and Lawson 1971]).

In any case, every Riemannian submersion $\pi : P \to N$ induces a *fiber volume function*

$$e^\mu : N \to (0, \infty), \qquad e^\mu(p) := |\pi^{-1}(p)|,$$

where $|\pi^{-1}(p)|$ is the k-dimensional volume of the fiber over p. A standard first-variation calculation then shows:

Observation 3.4. *The μ-mean curvature h_μ of a hypersurface $\Sigma \subset N$ gives the classical mean curvature h of its preimage $\pi^{-1}(\Sigma) \subset P$.*

In the context of an isometric $\mathscr{G}$-action as discussed above, one may then study G-invariant hypersurfaces of constant (classical) mean curvature $h \equiv H$ in X by considering, instead, hypersurfaces of constant μ-mean curvature $h_\mu \equiv H$ in the orbit space $X/\mathscr{G}$. This can be especially fruitful when $X/\mathscr{G}$ is just two- or three-dimensional. We consider examples involving twizzlers at the end of the paper.

In any case, the constrained first-variation formula (3-10) lets us extend the conservation law presented in [Korevaar et al. 1989]. As before, $\mathscr{L} \subset \mathscr{I}$ denotes a μ-preserving group of isometries on N, and the Killing fields that generate its identity component correspond to $L(\mathscr{L})$.

Theorem 3.5 (conservation law). *Suppose $\Sigma \subset N$ is an oriented hypersurface with $h_\mu \equiv H$, and B is a spine for the pair (N, Σ). Then the formula*

$$\phi_B[\boldsymbol{k}](Y) := \int_{\partial K - \beta} e^\mu \eta \cdot Y + H \int_K e^\mu \nu \cdot Y \tag{3-11}$$

yields a well-defined homomorphism

$$\phi_B : \mathscr{K}(N, \Sigma, B) \otimes L(\mathscr{L}) \to \mathbb{R}.$$

Here Y is any Killing field in $L(\mathscr{L})$, K is any cap in $\boldsymbol{k}$, and $\beta \in B$ is the spine of $\boldsymbol{k}$ given by Definition 3.3.1.

Proof. The basic linearity properties of the integral make ϕ_B a homomorphism once we establish well-definition: that $\phi_B[\boldsymbol{k}](Y)$ doesn't depend on which cap $K \in \boldsymbol{k}$ we use to compute it. We thus need to show, for all $Y \in L(\mathscr{L})$ and all $K, K' \in \boldsymbol{k}$, that

$$\int_{\partial K - \beta} e^\mu \eta \cdot Y + H \int_K e^\mu \nu \cdot Y = \int_{\partial K' - \beta} e^\mu \eta \cdot Y + H \int_{K'} e^\mu \nu \cdot Y \tag{3-12}$$

for any other $K' \in \boldsymbol{k}$. This follows easily from the constrained first-variation formula (3-10), however.

For μ is $\mathscr{L}$-invariant, and Y generates a flow that leaves both $|S|_\mu$ and $|U|_\mu$ unchanged, and hence the left-hand side of (3-10) must vanish. The integral over S on the right of (3-10) vanishes too, because $h_\mu \equiv H$. So (3-10) reduces to

$$0 = \int_{\partial K - \partial K'} e^\mu \eta \cdot Y + H \int_{K-K'} e^\mu \nu \cdot Y.$$

This is clearly equivalent to (3-12), since the integrals over β there cancel. □

Remark 3.5.1. The simplest case of Theorem 3.5, where $\mu \equiv 0$ and $\mathscr{L}$ is the full isometry group of N (so that $L(\mathscr{L})$ includes all Killing fields), already improves on the conservation law in [Korevaar et al. 1989] by eliminating the triviality assumptions there on $H_{n-1}(N)$ and $H_{n-2}(N)$.

Remark 3.5.2. The particular choice of spine B in Theorem 3.5 is of no real consequence. For when B and B' are both spines for (N, Σ), the well-definition of ϕ_B on a class in $\mathscr{K}(N, \Sigma, B)$ implies that of $\phi_{B'}$ on a corresponding class in $\mathscr{K}(N, \Sigma, B')$.

To see this, suppose ϕ_B is well-defined on a class $\boldsymbol{k}$ containing a cap K with boundary $\Gamma + \beta$, where $\beta \in B$ and Γ is supported in Σ. Then there exists a cycle $\beta' \in B'$ homologous to β, and hence an $(n-1)$-chain P with

$$\partial P = \beta' - \beta.$$

We claim $\phi_{B'}$ will now be well-defined on the class $\boldsymbol{k}'$ represented by $K + P$ in $\mathscr{K}(N, \Sigma, B')$.

Indeed, take any cap $\tilde{K}$ homologous to $K + P$ in the latter group. Then $\tilde{K} - P \in \boldsymbol{k} \in \mathscr{K}(N, \Sigma, B)$, and if ϕ_B is well-defined there for some $Y \in L(\mathscr{I})$, we have, on the one hand,

$$\phi_B(\tilde{K} - P, Y) = \phi_B(K, Y).$$

On the other hand, we have

$$\begin{aligned}
\phi_B(\tilde{K} - P, Y) &= \int_{\Gamma'} e^\mu \eta \cdot Y + H \int_{\tilde{K}-P} e^\mu \nu \cdot Y \\
&= \int_{\Gamma'} e^\mu \eta \cdot Y + H \int_{\tilde{K}} e^\mu \nu \cdot Y + H \int_P e^\mu \nu \cdot Y \\
&= \phi_{B'}(\tilde{K}, Y) + H \int_P e^\mu \nu \cdot Y.
\end{aligned}$$

Together, these facts yield

$$\phi_{B'}(\tilde{K}, Y) = \phi_B(K, Y) - H \int_P e^\mu \nu \cdot Y.$$

Since $\tilde{K}$ was arbitrary in $\boldsymbol{k}'$, while P is fixed, we see that $\phi_{B'}$ is well-defined on $\boldsymbol{k}' \in \mathcal{K}(N, \Sigma, B')$, as claimed.

4. Partial converse

Suppose the isometry group $\mathcal{I}$ of our ambient manifold N contains a closed, connected group $\mathcal{L}$ preserving a density function e^{μ} as above. Consider an immersed hypersurface $f : \Sigma \to N$, together with a spine B for the pair (N, Σ).

Above, we assumed constancy of μ-mean curvature on Σ, and deduced conservation of flux. We now seek a *converse* conservation law to the effect that well-definition of the flux functional ϕ_B implies constancy of μ-mean curvature. Well-definition of ϕ_B, however, means nothing without Killing fields on which to pose it, so the strength of any such converse must correlate with the abundance of Killing fields.

Similarly, one shouldn't need to assume well-definition of ϕ_B on *all* Killing fields to get a conservation law. We could restrict ϕ_B to a nonempty subset of $L(\mathcal{L})$ (even a singleton) and ask whether well-definition of ϕ_B there influences geometry.

Dually, we needn't assume constancy of ϕ_B on all caps. We have in mind the case where Σ is preserved by a closed, connected subgroup $\mathcal{G} \subset \mathcal{L}$ and ϕ_B takes a fixed value on a sufficiently "crowded" set of homologous $\mathcal{G}$-invariant caps.

Definition 4.0.1 ($\mathcal{G}$-crowded). A set of trivial caps $\mathcal{C} \subset \mathcal{B}_{n-1}(N, \Sigma)$ is a *$\mathcal{G}$-crowded set of boundaries* if, for every $\mathcal{G}$-orbit λ and every $\epsilon > 0$, we can find a cap $K \in C$ satisfying

$$K = \partial U - S.$$

Here U is an $(n+1)$-chain in N, and S is an n-chain in Σ, which, as an n-current, is represented by some constant multiple of a submanifold-with-boundary within distance ϵ of the orbit λ. In other words, for some constant c and any integrable function f, we have

$$\int_S f = c \int_{\operatorname{spt} S} f.$$

We say that a set $\mathcal{C}$ of *non*bounding caps in $\mathcal{Z}(N, \Sigma, B)$ is *$\mathcal{G}$-crowded* if the difference set $\{K - K' : K, K' \in \mathcal{C}\}$ forms a $\mathcal{G}$-crowded set of boundaries. Note that in this case, each $K \in \mathcal{C}$ represents the same reduced cap in $\mathcal{K}(N, \Sigma, B)$.

Example 4.0.2. For any point $p \in \Sigma$, and $\epsilon > 0$, let $V_\epsilon(p)$ be the $\mathcal{G}$-orbit of the ball $B_\epsilon(p)$. When ϵ is sufficiently small, Σ will separate V_ϵ into two open sets $V_\epsilon^{\pm}$. Then $K = \partial V_\epsilon^{+} - \Sigma \cap V_\epsilon$ will be a trivial cap, and the collection of all these K for $p \in \Sigma$ and $\epsilon > 0$ small will form a $\mathcal{G}$-crowded set.

Using this definition, we can state and prove our partial converse, which says (roughly) that when our hypersurface Σ and the density e^{μ} are preserved by a closed, connected subgroup $\mathscr{G} \subset \mathscr{I}$, and the flux is constant on a $\mathscr{G}$-crowded set of caps — with respect to Killing fields that commute with $\mathscr{G}$ — we can split Σ into two nice subsets: one with constant μ-mean curvature, and one preserved by the flows of those Killing fields. These subsets may overlap, and either can be empty, as seen in Examples 4.4.1 below.

Theorem 4.1. *Let $\Sigma \subset N$ be a complete oriented $\mathscr{G}$-invariant hypersurface, and B a spine for the pair (N, Σ). Suppose $\mathscr{C} \subset \mathscr{Z}(N, \Sigma, B)$ is a $\mathscr{G}$-crowded set of caps, and $\beta \in B$ is the spine of the reduced cap containing $\mathscr{C}$.*

If $\mathscr{G}$ preserves a Killing field Y, and the μ-weighted flux functional

$$\int_{\partial K - \beta} e^{\mu} \eta \cdot Y + H \int_K e^{\mu} \nu \cdot Y$$

is constant on $\mathscr{C}$, then the set

$$\Sigma' := \Sigma \setminus h_\mu^{-1}(H)$$

is preserved by the flow of Y.

Proof. Definition 4.0.1 and the form of the flux functional immediately show that constancy of flux on any $\mathscr{G}$-crowded set of caps in $\mathscr{Z}(N, \Sigma, B)$ forces *vanishing* of flux on a $\mathscr{G}$-crowded set of *boundaries*. So without losing generality, we may assume $\mathscr{C} \subset \mathscr{B}_{n-1}(N, \Sigma)$.

The heart of our argument then lies with the following:

Claim. *If $p \in \Sigma'$, then $Y_p \in T_p\Sigma$.*

The definition makes Σ' relatively open in Σ. Since $\mathscr{G}$ preserves Σ and μ, it preserves h_μ and hence Σ'. The $\mathscr{G}$-crowdedness of $\mathscr{C}$ ensures the existence of a cap

$$K = \partial U - S \in \mathscr{C},$$

with $S \subset \Sigma'$ supported within an arbitrarily small distance to the $\mathscr{G}$-orbit of p.

We use the volume-constrained first-variation formula with K as above, and $K' = 0$ since K bounds modulo $\Sigma' \subset \Sigma$. The first two integrals in (3-10) now vanish on our Killing field Y, since together they compute the flux of Y across a trivial cap.

This reduces the constrained first variation to a single integral:

$$\delta_Y(|S|_\mu - H\,|U|_\mu) = \int_S e^{\mu}(h_\mu - H) Y \cdot \nu.$$

Finally, since Y preserves μ, the left side of this equation must *vanish*, leaving the identity

$$\int_{\operatorname{spt} S} e^{\mu}(h_\mu - H) Y \cdot \nu = 0. \tag{4-1}$$

If $Y_p \notin T_p S$, then by assumption the integrand $(h_\mu - H)Y \cdot \nu$ is not 0 at p. Since all quantities are continuous and preserved by $\mathscr{G}$, it follows that $(h_\mu - H)Y \cdot \nu$ is strictly positive (or negative) in a neighborhood of the $\mathscr{G}$-orbit of p. Since the $\mathscr{G}$-crowdedness of $\mathscr{C}$ lets us confine the support of S to such a neighborhood, we can contradict (4-1), thereby proving the claim.

To finish proving the theorem, it suffices to show that whenever $p \in \Sigma'$, the entire Y-streamline with initial velocity Y_p lies in Σ'.

Let $T > 0$ be the maximal time such that $\Theta(p, t) \subset \Sigma'$ for *all* $t < T$. By Lemma 2.3 (with $q := p$), some such T exists. Since Y generates a μ-preserving *isometric* flow, we have $h_\mu(\Theta(p, t)) \equiv H'$ with H' constant for all $t \in [0, T)$. Moreover, $H' \neq H$, as we are in Σ'. We now claim $T = \infty$. For otherwise, the continuity of h_μ and the completeness of the larger hypersurface Σ immediately yields both $\Theta(p, T) \in \Sigma$ and $h_\mu(\Theta(p, T)) = H' \neq H$, so that $\Theta(p, T) \in \Sigma'$. But then Lemma 2.3 (with $q := \Theta(p, T)$) contradicts the maximality of T. In short, $\Theta(p, t) \in \Sigma'$ for *all* $t \geq 0$. Since the same reasoning shows that $\Theta(p, t) \in \Sigma'$ for all $t \leq 0$ too, the proof is complete. □

Remark 4.1.1. We emphasize again that our converse remains interesting even when $\mathscr{G}$ is trivial. Theorem 4.1 then implies, for instance, that when the flux across every sufficiently small trivial cap vanishes on the generators of a subgroup $\mathscr{L} \subset \mathscr{I}$, the part of Σ that does *not* have constant μ-mean curvature $h_\mu = H$ must be $\mathscr{L}$-invariant.

Corollary 4.2. *If, as in Theorem 4.1, the μ-weighted flux functional is constant on one $\mathscr{G}$-crowded set of caps, it actually extends as a well-defined conserved quantity to all of $\mathscr{K}(N, \Sigma, B)$.*

Proof. While the theorem assumes constancy of ϕ_B only on a $\mathscr{G}$-crowded set of caps, the proof then deduces that at every point $p \in \Sigma$, either $h_\mu = H$ or Y belongs to $T_p\Sigma$. In this case, the last integral in the volume-constrained first-variation formula (3-5) clearly vanishes on any $(n-1)$-chain S in Σ, so that $\phi_B(K, Y) = \phi_B(K', Y)$ for any two homologous caps $K, K' \in \mathscr{K}(N, \Sigma, B)$. □

Let us henceforth agree that when $\mathscr{G}$ is trivial, we call a $\mathscr{G}$-crowded set of caps simply *crowded*.

Corollary 4.3. *If N is homogeneous, μ is constant, and on some crowded set of caps, the flux functional is well-defined for all Killing fields on N, then Σ has mean curvature $h \equiv H$ everywhere.*

Proof. With $\mathscr{G}$ trivial in Theorem 4.1, well-definition on all Killing fields makes Σ' invariant under the entire isometry group $\mathscr{I}$. But in a homogeneous space, all nonempty $\mathscr{I}$-invariant sets have top dimension. So Σ', having codimension one, must be empty, forcing $h \equiv H$ throughout Σ. □

When $\mathcal{L} \subset \mathcal{I}$ is a subgroup, we say that N has *cohomogeneity* k with respect to $\mathcal{L}$ when the highest-dimensional orbits of $\mathcal{L}$ have codimension k in N. Cohomogeneity zero is the same as *homogeneity*.

Corollary 4.4. *Suppose a real-analytic Riemannian manifold N has cohomogeneity one with respect to a μ-preserving group $\mathcal{L}$, and on some crowded set of caps, the flux functional is well-defined on all of $L(\mathcal{L})$. Then either $h_\mu \equiv H$, or Σ is an orbit of $\mathcal{L}$. Either way, h_μ is constant on Σ.*

Proof. In an analytic ambient space, hypersurfaces with constant μ-mean curvature are analytic [Federer 1969, 5.2.16]. Cohomogeneity one means the only connected $\mathcal{L}$-invariant hypersurfaces are single orbits of $\mathcal{L}$, which clearly have constant μ-mean curvature. Since Σ is connected, the corollary now follows from Theorem 4.1. □

Examples 4.4.1. Take $N = \mathbb{R}^3$ and let $\mathcal{G}$ be the circular group acting by rotation about the x-axis. The Killing field $Y = (1, 0, 0)$ generates translational flow along that axis, and $\mathcal{G}$ commutes with this flow as required by Theorem 4.1. The noncylindrical Delaunay surfaces — CMC surfaces of revolution about the x-axis analyzed by C. Delaunay in 1841 — show that Theorem 4.1 may obtain with $\mathcal{G}$-invariant hypersurfaces having $h \equiv H$ and *no* flow-invariant subset Σ'.

In contrast, if we take Σ to be any cylinder centered about the x-axis with radius *not* equal to $1/H$, we get an example with $\Sigma' = \Sigma$. That is, Σ has mean curvature H nowhere, and yet the flux functional remains well-defined on Y, thanks to the global flow-invariance of Σ.

Of course, the cylinder of radius $1/H$ about the x-axis has *both* $h \equiv H$ *and* the extra translational symmetry.

All these possibilities arise in the family of twizzlers too, as we shall shortly see.

Case study 4.5 (first integrals for twizzlers). Consider the Riemannian product $N := \mathbb{C} \times \boldsymbol{S}^1_R$, where the complex plane $\mathbb{C}$ and $\boldsymbol{S}^1_R$ (the circle of radius R) have their standard metrics. Take $\mu \equiv 0$, and let $\mathcal{G} \approx \boldsymbol{S}^1$ act via screw-motion:

$$[e^{\mathrm{i}t}](z, Re^{\mathrm{i}\theta}) = (e^{\mathrm{i}t}z, Re^{\mathrm{i}(t+\theta)}).$$

In this situation, each helical orbit of the $\mathcal{G}$-action generates $H_1(N) \approx \mathbb{Z}$. Let $\Sigma \subset N$ be any connected $\mathcal{G}$-invariant surface, and with no loss of generality assume it does not contain the shortest orbit $\beta := \mathbf{0} \times \boldsymbol{S}^1_R$. Then β clearly generates a spine for (N, Σ).

We can parametrize Σ by letting $\mathcal{G}$ act on an immersed curve $\gamma : \mathbb{R} \to \mathbb{C} \times \{1\} \approx \mathbb{C}$ via the map

$$X(u, v) = (e^{\mathrm{i}v}\gamma(u), Re^{\mathrm{i}v}). \tag{4-2}$$

Assume the orientation of γ makes the natural frame $\{X_u, X_v\}$ *positively* oriented along Σ.

Now fix any point p on the generating curve γ, and join it to the origin in $\mathbb{C}$ by a line segment. This segment sweeps out a helicoidal cap K^p, invariant under the $\mathscr{G}$-action, and the reduced class of K^p in $\mathscr{K}(N, \Sigma, B)$ is clearly independent of p. One easily sees that as p varies over γ, the resulting caps K^p form a $\mathscr{G}$-crowded set $\mathscr{C}$ according to Definition 4.0.1.

Now let Y be the circular Killing field generating the purely "horizontal" isometric flow $[e^{is}](z, Re^{i\theta}) = (e^{is}z, Re^{i\theta})$. Note that Y commutes with $\mathscr{G}$ and preserves μ, as required by Theorem 4.1.

Finally, suppose that when we put $K = K^p$ and β as above in the flux formula of Theorem 3.5, the result is independent of p.

Since N has cohomogeneity one with respect to the extension of $\mathscr{G}$ by the flow of Y, Corollary 4.4 dictates that *either* Σ is a CMC twizzler with $h \equiv H$, or Σ is an orbit of the combined action, and thus a circular cylinder with $h \equiv 1/r$ (r giving the radius of the cylinder; typically $1/r \neq H$).

As an application of our theory, we now show that constancy of ϕ_B on the $\mathscr{G}$-crowded set of caps K^p described above "explains" the first-order ODE known to characterize generating curves of CMC twizzlers, as mentioned in our introduction.

Proposition 4.6. *A noncircular immersed curve γ in $\mathbb{C}$ generates a twizzler in $\mathbb{C} \times S^1_R$ with $h \equiv H$ if and only if, for some $c \in \mathbb{R}$, it solves*

$$\frac{2\pi R^2(\dot\gamma \cdot i\gamma)}{\sqrt{R^2|\dot\gamma|^2 + (\dot\gamma \cdot \gamma)^2}} - \pi R H |\gamma|^2 = c. \tag{4-3}$$

Proof. Since we assume γ is not circular, Theorem 3.5 and Corollary 4.4, as noted above, tell us that $h \equiv H$ if and only if the flux of the circular vector field

$$Y_{(z,\tau)} = (iz, 0)$$

across K^p is independent of p. That is,

$$\phi_B(K^p, Y) \equiv c \quad \text{for all } p \in \gamma. \tag{4-4}$$

Equation (4-3) merely evaluates this assertion.

To reach (4-3) from (4-4), we temporarily fix a point $p = \gamma(t)$ on the generating curve γ, and specify an orientation on the cap K^p, by declaring the frame field $\{K_u, K_v\}$ associated with the parametrization

$$K(u, v) = (ue^{iv}p, Re^{iv}), \quad (u, v) \in (0, 1) \times (0, 2\pi)$$

to be positively oriented.

Now consider the second integral in the flux formula (3-11) — the one that pairs Y with the unit normal ν along K^p. The correctly oriented unit normal will be a positive multiple of

$$K_u \wedge K_v = (-R\mathrm{i}e^{\mathrm{i}v}p, u|p|^2\mathrm{i}e^{\mathrm{i}v}).$$

The length of $K_u \wedge K_v$ is actually irrelevant: we divide by it to normalize, but then multiply it back in as the Jacobian in the flux integral, namely

$$\int_{K^p} \nu \cdot Y = \int_0^{2\pi}\int_0^1 (K_u \wedge K_v)\cdot Y\big|_{K(u,v)}\,du\,dv.$$

At $K(u,v)$, we have $Y = (u\mathrm{i}e^{\mathrm{i}v}p, 0)$, so the corresponding flux term evaluates easily to

$$H\int_{K^p} \nu \cdot Y = -2\pi HR|p|^2\int_0^1 u\,du = -\pi HR|p|^2. \tag{4-5}$$

Now consider the other integral in the flux formula (3-11), the integral over $\Gamma := \partial K - \beta$, where K^p meets Σ. This curve is the helical $\mathscr{G}$-orbit of p, and one easily computes its length as

$$|\Gamma| = 2\pi\sqrt{R^2+|p|^2}.$$

Our chosen orientation of K induces an orientation on Γ. Since K_u at Γ is parallel to the outer conormal in K, the velocity Γ' of Γ is equal to a positive multiple of X_v. The outer conormal in Σ along Γ, which we called η, must then give the pair $\{\eta, \Gamma'\}$ positive orientation, so we can obtain η by orthonormalizing X_u along Γ, i.e., by normalizing

$$|X_v|^2X_u - (X_u\cdot X_v)X_v.$$

Both η and Y are $\mathscr{G}$-invariant, making $\eta\cdot Y$ constant along Γ, and careful calculation then shows that indeed,

$$\eta\cdot Y \equiv \frac{R^2\dot\gamma\cdot \mathrm{i}p}{\sqrt{R^2+|p|^2}\sqrt{(R^2+|p|^2)|\dot\gamma|^2-(\gamma\cdot\mathrm{i}p)^2}},$$

where we evaluate $\dot\gamma$ at p. We can simplify the second square root in the denominator here via the elementary identity

$$(\dot\gamma\cdot\mathrm{i}p)^2 = |\dot\gamma|^2|p|^2 - (\dot\gamma\cdot p)^2.$$

This lets us express the conormal flux integral as

$$\int_\Gamma \eta\cdot Y = \frac{2\pi R^2(\dot\gamma\cdot\mathrm{i}p)}{\sqrt{R^2|\dot\gamma|^2+(\dot\gamma\cdot p)^2}}. \tag{4-6}$$

Setting $p = \gamma(t)$ and recalling (3-11), we now get $\phi_B(K^p, Y)$ by adding (4-5) to (4-6). □

Remark 4.6.1. If we parametrize a convex arc of the generating curve γ using its *support function*, namely

$$k(t) := \sup_\theta \gamma(t) \cdot e^{\mathrm{i}\theta},$$

then

$$\gamma(t) = (k(t) + \mathrm{i}\dot{k}(t))e^{\mathrm{i}t}.$$

It now follows from Proposition 4.6 that when γ generates a pitch-R twizzler with $h \equiv H$, its support function satisfies a simple nonlinear ODE:

$$\frac{2Rk}{\sqrt{R^2 + \dot{k}^2}} - H(k^2 + \dot{k}^2) = C.$$

In other words, the phase portrait of k lies on one of the "heart-shaped" level curves of the function

$$F(x, y) := \frac{2Rx}{\sqrt{R^2 + y^2}} - H(x^2 + y^2).$$

Perdomo [2012; 2013] based his dynamical characterization of twizzler generating curves and his study of their moduli space on this observation.

Remark 4.6.2 (twizzlers in other 3D-space forms). It is natural to see the curve γ in Case study 4.5 as the projection of the hypersurface Σ into the *orbit space* $N/\mathcal{G} \approx \mathbb{C}$. The length of the orbit above $z \in \mathbb{C}$ is easily computed as $|\Gamma_z| = 2\pi\sqrt{R^2 + |z|^2}$, and if we adopt this as our density function, i.e., $e^{\mu(z)} = |\Gamma_z|$, on the orbit space (see Definition 2.1.2), a simple reworking of Proposition 4.6 reinterprets the first integral there as the condition for γ to have $h_\mu \equiv H$ as a "hypersurface" in the two-dimensional orbit space.

Similarly, one can seek CMC "twizzlers" in the 3-sphere $\boldsymbol{S}^3 \subset \mathbb{R}^4$ invariant under one of the helical (k, l) "torus knot" circle actions given by

$$[e^{\mathrm{i}t}](z, w) = (e^{\mathrm{i}kt}z, e^{\mathrm{i}lt}w).$$

This is the standard Hopf action when $k = l = 1$, in which case the orbit space $\boldsymbol{S}^3/\mathcal{G}$ is of course the standard 2-sphere $\boldsymbol{S}^2$. More generally, when $\gcd(k, l) = 1$, one can realize the orbit space as an eccentric "football" or "teardrop" shaped surface of revolution in $\mathbb{R}^3$, smooth except for conical singularities at one or both ends. The $\mathcal{G}$-invariant CMC twizzlers in $\boldsymbol{S}^3$ then correspond one-to-one with curves having constant μ-mean curvature in the orbit space, where the density function is again given by orbit length: $e^{\mu(p)} = |\pi^{-1}(p)|$ for p in the orbit space. By Theorem 4.1, these $h_\mu \equiv H$ curves are precisely the noncircular curves that conserve flux along the Killing fields that generate the rotational symmetry of the orbit space. It is then straightforward to use this fact, as in Proposition 4.6, to derive the first integral they satisfy. See [Edelen 2011] for the resulting expression. We should note here that the special case $h_\mu \equiv 0$ (*minimal* twizzlers in S^3) was analyzed using Hamilton–Jacobi theory in [Hsiang and Lawson 1971, Chapter IV].

Analogous helical actions exist in the hyperbolic space form $\mathbb{H}^3$, and the resulting CMC twizzlers have a first integral derivable in precisely the same way. The reader may consult [Edelen 2011] for a description of the group action and the resulting first integral in this case as well.

Acknowledgments

The Summer 2011 Math REU program at Indiana University, Bloomington initiated and supported this work. We gratefully acknowledge the National Science Foundation for funding that program.

References

[do Carmo and Dajczer 1982] M. P. do Carmo and M. Dajczer, "Helicoidal surfaces with constant mean curvature", *Tôhoku Math. J.* (2) **34**:3 (1982), 425–435. MR 84f:53003 Zbl 0501.53003

[Edelen 2011] N. Edelen, "A conservation approach to helicoidal constant mean curvature surfaces in $\mathbb{R}^3$, $\mathbb{S}^3$ and $\mathbb{H}^3$", preprint, 2011. arXiv 1110.1068

[Federer 1969] H. Federer, *Geometric measure theory*, Grundlehren der mathematischen Wissenschaften **153**, Springer, New York, 1969. MR 41 #1976 Zbl 0176.00801

[Halldorsson 2013] H. P. Halldorsson, "Helicoidal surfaces rotating/translating under the mean curvature flow", *Geom. Dedicata* **162** (2013), 45–65. MR 3009534 Zbl 1261.53007

[Hsiang and Lawson 1971] W.-Y. Hsiang and H. B. Lawson, Jr., "Minimal submanifolds of low cohomogeneity", *J. Differential Geom.* **5** (1971), 1–38. MR 45 #7645 Zbl 0219.53045

[Korevaar et al. 1989] N. J. Korevaar, R. Kusner, and B. Solomon, "The structure of complete embedded surfaces with constant mean curvature", *J. Differential Geom.* **30**:2 (1989), 465–503. MR 90g:53011 Zbl 0726.53007

[Kusner 1991] R. Kusner, "Bubbles, conservation laws, and balanced diagrams", pp. 103–108 in *Geometric analysis and computer graphics* (Berkeley, CA, 1988), edited by P. Concus et al., Math. Sci. Res. Inst. Publ. **17**, Springer, New York, 1991. MR 91i:53012 Zbl 0776.68009

[Perdomo 2012] O. M. Perdomo, "A dynamical interpretation of the profile curve of CMC twizzler surfaces", *Pacific J. Math.* **258**:2 (2012), 459–485. MR 2981962 Zbl 1252.53007

[Perdomo 2013] O. M. Perdomo, "Helicoidal minimal surfaces in $\mathbb{R}^3$", *Illinois J. Math.* **57**:1 (2013), 87–104. MR 3224562 Zbl 1294.53015

[Simon 1983] L. Simon, *Lectures on geometric measure theory*, Proceedings of the Centre for Mathematical Analysis **3**, Australian National University, Canberra, 1983. MR 87a:49001 Zbl 0546.49019

[Wunderlich 1952] W. Wunderlich, "Beitrag zur Kenntnis der Minimalschraubflächen", *Compositio Math.* **10** (1952), 297–311. MR 14,1014a Zbl 0047.40502

Received March 24, 2014. Revised August 6, 2014.

NICK EDELEN
STANFORD UNIVERSITY
STANFORD, CA 94305
UNITED STATES
nedelen@math.stanford.edu

BRUCE SOLOMON
INDIANA UNIVERSITY
BLOOMINGTON, IN 47405
UNITED STATES
solomon@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.73

THE CYLINDRICAL CONTACT HOMOLOGY OF UNIVERSALLY TIGHT SUTURED CONTACT SOLID TORI

ROMAN GOLOVKO

We calculate the sutured version of cylindrical contact homology of a sutured contact solid torus $(S^1 \times D^2, \Gamma, \xi)$, where Γ consists of $2n$ parallel sutures of arbitrary slope and ξ is a universally tight contact structure. In particular, we show that it is nonzero. This computation is one of the first computations of the sutured version of cylindrical contact homology and does not follow from computations in the closed case.

1. Introduction

The cylindrical contact homology of a (closed) contact manifold was introduced by Eliashberg and Hofer and is the simplest version of the symplectic field theory of Eliashberg, Givental and Hofer [Eliashberg et al. 2000]. It is the homology of a differential graded module whose differential counts genus zero holomorphic curves in the symplectization with one positive puncture and one negative puncture.

In the early 1980s, Gabai [1983] developed the theory of sutured manifolds, which became a powerful tool in studying 3-manifolds with boundary. It turns out that there is a way to generalize cylindrical contact homology to sutured manifolds. This is possible by imposing a certain convexity condition on the contact form. This construction is described in the paper of Colin, Ghiggini, Honda and Hutchings [Colin et al. 2011] and will be summarized in Section 2.

In this paper, we construct a sutured contact solid torus with $2n$ parallel sutures of slope $-k/l$ using the gluing method of [Colin et al. 2011], and calculate the sutured cylindrical contact homology of it. Here $n \in \mathbb{N}$, $(k, l) = 1$ and $k > l > 0$. In order to define the slope, we choose an oriented identification $\partial(S^1 \times D^2) \simeq T^2 = (\mathbb{R}/\mathbb{Z})^2$ as follows: map $\{pt\} \times \partial D^2$ (the meridian) to $(1, 0)$ (slope is 0) and $S^1 \times \{pt\}$ (a longitude) to $(0, 1)$.

This calculation, together with the calculation of the sutured cylindrical contact homology of the sutured contact solid torus with $2n$ parallel longitudinal sutures, where $n \geq 2$, that has been done in [Golovko 2011], finishes the calculation of the cylindrical contact homology of $(S^1 \times D^2, \Gamma, \xi)$, where Γ consists of $2n$ parallel

MSC2010: primary 53D42; secondary 57M50, 53D10.
Keywords: sutured manifolds, contact homology.

sutures of arbitrary slope, ξ is a universally tight contact structure and such that if one cuts along the meridian disk, the sutures on the disk are ∂-parallel. In particular, this gives a complete calculation of the cylindrical contact homology of $(S^1 \times D^2, \Gamma, \xi)$, where Γ consists of 2 parallel sutures of arbitrary slope and ξ is a universally tight contact structure (observe that in this situation there are only two isomorphic (but not isotopic) universally tight contact structures; see [Honda 2002, Section 2]). These are not all the universally tight contact structures on the solid torus, but all of them can be obtained from the $\#\Gamma = 2$ case by successively applying the folding operation.

Our goal is to prove the following theorem:

Theorem 1.1. *Let $(S^1 \times D^2, \Gamma)$ be a sutured manifold, where Γ is a set of $2n$ parallel closed curves of slope $-k/l$, where $(k, l) = 1$, $k > l > 0$ and $n \in \mathbb{N}$. Then there is a contact form α which makes $(S^1 \times D^2, \Gamma, \alpha)$ a sutured contact manifold with a universally tight contact structure $\xi = \ker \alpha$, $HC^{\mathrm{cyl}}(S^1 \times D^2, \Gamma, \alpha)$ is defined, is independent of the contact form α for $\xi = \ker \alpha$ and the almost complex structure J and*

$$HC^{\mathrm{cyl},h}(S^1 \times D^2, \Gamma, \xi) \simeq \begin{cases} \mathbb{Q} & \text{for } k \nmid h > 0, \\ \mathbb{Q}^{n-1} & \text{for } k \mid h > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here h corresponds to the homological grading.

2. Background

The goal of this section is to review definitions of sutured contact manifold and the relative version of cylindrical contact homology. This section can be considered as a summary of [Colin et al. 2011].

2A. *Review of sutured contact manifolds.* In this section, we recall some definitions and describe some constructions from [Colin et al. 2011]. We first start with the notion of a Liouville manifold.

Definition 2.1. A *Liouville manifold* (often also called a Liouville domain) is a pair (W, β) consisting of a compact, oriented $2n$-dimensional manifold W with boundary and a 1-form β on W, where $\omega = d\beta$ is a positive symplectic form on W and the *Liouville vector field Y* given by $i_Y(\omega) = \beta$ is positively transverse to ∂W. It follows that the 1-form $\beta_0 = \beta|_{\partial W}$ (this notation means β pulled back to ∂W) is a positive contact form with kernel ζ.

We now recall the definition of a sutured contact manifold.

Definition 2.2. A compact oriented $2n+1$-dimensional manifold M with boundary and corners is a *sutured contact manifold* if it comes with an oriented, not necessarily connected submanifold $\Gamma \subset \partial M$ of dimension $2n-1$ (called the *suture*), together with a neighborhood $U(\Gamma) = [-1, 0] \times [-1, 1] \times \Gamma$ of $\Gamma = \{0\} \times \{0\} \times \Gamma$ in M, with coordinates $(\tau, t) \in [-1, 0] \times [-1, 1]$, such that the following holds:

(1) $U \cap \partial M = (\{0\} \times [-1, 1] \times \Gamma) \cup ([-1, 0] \times \{-1\} \times \Gamma) \cup ([-1, 0] \times \{1\} \times \Gamma)$.

(2) $\partial M \setminus (\{0\} \times (-1, 1) \times \Gamma) = R_-(\Gamma) \sqcup R_+(\Gamma)$, where the orientation of ∂M agrees with that of $R_+(\Gamma)$ and is opposite that of $R_-(\Gamma)$ and the orientation of Γ agrees with the boundary orientation of $R_+(\Gamma)$.

(3) The corners of M are precisely $\{0\} \times \{\pm 1\} \times \Gamma$.

In addition, M is equipped with a contact structure ξ, which is given by the kernel of a positive contact 1-form α such that

(i) $(R_\pm(\Gamma), \beta_\pm = \alpha|_{R_\pm(\Gamma)})$ is a Liouville manifold;

(ii) $\alpha = C\,dt + \beta$ inside $U(\Gamma)$, where $C > 0$ and β is independent of t and does not have a dt-term;

(iii) $\partial_\tau = Y_\pm$, where $Y_\pm$ is a Liouville vector field for $\beta_\pm$.

Such a contact form α is said to be *adapted* to $(M, \Gamma, U(\Gamma))$.

Here we briefly describe the way to glue sutured contact manifolds. This procedure was first described by Colin and Honda [2005] and then generalized by Colin et al. [2011].

Let $(M', \Gamma', U(\Gamma'), \xi')$ be a sutured contact 3-manifold with an adapted contact form α'. We denote by π the projection along ∂_t defined on $U(\Gamma')$.

Take 2-dimensional submanifolds $P_\pm \subset R_\pm(\Gamma')$ such that $\partial P_\pm$ is the union of $(\partial P_\pm)_\partial \subset \partial R_\pm(\Gamma')$, $(\partial P_\pm)_{\mathrm{int}} \subset \mathrm{int}(R_\pm(\Gamma'))$ and $\partial P_\pm$ is positively transversal to the Liouville vector field $Y'_\pm$ on $R_\pm(\Gamma')$. Whenever we refer to $(\partial P_\pm)_{\mathrm{int}}$ and $(\partial P_\pm)_\partial$, we assume that closures are taken as appropriate. Moreover we make the assumption that $\pi((\partial P_-)_\partial) \cap \pi(\partial P_+)_\partial) = \varnothing$.

Let φ be a diffeomorphism which sends $(P_+, \beta'_+|_{P+})$ to $(P_-, \beta'_-|_{P_-})$ and takes $(\partial P_+)_{\mathrm{int}}$ to $(\partial P_-)_\partial$ and $(\partial P_+)_\partial$ to $(\partial P_-)_{\mathrm{int}}$. Note that, since $\dim M = 3$, we only need $\beta'_+|_{P_+}$ and $\varphi^*(\beta'_-|_{P_-})$ to match up on ∂P_+, since we can linearly interpolate between primitives of positive area forms on a surface.

Topologically, we construct the sutured manifold (M, Γ) from (M', Γ') and the gluing data (P_+, P_-, φ) as follows: Let $M = M'/\sim$, where

- $x \sim \varphi(x)$ for all $x \in P_+$;
- $x \sim x'$ if $x, x' \in \pi^{-1}(\Gamma')$ and $\pi(x) = \pi(x') \in \Gamma'$.

Then

$$R_\pm(\Gamma) = \frac{\overline{R_\pm(\Gamma') \setminus P_\pm}}{(\partial P_\pm)_{\mathrm{int}}} \sim \pi_\pm((\partial P_\mp)_\partial)$$

and

$$\Gamma = \frac{\overline{\Gamma' \setminus \pi(\partial P_+ \sqcup \partial P_-)}}{\pi((\partial P_+)_{\text{int}} \cap (\partial P_+)_\partial)} \sim \pi((\partial P_-)_{\text{int}} \cap (\partial P_-)_\partial).$$

For a detailed description of the gluing procedure, see [Colin et al. 2011].

Finally, we describe the way to complete sutured contact manifold (M, α) to a noncompact contact manifold (M^*, α^*). This construction was first described in [Colin et al. 2011].

Let $(M, \Gamma, U(\Gamma), \xi)$ be a sutured contact manifold with an adapted contact form α. The form α is then given by

$$C\,dt + \beta_\pm$$

on collar neighborhoods $[1-\varepsilon, 1] \times R_+(\Gamma)$ and $[-1, -1+\varepsilon] \times R_-(\Gamma)$ of $R_+(\Gamma) = \{1\} \times R_+(\Gamma)$ and $R_-(\Gamma) = \{-1\} \times R_-(\Gamma)$, where $t \in [-1, -1+\varepsilon] \cup [1-\varepsilon, 1]$ extends the t-coordinate on U. On U we have $\alpha = C\,dt + \beta$, $\beta = \beta_+ = \beta_-$ and ∂_τ is a Liouville vector field Y for β. We first extend α to $[1, \infty) \times R_+(\Gamma)$ and $(-\infty, -1] \times R_-(\Gamma)$ by taking $C\,dt + \beta_\pm$ as appropriate. The boundary of this new manifold is $\{0\} \times \mathbb{R} \times \Gamma$. Notice that since $\partial_\tau = Y$, the form $d\beta|_{[-1,0]\times\{t\}\times\Gamma}$ is the symplectization of $\beta|_{\{0\}\times\{t\}\times\Gamma}$ in the positive τ-direction. We glue $[0, \infty) \times \mathbb{R} \times \Gamma$ with the form $C\,dt + e^\tau \beta_0$, where β_0 is the pullback of β to $\{0\} \times \{t\} \times \Gamma$.

We denote by M^* the noncompact extension of M described above and by α^* the extension of α to M^*.

2B. ***Review of cylindrical contact homology.*** In this section, we review the definition of cylindrical contact homology for sutured manifolds. We refer to [Colin et al. 2011] for more details of this construction.

Let $(M, \Gamma, U(\Gamma), \xi)$ be a sutured contact manifold with an adapted contact form α and (M^*, α^*) be its completion.

The *Reeb vector field* R_{α^*} that is associated to a contact form α^* is given by $d\alpha^*(R_{\alpha^*}, \cdot) = 0$ and $\alpha^*(R_{\alpha^*}) = 1$. We assume that R_{α^*} is *nondegenerate*, i.e., the first return map along each (not necessarily simple) periodic orbit does not have 1 as an eigenvalue. Observe that nondegeneracy can always be achieved by a small perturbation.

Remark 2.3. Every periodic orbit of R_{α^*} lies in M. Hence, the set of periodic Reeb orbits of R_{α^*} coincides with the set of periodic Reeb orbits of R_α.

A Reeb orbit γ is called *elliptic* or *positive* (respectively *negative*) *hyperbolic* if the eigenvalues of P_γ are on the unit circle or the positive (resp. negative) real line respectively.

If τ is a trivialization of ξ over γ, we can then define the Conley–Zehnder index. In 3-dimensional situation, we can explicitly describe the Conley–Zehnder index and its behavior under multiple covers as follows:

Proposition 2.4 [Hutchings 2002]. *If γ is elliptic, then there is an irrational number $\phi \in \mathbb{R}$ such that P_γ is conjugate in $SL_2(\mathbb{R})$ to a rotation by angle $2\pi\phi$ and*

$$\mu_\tau(\gamma^k) = 2\lfloor k\phi \rfloor + 1,$$

where $2\pi\phi$ is the total rotation angle with respect to τ of the linearized flow around the orbit.

If γ is positive (respectively negative) hyperbolic, then there is an even (respectively odd) integer r such that the linearized flow around the orbit rotates the eigenspaces of P_γ by angle πr with respect to τ and

$$\mu_\tau(\gamma^k) = kr.$$

A closed orbit of R_{α^*} is said to be *good* if it does not cover a simple orbit γ an even number of times, where the first return map $\xi_{\gamma(0)} \to \xi_{\gamma(T)}$ has an odd number of eigenvalues in the interval $(-1, 0)$. Here T is the period of the orbit γ. An orbit that is not good is called *bad*.

We now recall the notion of an almost complex structure on $\mathbb{R} \times M^*$ that is tailored to (M^*, α^*).

Let (W, β) be a Liouville manifold and ζ be the contact structure given on ∂W by $\ker(\beta_0)$, where $\beta_0 = \beta|_{\partial W}$. In addition, An almost complex structure J_0 on $\widehat{W}$ is *$\hat{\beta}$- adapted* if J_0 is β_0-adapted on $[0, \infty) \times \partial W$; and $d\beta(v, J_0 v) > 0$ for all nonzero tangent vectors v on W.

Definition 2.5. Let $(M, \Gamma, U(\Gamma), \xi)$ be a sutured contact manifold, α be an adapted contact form and (M^*, α^*) be its completion. We say that an almost complex structure J on $\mathbb{R} \times M^*$ is *tailored to* (M^*, α^*) if the following conditions hold:

(1) J is α^*-adapted, i.e., J is $\mathbb{R}$-invariant, $J(\xi) = \xi$, $d\alpha(v, Jv) > 0$ for nonzero $v \in \xi$ and $J(\partial_s) = R_{\alpha^*}$, where s denotes the $\mathbb{R}$-coordinate.

(2) J is ∂_t-invariant in a neighborhood of $M^* \setminus \mathrm{int}(M)$.

(3) The projection of J to $T\widehat{R_\pm(\Gamma)}$ is a $\hat{\beta}_\pm$-adapted almost complex structure J_0 on the completion $(\widehat{R_+(\Gamma)}, \hat{\beta}_+) \sqcup (\widehat{R_-(\Gamma)}, \hat{\beta}_-)$ of the Liouville manifold $(R_+(\Gamma), \beta_+) \sqcup (R_-(\Gamma), \beta_-)$. Moreover, the flow of ∂_t identifies $J_0|_{\widehat{R_+(\Gamma)} \setminus R_+(\Gamma)}$ and $J_0|_{\widehat{R_-(\Gamma)} \setminus R_-(\Gamma)}$.

Given a sutured contact manifold $(M, \Gamma, U(\Gamma), \alpha)$ and an α^*-adapted almost complex structure J, we define the *sutured cylindrical contact homology group* $HC^{\mathrm{cyl}}(M, \Gamma, \alpha, J)$ to be the cylindrical contact homology of (M^*, α^*, J). The cylindrical contact homology chain complex $C(\alpha, J)$ is a $\mathbb{Q}$-module freely generated by all good Reeb orbits, where the grading $|\cdot|$ and the boundary map ∂ are defined as in [Bourgeois 2009] with respect to the α^*-adapted almost complex structure

J. The homology of $C(\alpha, J)$ is the sutured cylindrical contact homology group $HC^{\mathrm{cyl}}(M, \Gamma, \alpha, J)$.

For our calculations we need the following construction of a "global" symplectic trivialization described in [Bourgeois 2009]. Assume that all the Reeb orbits of R_α are good. Let us now choose trivializations $\tau(\gamma)$ consistently for all Reeb orbits γ. Assume that $H_1(M; \mathbb{Z})$ is a free module. We pick representatives $C_1, \ldots, C_s$ in $H_1(M; \mathbb{Z})$ for a basis of $H_1(M; \mathbb{Z})$, together with a trivialization of ξ along each representative C_i, $i = 1, \ldots, s$. Now for a Reeb orbit γ, we distinguish the following cases:

(1) $[\gamma] = 0 \in H_1(M; \mathbb{Z})$. Choose a spanning surface S_γ and use it to trivialize ξ along γ.

(2) $0 \neq [\gamma] \in H_1(M; \mathbb{Z})$. We choose a surface S_γ realizing a homology between γ and a linear combination of the representatives C_i, $i = 1, \ldots, s$. We then use S_γ to extend the chosen trivializations of ξ along the C_i to γ.

We denote the obtained trivialization by τ.

To a J-holomorphic cylinder in $\mathcal{M}^J(\gamma; \gamma')$, we can glue the chosen surfaces S_γ and $S_{\gamma'}$ and obtain a closed surface in M (here $\mathcal{M}^J(\gamma; \gamma')$ is a moduli space of J-holomorphic cylinders considered in cylindrical contact homology theory). Let $A \in H_2(M; \mathbb{Z})$ be its homology class; we can use it to decorate the corresponding connected component $M^J_A(\gamma; \gamma')$ of the moduli space. Using τ we can write

$$\mathrm{ind}(u) = |\gamma| - |\gamma'| + 2\langle c_1(\xi), A\rangle \tag{2B.1}$$

for $u \in \mathcal{M}^J_A(\gamma; \gamma')$, where $|\gamma|$ is the *Conley–Zehnder grading of* γ defined by

$$|\gamma| := \mu_\tau(\gamma) - 1. \tag{2B.2}$$

We will use (2B.1) and (2B.2) for our calculations.

In addition, we will need the following fact, which is a consequence of Lemma 5.4 in [Bourgeois et al. 2003]:

Fact 2.6. *Let (M, α) be a closed, oriented contact manifold with nondegenerate Reeb orbits and $u \in \mathcal{M}^J(\gamma; \gamma')$, where γ and γ' are good Reeb orbits and J is an α-adapted almost complex structure on $\mathbb{R} \times M$. Then $\mathcal{A}(\gamma) := \int_\gamma \alpha \geq \int_{\gamma'} \alpha =: \mathcal{A}(\gamma')$ with equality if and only if $\gamma = \gamma'$ and in this case the moduli space consists of a single element $\mathbb{R} \times \gamma$.*

Theorem 2.7 [Bourgeois 2009]. *Let (M, α) be a closed, oriented contact manifold with nondegenerate Reeb orbits. Let $C^h_m(M, \alpha)$ be the cylindrical contact homology complex, where h is a homotopy class of Reeb orbits and m corresponds to the Conley–Zehnder grading. If $C^0_k(M, \alpha) = 0$ for $k = -1, 0, 1$, we have for every free homotopy class h:*

(1) $\partial^2 = 0$.

(2) $H(C_*^h(M, \alpha), \partial)$ *is independent of the contact form* α *for* ξ, *the almost complex structure* J *and the choice of perturbation for the moduli spaces.*

When M is closed and $\mathbb{R} \times M$ is 4-dimensional, the following transversality result has been proven by Momin [2011, Proposition 2.10]:

Theorem 2.8 [Momin 2011]. *Let* $u \in \mathcal{M}^J(\gamma; \gamma')$ *be such that* $\operatorname{ind}(u) = 1$. *Then the linearization of the Cauchy–Riemann operator is surjective at* u.

Remark 2.9. Theorem 2.8 does not require J to be generic. In addition, note that Theorem 2.8 can be considered as a consequence of the automatic transversality result of Wendl [2010, Theorem 0.1].

Finally, we recall the following result of Colin, Ghiggini, Honda and Hutchings:

Theorem 2.10 [Colin et al. 2011]. *Let* $(M, \Gamma, U(\Gamma), \xi)$ *be a sutured contact* 3-*manifold with an adapted contact form* α, (M^*, α^*) *be its completion and* J *be an almost complex structure on* $\mathbb{R} \times M^*$ *which is tailored to* (M^*, α^*). *Then the contact homology algebra* $HC(M, \Gamma, \xi)$ *is defined and independent of the choice of contact* 1-*form* α *with* $\ker(\alpha) = \xi$, *adapted almost complex structure* J, *and abstract perturbation.*

Remark 2.11. Fact 2.6, Theorems 2.7 and 2.8 and formulas (2B.1) and (2B.2) hold for J-holomorphic curves in the symplectization of the completion of a sutured contact manifold, provided that we choose the almost complex structure J on $\mathbb{R} \times M^*$ to be tailored to (M^*, α^*).

Remark 2.12. Theorem 2.10 and Remark 2.11 rely on the assumption that the machinery, needed to prove the analogous properties for contact homology and cylindrical contact homology in the closed case, works.

3. Construction

The goal of this section is to construct the sutured contact solid torus $(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta)$, where $\tilde{\Gamma}$ consists of $2n$ parallel sutures of slope $-k/l$, $(k, l) = 1$, $k > l > 0$ and $n \in \mathbb{N}$. Here $\tilde{\alpha}_\delta$ is a contact form such that $\xi = \ker \tilde{\alpha}_\delta$ is a universally tight contact structure and the set of embedded orbits of $R_{\tilde{\alpha}_\delta}$ consists of an elliptic orbit γ and hyperbolic orbits $\gamma_1, \dots, \gamma_n$ with

$$\begin{aligned} &[\gamma] = 1, && \mu_\tau(\gamma_i^s) = -2ls, && \mathcal{A}(\gamma^k) > \mathcal{A}(\gamma_i), \\ &[\gamma_i] = k \in \mathbb{Z} \simeq H_1(S^1 \times D^2; \mathbb{Z}), && \mu_\tau(\gamma^t) = -2ml + 1, \end{aligned}$$

where $(m-1)k < t \leq mk$, for some "global" symplectic trivialization τ. Here $i = 1, \dots, n$, $t \leq N_\delta$, $s \leq N_\delta/k$, $N_\delta \gg 0$.

3A. ***Gluing map.*** First we construct $H \in C^\infty(\mathbb{R}^2)$. The time-1 flow of the Hamiltonian vector field associated to H composed with an appropriate rotation will play a role of the gluing map when we will apply the gluing construction described in Section 2A to the sutured contact solid cylinder constructed in Section 3B.

We fix $p \in \mathbb{R}^2$ and consider $H_{\text{sing}} : \mathbb{R}^2 \to \mathbb{R}$ given by $H_{\text{sing}} = \mu r^2 \cos(nk\theta)$ in polar coordinates (r, θ) about p, where $\mu > 0$, $n \geq 1$ and $k \in \mathbb{N} \setminus \{1\}$. Note that H_{sing} is singular only at p.

Lemma 3.1. *There exists a function $H \in C^\infty(\mathbb{R}^2)$ which satisfies the following properties*:

- *$H = H_{\text{sing}}$ on $\mathbb{R}^2 \setminus D(r_{\text{sing}})$ for some $r_{\text{sing}} > 0$.*
- *H is $\frac{2\pi}{nk}$-symmetric with respect to θ.*
- *The set of critical points of H consists of equally spaced saddle points $p_1, \ldots, p_{nk}$ and a critical point p.*
- *There exists a neighborhood U_s of p_s with coordinates (x, y) such that $H = axy$ on U_s with $a > 0$, and such that $\frac{2\pi}{nk}$-rotation about p that we call R_{nk} maps U_s with the corresponding coordinate system to U_{s+1} with the corresponding coordinate system for $s = 1, \ldots, nk$.*
- *There exists a neighborhood U of p such that $H = \tilde{B} r^2 - \tilde{C}$ on U, where $\tilde{C} > 0$ and $\tilde{B}$ is a small positive number.*

Proof. We construct $H \in C^\infty(\mathbb{R}^2)$ from H_{sing} by perturbing H_{sing} on a disk $D(r_{\text{sing}})$ about p in such a way that H has nk equally spaced saddle points, critical point at p and interpolates with no other critical points with H_{sing}. In other words, $H = H_{\text{sing}}$ on $\mathbb{R}^2 \setminus D(r_{\text{sing}})$ for some $r_{\text{sing}} > 0$. For the level sets of H_{sing} and H in the case $n = 1$, $k = 3$ we refer to Figure 1.

The construction of H is a modification of the construction described in [Cotton-Clay 2009].

We proceed in four steps.

(1) We consider

$$
\begin{aligned}
H_1 &= H_{\text{sing}} + f(r, \theta) = H_{\text{sing}} + f_{\exp}(r, \theta) + g(r, \theta) \\
&= \mu r^2 \cos(nk\theta) - A e^{-mr^2} + g(r, \theta),
\end{aligned}
$$

where A and m are positive constants, and $g(r, \theta)$ is a smooth function to be chosen later. We are interested in the critical points of H_1 away from the origin.

We calculate

$$
\frac{\partial H_1}{\partial r} = 2\mu r \cos(nk\theta) + 2mrAe^{-mr^2} + \frac{\partial g}{\partial r}, \qquad \frac{\partial H_1}{\partial \theta} = -nk\mu r^2 \sin(nk\theta).
$$

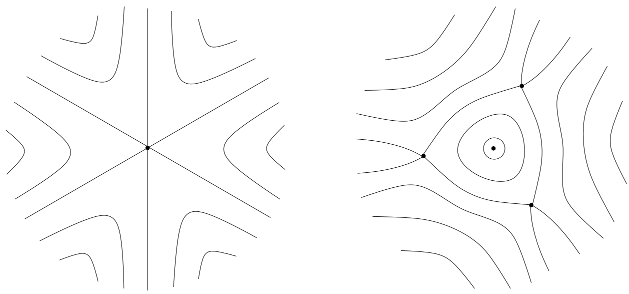

Figure 1. The level sets of H_{sing} (left) and the level sets of H (right) in the case $n = 1$, $k = 3$.

Thus, at the critical points of H_1 we must have $\sin(nk\theta) = 0$. In this case, $\cos(nk\theta) = \pm 1$. If $\cos(nk\theta) = 1$, then $\partial H_1/\partial r - \partial g/\partial r$ cannot be zero. When $\cos(nk\theta) = -1$, $\partial H_1/\partial r - \partial g/\partial r = -2\mu r + 2mrAe^{-mr^2}$. For $r > 0$, $\partial H_1/\partial r - \partial g/\partial r = 0$ when $e^{mr^2} = mA/\mu$, i.e., when

$$r = r_c := \sqrt{\frac{1}{m} \ln \frac{mA}{\mu}}.$$

We impose the restriction that $mA > \mu$. By making m large, we can make r_c arbitrarily small. When $\cos(nk\theta) = -1$, $H_1 - g(r,\theta) = -\frac{\mu}{m}\big(\ln(mA/\mu) + 1\big)$. Let $g(r)$ be equal to $\frac{\mu}{m}\big(\ln(mA/\mu)+1\big)$ on the annular neighborhood of $r = r_c$. For such g, H_1 is 0 at the critical points, i.e., at the points (r_c, θ), where $\cos(nk\theta) = -1$.

In summary, we get critical points at one value of r at the values of θ when $\cos(nk\theta) = -1$, that is, for nk values of θ. These are our nk saddle points (it's not hard to see they are saddle points; alternatively, we can deduce that they must be for index reasons).

(2) Keeping f_{exp} solely a function of r and keeping g constant, we cut off f_{exp} smoothly starting at some point past r_c to give a Hamiltonian H_2 which agrees with $H_{\text{sing}} + g$ outside a ball. As long as $\partial f_{\text{exp}}/\partial r < 2\mu r$, there are no new critical points.

Note that $f_{\text{exp}}(r_c) = -\mu/m$. Keeping $\partial f_{\text{exp}}/\partial r$ near μr_c (which, using, e.g., $A = e\mu/m$, is $1/\sqrt{m}$), we can bring f_{exp} to zero in a radial distance of a constant times $1/\sqrt{m}$; that is, for m large we can make H_2 agree with $H_{\text{sing}} + g$ outside an arbitrarily small ball.

For $A = e\mu/m$, $g = 2\mu/m$. Then keeping g solely a function of r, we cut off $g(r,\theta)$ smoothly starting at some point past the point where $H_2 = H_{\text{sing}} + g$ to give Hamiltonian H_3. As long as $\partial g/\partial r > -2\mu r$, there are no new critical points. We can make it in such a way that H_3 agrees with H_{sing} outside a small ball.

(3) Recall that $H_3 = H_{\text{sing}} + f_{\text{exp}} + g$ near the origin and $g(r,\theta) = 2\frac{\mu}{m} > 0$. Note that $g(r,\theta)$ is small for large m. Now keeping g constant we modify $H_{\text{sing}} + f_{\text{exp}} + g$ near the origin to give us H_4 which is $Br^2 - C$ near the origin (for $B > 0$), which corresponds to the Hamiltonian flow rotating at a constant angular rate. Since

$$\frac{\partial H_3}{\partial r} = \frac{\partial(H_{\text{sing}} + f_{\text{exp}})}{\partial r} > 0 \quad \text{for } r < r_c,$$

we can patch together $Br^2 - C$ near the origin with H_2 outside a small ball of radius less than r_c in a radially symmetric manner to get H_4 such that $\partial H_4/\partial r > 0$ for $r < r_c$ (we do this by choosing C sufficiently large). Note that H_4 has a critical point at the origin.

(4) Finally, to ensure no fixed points of the time-1 flow of the Hamiltonian vector field of H, we let H be H_4 multiplied by a radially symmetric function which is ϵ for $r < R$ (for ϵ sufficiently small that the only fixed points of the time-1 flow inside radius R are the critical points and for R large enough that H_4 agrees with H_{sing} for $r > R$) and 1 for $r > 2R$. This creates no new fixed points in the region $R < r < 2R$ because H_4 and $\partial H_4/\partial r$ have the same sign there. Now there are no fixed points of the time-1 flow of the Hamiltonian vector field of H, except for the $nk+1$ critical points of H because outside radius R there are no compact flow lines.

Let $p_1, \dots, p_{nk}$ denote the equally spaced saddle points of H ordered counterclockwise, i.e., $R_{nk}(p_i) = p_{i+1}$, where R_{nk} corresponds to the $\frac{2\pi}{nk}$-rotation around p. We note that $H(p_s) = 0$ for $s = 1, \dots, nk$. Hence, by Morse lemma (arguing the same way as in Lemma 3.2 in [Golovko 2011]) we get that there is a neighborhood U_s of p_s such that $H = axy$ on U_s, where $s = 1, \dots, nk$ and $a > 0$. In addition, observe that H is $\frac{2\pi}{nk}$-symmetric with respect to θ. Therefore, the U_s together with coordinates (x, y) are $\frac{2\pi}{nk}$-symmetric with respect to θ, i.e., $R_{nk}(U_s) = U_{s+1}$ and coordinates on U_s maps to the coordinate on U_{s+1}. Finally, note that $H = \tilde{B}r^2 - \tilde{C}$ on a neighborhood of the center of $D(r_{\text{sing}})$, which we call U, where $\tilde{C} > 0$ and $\tilde{B}$ is a small positive number and hence Hamiltonian flow rotates at a constant rate near the origin. □

3B. ***Sutured contact solid cylinder.*** In this section, we construct the sutured contact solid cylinder that we later will glue to get the sutured contact solid torus with $2n$ sutures of slope $-k/l$, where $n \in \mathbb{N}$, $(k, l) = 1$ and $k > l > 0$.

Let γ_{p,p_s} be an embedded curve in $\mathbb{R}^2$ which starts at p and ends at p_s for $s = 1, \dots, nk$. For the time being, we can think about γ_{p,p_s} as about the segment connecting p and p_s.

Lemma 3.2. *There exists a* 1*-form* β *on* $\mathbb{R}^2$ *satisfying the following*:

(1) $d\beta > 0$.

(2) *Its singular foliation given by* $\ker\beta$ *has isolated singularities and no closed orbits.*

(3) $\beta = \frac{1}{2}\varepsilon_c r^2 d\theta$ *on* U *with respect to the polar coordinates whose origin is at the center of* $D(r_{\text{sing}})$; $\beta = \frac{1}{2}\varepsilon_{\text{sym}}(x\,dy - y\,dx)$ *on* U_s *with respect to the coordinates from Lemma 3.1, where* $s \in \{1, \dots, nk\}$; $\beta = \frac{1}{2}r^2 d\theta$ *on* $\mathbb{R}^2 \setminus D(r_{\text{sing}})$ *with respect to the polar coordinates whose origin is at the center of* $D(r_{\text{sing}})$; *here* $0 < \varepsilon_c \ll \varepsilon_{\text{sym}} \ll 1$.

(4) *The set of hyperbolic points of the singular foliation of* β *is given by* $\{q_s\}_{s=1}^{nk}$ *such that* q_s *lies on* γ_{p,p_s} *outside of* U_s *and* U.

(5) β *is* $\frac{2\pi}{nk}$*-symmetric, i.e.,* $R_{nk}^*(\beta) = \beta$.

Proof. Consider a singular foliation $\mathcal{F}$ on $\mathbb{R}^2$ which satisfies the following:

(1) $\mathcal{F}$ is Morse–Smale and has no closed orbits.

(2) The singular set of $\mathcal{F}$ consists of elliptic points and hyperbolic points. The elliptic points are the equally spaced saddle points of H and the center of $D(r_{\text{sing}})$. The set of hyperbolic points of the singular foliation of β is given by $\{q_s\}_{s=1}^{nk}$ such that q_s lies on γ_{p,p_s} outside of U_s and U.

(3) $\mathcal{F}$ is oriented and for one choice of orientation the flow is transverse to and exits from $\partial D(r_{\text{sing}})$.

(4) $\mathcal{F}$ is $\frac{2\pi}{nk}$-symmetric with respect to θ.

Next, we modify $\mathcal{F}$ near each of the singular points so that $\mathcal{F}$ is given by $\beta_0 = \frac{1}{2}(x\,dy - y\,dx)$ on U_s with respect to the coordinates from Lemma 3.1 and $\beta_0 = 2x\,dy + y\,dx$ near a hyperbolic point. On $\mathbb{R}^2 \setminus D(r_{\text{sing}})$, $\beta_0 = \frac{1}{2}r^2 d\theta$ with respect to the polar coordinates whose origin is at the center of $D(r_{\text{sing}})$. In addition, on U, $\beta_0 = \frac{1}{2}r^2 d\theta$ with respect to the polar coordinates whose origin is at the center of $D(r_{\text{sing}})$. From Lemma 3.1 it follows that we can do it in such a way that the modification of $\mathcal{F}$ is still $\frac{2\pi}{nk}$-symmetric. Finally, we get $\mathcal{F}$ given by β_0, which satisfies $d\beta_0 > 0$ near the singular points and on $\mathbb{R}^2 \setminus D(r_{\text{sing}})$. Now let $\beta = g\beta_0$, where g is a positive function with $dg(X) \gg 0$ outside of

$$U \cup \left(\bigcup_{s=1}^{nk} U_s\right) \cup (\mathbb{R}^2 \setminus D(r_{\text{sing}})),$$

$g|_{\bigcup_{s=1}^{nk} U_s} = \varepsilon_{\text{sym}},\ g|_U = \varepsilon_c,\ g|_{\mathbb{R}^2 \setminus D(r_{\text{sing}})} = 1$ and X is an oriented vector field for $\mathcal{F}$ (nonzero away from the singular points). Here $0 < \varepsilon_c \ll \varepsilon_{\text{sym}} \ll 1$. Since $d\beta = dg \wedge \beta_0 + g \wedge d\beta_0$, $dg(X) \gg 0$ guarantees that $d\beta > 0$. $\square$

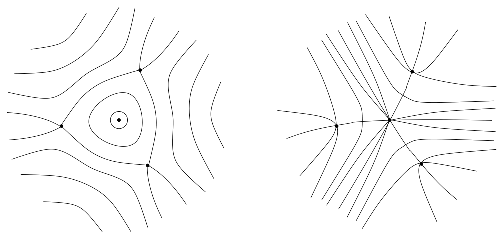

Figure 2. The level sets of H (left) and the characteristic foliation of β (right) in the case $n=1$, $k=3$.

For the comparison of the level sets of H with the singular foliation of β in the case $n=1$, $k=3$ we refer to Figure 2.

Lemma 3.3. *Let β be a 1-form from Lemma 3.2. The Hamiltonian vector field X_H of H with respect to the area form $d\beta$ satisfies $\beta(X_H)=H$ on*

$$\left(\bigcup_{s=1}^{nk} U_s\right)\cup(\mathbb{R}^2\setminus D(r_{\mathrm{sing}})).$$

In addition, the Hamiltonian vector field X_H of H with respect to the area form $d\beta$ satisfies $\beta(X_H)-H=\tilde{C}$ on U.

Proof. First, Lemmas 3.1 and 3.2 imply that $\beta=\frac{1}{2}\varepsilon_c r^2 d\theta$, $H=\tilde{B}r^2-\tilde{C}$ on U and ε_c is a small positive number. Now we show that $X_H=\frac{2\tilde{B}}{\varepsilon_c}\frac{\partial}{\partial\theta}$ is a solution of $\beta(X_H)-H=\tilde{C}$ on U. We calculate

$$i_{X_H}(d\beta)=\left(\frac{2\tilde{B}}{\varepsilon_c}\frac{\partial}{\partial\theta}\right)\lrcorner(\varepsilon_c r\,dr\wedge d\theta)=-2\tilde{B}r\,dr=-dH$$

and

$$\beta(X_H)-H=\tfrac{1}{2}\varepsilon_c r^2 d\theta\left(\frac{2\tilde{B}}{\varepsilon_c}\frac{\partial}{\partial\theta}\right)-\tilde{B}r^2+\tilde{C}=\tilde{C}.$$

Next, we work on U_s, where $s=1,\dots,nk$. From Lemmas 3.1 and 3.2 it follows that $\beta=\frac{1}{2}\varepsilon_{\mathrm{sym}}(x\,dy-y\,dx)$ and $H=axy$ on U_s. Let X_H be a Hamiltonian vector field defined by $i_{X_H}d\beta=-dH$.

We show that

$$X_H=-\frac{ax}{\varepsilon_{\mathrm{sym}}}\frac{\partial}{\partial x}+\frac{ay}{\varepsilon_{\mathrm{sym}}}\frac{\partial}{\partial y}$$

is a solution of the equation

$$\beta(X_H)=H \tag{3B.1}$$

on U_s. We calculate

$$i_{X_H}(d\beta)=\left(-\frac{ax}{\varepsilon_{\mathrm{sym}}}\frac{\partial}{\partial x}+\frac{ay}{\varepsilon_{\mathrm{sym}}}\frac{\partial}{\partial y}\right)\lrcorner(\varepsilon_{\mathrm{sym}}dx\wedge dy)=-ax\,dy-ay\,dx=-dH$$

and

$$\beta(X_H) = \tfrac{1}{2}\varepsilon_{\text{sym}}(x\,dy - y\,dx)\left(-\frac{ax}{\varepsilon_{\text{sym}}}\frac{\partial}{\partial x} + \frac{ay}{\varepsilon_{\text{sym}}}\frac{\partial}{\partial y}\right) = axy = H.$$

Finally, Lemmas 3.1 and 3.2 say that $\beta = \frac{1}{2}r^2 d\theta$ and $H = \mu r^2\cos(nk\theta)$ on $\mathbb{R}^2 \setminus D(r_{\text{sing}})$. As in the previous case, we show that

$$X_H = nk\mu r\sin(nk\theta)\frac{\partial}{\partial r} + 2\mu\cos(nk\theta)\frac{\partial}{\partial\theta}$$

is a solution of Equation (3B.1) on $\mathbb{R}^2 \setminus D(r_{\text{sing}})$.

We calculate

$$\begin{aligned} i_{X_H}(d\beta) &= (nk\mu r\sin(nk\theta)\partial_r + 2\mu\cos(nk\theta)\partial_\theta)\lrcorner(r\,dr \wedge d\theta) \\ &= -2\mu r\cos(nk\theta)dr + nk\mu r^2\sin(nk\theta)\,d\theta = -dH, \end{aligned}$$

and

$$\begin{aligned} \beta(X_H) &= \tfrac{1}{2}r^2 d\theta\left(nk\mu r\sin(nk\theta)\frac{\partial}{\partial r} + 2\mu\cos(nk\theta)\frac{\partial}{\partial\theta}\right) \\ &= \mu r^2\cos(nk\theta) = H. \end{aligned}$$

□

Let X_H be the Hamiltonian vector field of H with respect to $d\beta$ and $\varphi^s_{X_H}$ be the time-s flow of X_H. Now we introduce the following notations:

$$\begin{aligned} S &:= \left\{x \in \mathbb{R}^2 \setminus D(r_{\text{sing}}) \mid \varphi^s_{X_H}(x) \in \mathbb{R}^2 \setminus D(r_{\text{sing}}) \text{ for all } s \in [0,1]\right\}, \\ V &:= \left\{x \in U \mid \varphi^s_{X_H}(x) \in U \text{ for all } s \in [-1,1]\right\}, \\ V_i &:= \left\{x \in U_i \mid \varphi^s_{X_H}(x) \in U_i \text{ for all } s \in [-1,1]\right\}. \end{aligned}$$

For simplicity, let us denote $\varphi_{X_H} := \varphi^1_{X_H}$.

Remark 3.4. Using the form of X_H on U_i, where $i = 1, \ldots, nk$, we may assume that the curves γ_{p,p_i} in Lemma 3.2 satisfy the following list of properties:

(1) γ_{p,p_i} is an embedded curve which starts at p and ends at p_i;

(2) γ_{p,p_i} is a part of one of the curves of the singular foliation given by $\ker\beta$;

(3) γ_{p,p_i} coincides with one of the level sets of H on V_i and near p_i can be presented as $W^s(\varphi_{X_H}, p_i) = \{x \mid (\varphi_{X_H})^n(x) \to p \text{ as } n \to \infty\}$.

Recall that the following claim was proven in [Golovko 2011]:

Claim 3.5. *If (M, ω) is an exact symplectic manifold, i.e., $\omega = d\beta$, then the flow $\varphi^t_{X_H}$ of a Hamiltonian vector field X_H consists of exact symplectic maps, i.e.,*

$$(\varphi^t_{X_H})^*\beta - \beta = df_t,$$

where

$$f_t = \int_0^t (-H + \beta(X_H)) \circ \varphi^s_{X_H}\, ds.$$

Remark 3.6. From Lemma 3.3 and Claim 3.5 it follows that

$$\varphi^*_{X_H}(\beta) - \beta = dh,$$

where $h := f_1 = 0$ on $S \cup \bigcup_{i=1}^{nk} V_i$ and $h = \tilde{C} > 0$ on V. Hence, we get $\varphi^*_{X_H}(\beta) = \beta$ on $S \cup V \cup \bigcup_{i=1}^{nk} V_i$.

Now we define $\varphi_{-k/l} := R_{-k/l} \circ \varphi_{X_H}$, where $R_{-k/l} : \mathbb{R}^2 \to \mathbb{R}^2$ is a rotation by $-2\pi l/k$ around p.

Remark 3.7. Since $R^*_{nk}(\beta) = \beta$, we get $R^*_{-k/l}(\beta) = \beta$ and hence

$$\varphi^*_{-k/l}(\beta) = (R_{-k/l} \circ \varphi_{X_H})^*(\beta) = \varphi^*_{X_H}(R^*_{-k/l}(\beta)) = \varphi^*_{X_H}(\beta).$$

Fix $R_* \gg r_{\text{sing}}$ such that there is an annular neighborhood V_{R_*} of $\partial D(R_*)$ in $\mathbb{R}^2$ with $V_{R_*} \subset S$. Consider $D(R_*)$ with

$$\beta_0 := \beta|_{D(R_*)} \quad \text{and} \quad \beta_1 := \varphi^*_{X_H}(\beta)|_{D(R_*)} (= \varphi^*_{-k/l}(\beta)|_{D(R_*)}).$$

Note that

$$d\beta_1 = d(\varphi^*_{X_H}(\beta)|_{D(R_*)}) = \varphi^*_{X_H}(d\beta)|_{D(R_*)} = (d\beta)|_{D(R_*)} = d\beta_0 > 0. \tag{3B.2}$$

In addition, from the definitions of $V(R_*)$ and $D(R_*)$ it follows that

$$\beta_0 = \beta_1 \quad \text{on} \quad V_{R_*} \cap D(R_*). \tag{3B.3}$$

Now we recall the construction of the contact 1-form on $[-1, 1] \times D^2$.

Lemma 3.8 [Golovko 2011, Lemma 3.10]. *Let β_0 and β_1 be two 1-forms on D^2 such that $\beta_0 = \beta_1$ in a neighborhood of ∂D^2 and $d\beta_0 = d\beta_1 = \omega > 0$. Then there exists a contact 1-form α and a Reeb vector field R_α on $[-1, 1] \times D^2$ with coordinates (t, x), where t is a coordinate on $[-1, 1]$ and x is a coordinate on D^2, with the following properties*:

1. *$\alpha = dt + \varepsilon\beta_0$ in a neighborhood of $\{-1\} \times D^2$.*
2. *$\alpha = dt + \varepsilon\beta_1$ in a neighborhood of $\{1\} \times D^2$.*
3. *R_α is collinear to $\partial/\partial t$ on $[-1, 1] \times D^2$.*
4. *$R_\alpha = \partial/\partial t$ in a neighborhood of $[-1, 1] \times \partial D^2$.*

Here ε is a small positive number.

In addition, recall that

$$\alpha = (1 + \varepsilon\chi_1(t)h)\, dt + \varepsilon((1 - \chi_0(t))\beta_0 + \chi_0(t)\beta_1), \tag{3B.4}$$

where $h \in C^\infty(D^2)$ such that $\beta_1 - \beta_0 = dh$; $\chi_0 : [-1, 1] \to [0, 1]$ is a smooth map for which $\chi_0(t) = 0$ for $-1 \le t \le -1 + \varepsilon_{\chi_0}$, $\chi_0(t) = 1$ for $1 - \varepsilon_{\chi_0} \le t \le 1$, $\chi_0'(t) \ge 0$ for $t \in [-1, 1]$ and ε_{χ_0} is a small positive number; $\chi_1(t) := \chi_0'(t)$; ε is a sufficiently small positive number.

Remark 3.9. Note that $d\alpha = \varepsilon\omega$, where α is a 1-form given by (3B.4) and $\omega = d\beta_0 = d\beta_1 > 0$ on D^2.

Observe that from (3B.2) and (3B.3) it follows that β_0 and β_1 described above satisfy the conditions of Lemma 3.8. We now take $[-1, 1] \times D(R_*)$ equipped with the contact 1-form α given by (3B.4). For simplicity, let us denote $\beta_- := \varepsilon\beta_0$ and $\beta_+ := \varepsilon\beta_1$, where ε is a constant from Lemma 3.8 which makes α contact.

3C. *Gluing.* We now construct P_+, P_- and D in the way described in [Golovko 2011]. Recall that

$$P_+,\ P_-,\ D \subset D(R_*) \subset \mathbb{R}^2$$

are surfaces with boundary which satisfy the following properties:

(1) $P_\pm \subset D$.

(2) $(\partial P_\pm)_\partial \subset \partial D$ and $(\partial P_\pm)_{\mathrm{int}} \subset \mathrm{int}(D)$.

(3) φ_{X_H} maps P_+ to P_- taking $(\partial P_+)_{\mathrm{int}}$ onto $(\partial P_-)_\partial$ and $(\partial P_+)_\partial$ onto $(\partial P_-)_{\mathrm{int}}$.

(4) $(\partial P_-)_\partial \cap (\partial P_+)_\partial = \varnothing$.

Note that

- $\partial P_+ = \left(\bigcup_{s=0}^{nk-1} a_s^+\right) \cup \left(\bigcup_{s=0}^{nk-1} b_s^+\right)$,
- $\partial P_- = \left(\bigcup_{s=0}^{nk-1} a_s^-\right) \cup \left(\bigcup_{s=0}^{nk-1} b_s^-\right)$,
- $\partial D = \left(\bigcup_{s=0}^{nk-1} a_s^+\right) \cup \left(\bigcup_{s=0}^{nk-1} b_s^-\right) \cup \left(\bigcup_{s=0}^{nk-1} c_s^+\right) \cup \left(\bigcup_{s=0}^{nk-1} c_s^-\right)$.

See Figure 3 for the schematic visualization of P_+ (bounded by the bold line), P_- and D. For more details of this construction we refer to [Golovko 2011].

Remark 3.10. Note that the $a_i^\pm$, $b_i^\pm$ and $c_i^\pm$ are constructed in such a way that

$$a_i^\pm,\ b_i^\pm,\ c_i^\pm \subset D(R_*) \cap S$$

for $i = 0, \dots, nk - 1$. Hence, we see that $\partial P_+, \partial P_-, \partial D \subset D(R_*) \cap S$. In addition, $R_{nk}(a_i^\pm) = a_{i+1}^\pm$ and $R_{nk}(b_i^\pm) = b_{i+1}^\pm$, where $i, i+1$ are considered modulo nk.

We take $[-1, 1] \times D$ with a contact form $\alpha := \alpha|_{[-1,1]\times D}$. Let $\Gamma = \{0\} \times \partial D$ in $[-1, 1] \times D$ and $U(\Gamma) := [0, 1] \times [-1, 1] \times \Gamma$ be a neighborhood of Γ with coordinates $(\tau, t) \in [0, 1] \times [-1, 1]$, where t is a usual t-coordinate on $[-1, 1] \times D$. From the definition of S and Remark 3.10 it follows that we may assume that $U(\Gamma) \subset [-1, 1] \times (S \cap D)$.

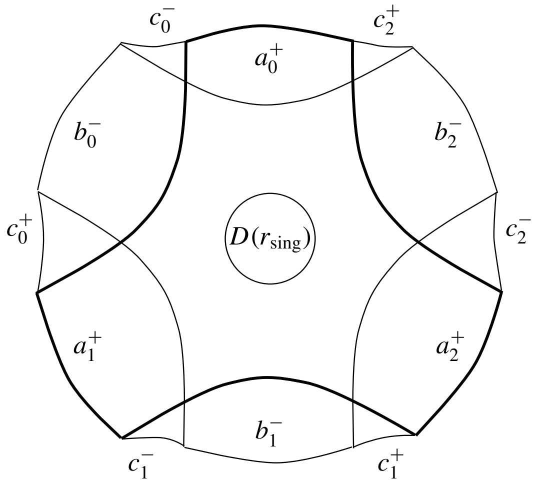

Figure 3. Construction of P_+, P_- and D in the case $n = 1$, $k = 3$.

Lemma 3.11. $([-1, 1] \times D, \Gamma, U(\Gamma), \xi)$ *is a sutured contact manifold and α is an adapted contact form.*

Proof. First note that $\alpha|_{R-} = \beta_-$ and $\alpha|_{R_+} = \beta_+$. Let us check that (R_-, β_-) and (R_+, β_+) are Liouville manifolds. From the construction of $\beta_\pm$ it follows that $d(\beta_-) = d(\beta_+) > 0$. Since $\beta_- = \beta_+$ on $D \cap S$ and by (3B.4), $\alpha = dt + \beta_-$ on $U(\Gamma)$. Recall that $\beta_- = \beta_+ = \frac{1}{2}\varepsilon r^2 d\theta$ on $D \cap S$. Hence, $\alpha|_{U(\Gamma)} = dt + \frac{1}{2}\varepsilon r^2 d\theta$. The calculation

$$i_{Y_\pm|_{R_\pm \cap U(\Gamma)}}(d\beta_\pm) = \left(\tfrac{1}{2} r\partial_r\right)\lrcorner(\varepsilon r\, dr \wedge d\theta) = \tfrac{1}{2}\varepsilon r^2 d\theta = \beta_\pm$$

implies that the Liouville vector fields $Y_\pm|_{R_\pm \cap U(\Gamma)}$ are equal to $\frac{1}{2} r\partial_r$. From the construction of D it follows that $Y_\pm$ is positively transverse to $\partial R_\pm$. Thus, $(R_-, \varepsilon\beta_0)$ and $(R_+, \varepsilon\beta_1)$ are Liouville manifolds. As already mentioned, we have $\alpha = dt + \beta_-$ on $U(\Gamma)$. Finally, if we take τ such that $\partial_\tau = \frac{1}{2} r\partial_r$, then $([-1, 1] \times D, \Gamma, U(\Gamma), \xi)$ becomes a sutured contact manifold with an adapted contact form α. □

Then we use $\varphi_{-k/l}$ for the gluing construction. Note that φ_{X_H} maps a_s^+ to a_s^- and b_s^+ to b_s^-. Hence, using Remark 3.10, we see that $\varphi_{-k/l}$ maps a_s^+ to a_{s-nl}^- and b_s^+ to b_{s-nl}^-. Then we follow the gluing procedure briefly described in Section 2A and completely written in [Colin et al. 2011]. Finally, we get a sutured contact solid torus $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}))$ with a contact form $\tilde{\alpha}_\delta$, where $\tilde{\Gamma}$ is a set of $2n$ parallel closed curves of slope $-k/l$, where $n \in \mathbb{N}$, $(k, l) = 1$, $k > l > 0$ and δ is the rotation angle of the map φ_{X_H} near p.

Remark 3.12. We have constructed $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}))$ using the gluing construction for sutured manifolds. However, since there is a close connection between sutured contact manifolds and contact manifolds with convex boundary, we observe that the gluing construction we used for the sutured contact solid cylinder corresponds to the gluing construction for the contact 3-ball with convex boundary and one dividing curve on the boundary. The corresponding gluing construction for the contact 3-ball with convex boundary corresponds (is inverse) to the convex decomposition of the contact solid torus $S^1 \times D^2$ with convex boundary with respect to the convex meridional disk $\{pt\} \times D^2$ with ∂-parallel dividing curves. Hence, the constructed sutured contact solid tori are universally tight sutured contact manifolds by the gluing/classification result from Section 2 in [Honda 2002] (more precisely, Corollary 2.3, Theorem 2.5 and Corollary 2.6).

3D. ***Reeb orbits.*** Note that $\varphi_{-k/l}|_{P_+}$ has n orbits of period k obtained from the equally spaced saddle points of H. Lemma 3.8 and the gluing procedure briefly described in Section 2A imply that these orbits correspond to the Reeb orbits, which we call $\gamma_1, \dots, \gamma_n$ such that

$$[\gamma_s] = [\gamma_t] = k \in H_1(S^1 \times D^2; \mathbb{Z})$$

for $s, t = 1, \dots, n$. In addition, $\varphi_{-k/l}|_{P_+}$ has a periodic point of period 1, which is p. It corresponds to the Reeb orbit, which we call γ, such that $[\gamma] = 1 \in H_1(S^1 \times D^2; \mathbb{Z})$.

Lemma 3.13. *$\int_{\gamma_s} \tilde{\alpha}_\delta = \int_{\gamma_t} \tilde{\alpha}_\delta$ and $k \int_{\gamma} \tilde{\alpha}_\delta > \int_{\gamma_s} \tilde{\alpha}_\delta$, where $s, t = 1, \dots, n$.*

Proof. Let

$$M^{(0)} = (([-1, 1] \times D) \cup (R_+(\Gamma) \times [1; \infty)) \cup (R_+(\Gamma) \times (-\infty; -1])),$$
$$\tilde{M} = M^{(0)} \setminus ((P_+ \times (N, \infty) \cup (P_- \times (-\infty, -N)).$$

In addition, let $\alpha_{\tilde{M}}$ denote the contact form on $\tilde{M}$ and let $\xi_{\tilde{M}}$ denote the contact structure defined by $\alpha_{\tilde{M}}$.

Consider $[-1, 1] \times D \subset \tilde{M}$. From the construction of α it follows that $\beta_+ = \beta_-$ on V_s and $\alpha|_{[-1,1]\times V_s} = dt + \beta_-$ for $s = 1, \dots, nk$. Hence, since the contact structure on $[1, \infty) \times P_+$ is given by $dt + \beta_+$ and the contact structure on $(-\infty, -1] \times P_-$ is given by $dt + \beta_-$, $\alpha_{\tilde{M}}|_{[-N,N]\times V_s} = dt + \beta_-$ on $[-N, N] \times V_s \subset \tilde{M}$ for $s = 1, \dots, nk$. Therefore, we get

$$\int_{[-N,N]\times\{p_s\}} \alpha_{\tilde{M}} = 2N \tag{3D.1}$$

for $s = 1, \dots, nk$. From the gluing construction and (3D.1) it follows that

$$\int_{\gamma_s} \tilde{\alpha}_\delta = 2Nk$$

for $s = 1, \dots, n$. Note that $\int_{\gamma_s} \tilde{\alpha}_\delta$ does not depend on s. Hence, $\int_{\gamma_s} \tilde{\alpha}_\delta = \int_{\gamma_t} \tilde{\alpha}_\delta$ for $s, t = 1, \dots, n$.

Now from the fact that $\alpha = (1 + \varepsilon\chi_1(t)h)\,dt + \beta_-$ on $[-1, 1] \times V$, where $h > 0$ and $\chi_1(t) > 0$, we get that

$$R_\alpha = \frac{1}{1 + \varepsilon\chi_1(t)h}\frac{\partial}{\partial t}$$

on $[-1, 1] \times V$. Hence, from the gluing construction we obtain $k\int_\gamma \tilde{\alpha}_\delta > 2Nk$. Thus,

$$\int_{\gamma_s} \tilde{\alpha}_\delta = \int_{\gamma_t} \tilde{\alpha}_\delta \quad \text{and} \quad k\int_\gamma \tilde{\alpha}_\delta > \int_{\gamma_s} \tilde{\alpha}_\delta,$$

where $s, t = 1, \ldots, n$. $\square$

Lemma 3.14. *All closed orbits of $R_{\tilde{\alpha}_\delta}$ are nondegenerate. Moreover, γ is an elliptic orbit and γ_i is a hyperbolic orbit such that γ^t and γ_i^s are good orbits for $i = 1, \ldots n$; $s, t \in \mathbb{N}$. There exists a symplectic trivialization τ of ξ along γ and the γ_i, constructed in the consistent way as described in Section 2B, and $N_\delta \in \mathbb{N}$ such that*

$$\mu_\tau(\gamma_i^s) = -2ls,$$
$$\mu_\tau(\gamma^t) = -2ml + 1,$$

where $(m-1)k < t \le mk$ and $i = 1, \ldots, n$, $t \le N_\delta$, $s \le N_\delta/k$.

Proof. For simplicity, assume that $l = 1$. The general calculation can be done in the analogous way.

Fix $i = 1, \ldots, n$. We first observe that $H|_{V_i} = axy$, where $a > 0$ and hence

$$\varphi_{X_H}|_{V_i} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

where $\lambda = e^a \ne 1$. Let the symplectic trivialization of $\xi_{\tilde{M}}$ along $[-N, N] \times \{p_i\}$ be given by the framing $(\lambda^{\frac{-N-t}{2N}}\partial_x, \lambda^{\frac{t+N}{2N}}\partial_y)$, where $i = 1, \ldots, nk$ and (x, y) are coordinates on V_i which coincide with the coordinates on U_i from Lemma 3.1. Since Lemma 3.1 implies that R_{nk} maps coordinates on V_i to the coordinate on V_{i+1}, where $i, i+1$ are considered modulo nk, we conclude that the symplectic trivializations of $\xi_{\tilde{M}}$ along each $[-N, N] \times \{p_{i+nm}\}$ for $m = 0, \ldots, k-1$ and fixed $i = 1, \ldots, n$ give rise to the symplectic trivialization τ_{γ_i} of $\tilde{\xi}$ along γ_i. It is easy to see that the linearized return map P_{γ_i} with respect to this trivialization is given by

$$P_{\gamma_i} = \begin{pmatrix} \lambda^k & 0 \\ 0 & \lambda^{-k} \end{pmatrix}.$$

Since the eigenvalues of P_{γ_i} are positive real numbers different from 1, γ_i is a positive hyperbolic orbit. In addition, $P_{\gamma_i^s} = P_{\gamma_i}^s$. Therefore, the eigenvalues of $P_{\gamma_i^s}$ are different from 1. Hence, γ_i^s is a nondegenerate orbit for $s \in \mathbb{N}$ and $i = 1, \ldots, n$.

We now observe that the linearized Reeb flow around γ_i (with respect to τ_{γ_i}) rotates the eigenspaces of P_{γ_i} by angle -2π. Hence, we get

$$\mu_{\tau_{\gamma_i}}(\gamma_i^s) = -2s \tag{3D.2}$$

for $s \in \mathbb{N}$ and $i = 1, \dots, n$.

Now let the symplectic trivialization of $\xi_{\tilde{M}}$ along $[-N, N] \times \{p\}$ be given by the framing

$$\big(\cos(\theta_{\delta,k,N}(t))\partial_x + \sin(\theta_{\delta,k,N}(t))\partial_y, -\sin(\theta_{\delta,k,N}(t))\partial_x + \cos(\theta_{\delta,k,N}(t))\partial_y\big),$$

where $\theta_{\delta,k,N}(t) = \pi(1-\delta k)(t+N)/(Nk)$ and $t \in [-N, N]$. Note that $R_{-k} \circ \varphi_{X_H}|_V$ is a rotation through $2\pi(-1/k+\delta)$, where R_{-k} is a $-2\pi/k$-rotation about p and δ is a small positive irrational number. It is easy to see that with respect to this framing P_γ is a rotation by $2\pi(-1/k+\delta)$. Hence, since δ is irrational, we see that γ is an elliptic orbit and γ^t is nondegenerate for $t \in \mathbb{N}$. Let

$$N_\delta := \max\big\{m \in \mathbb{N} \mid m\delta < 1/k\big\}.$$

Note that we get

$$\mu_{\tau_\gamma}(\gamma^t) = -2m + 1, \tag{3D.3}$$

where $(m-1)k < t \le mk$ and $t \le N_\delta$. Formulas (3D.2) and (3D.3) and the fact that δ is irrational imply that the parity of $\mu_{\tau_{\gamma_i}}(\gamma_i^s)$ is independent of s for given i and the parity of $\mu_{\tau_\gamma}(\gamma^t)$ is independent of t. Hence, we conclude that the γ_i^s and γ^t are good Reeb orbits for $i = 1, \dots, n$ and $s, t \in \mathbb{N}$.

It is not difficult to see that the symplectic trivialization τ_{γ^k} (induced from τ_γ) can be extended to the τ_{γ_i} (are consistent in terms of Section 2B) along the surfaces obtained from $(\varphi_{X_H}^{(-N-t)/2N}(\gamma_{p,p_i}))_{i=1}^{nk}$ by gluing them with φ_{-k} and gives rise to the global symplectic trivialization that we call τ. □

4. Calculation

In this section, we calculate the sutured version of cylindrical contact homology of the sutured contact solid torus that we have constructed in Section 3.

Remark 4.1. There are no contractible Reeb orbits. Hence, from Theorem 2.7, Remark 2.11, and the fact that $\pi_1(S^1 \times D^2; \mathbb{Z}) \simeq H_1(S^1 \times D^2; \mathbb{Z}) \simeq \mathbb{Z}$ it follows that for all $h \in H_1(S^1 \times D^2; \mathbb{Z})$, $HC_*^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta, J)$ is defined, i.e., $\partial^2 = 0$, and is independent of contact form $\tilde{\alpha}_\delta$ for the given contact structure $\tilde{\xi}$ and the almost complex structure J.

For simplicity, assume that $l = 1$. The calculation for $l > 1$ can be made in the completely analogous way.

Lemma 3.14 implies that all Reeb orbits are good and

$$|\gamma_i^s| = -2s - 1, \quad |\gamma^t| = -2m, \tag{4A.4}$$

where $m - 1 < t/k \le m$ and $i = 1, \dots, n,\ s \le N_\delta/k,\ t \le N_\delta$. Hence, we get

$$C_m^h(\tilde{\alpha}_\delta, J) = \begin{cases} \mathbb{Q}\langle\gamma^h\rangle & \text{for } h > 0 \text{ and } m = 2\lfloor h(-1/k+\delta)\rfloor, \\ \mathbb{Q}\langle\gamma_1^{h/k}, \dots, \gamma_n^{h/k}\rangle & \text{for } k \mid h > 0 \text{ and } m = -2h/k - 1, \\ 0, & \text{otherwise,} \end{cases} \tag{4A.5}$$

for $h \le N_\delta$.

Now, since by Lemma 3.13 $\mathcal{A}(\gamma^k) > \mathcal{A}(\gamma_i)$ for $i = 1, \dots, n$, we can use Fact 2.6 and Remark 2.11 and conclude that $\partial(\gamma_i^s) = 0$ for $i = 1, \dots, n$ and $s > 0$. Then, we prove that $\partial(\gamma^t) = 0$ for $k \nmid t \le N_\delta$. Since $[\gamma_i] = k[\gamma]$ in $H_1(S^1 \times D^2; \mathbb{Z}) \cong \mathbb{Z}$, the cylindrical contact homology differential at γ^t counts only cylinders with negative end at γ^t. Then, similarly to the previous case, Fact 2.6 and Remark 2.11 imply that $\partial(\gamma^t) = 0$ for $k \nmid t \le N_\delta$.

We now consider the case when $k \mid t$ and will show that $\partial(\gamma^t) \ne 0$ for $k \mid t \le N_\delta$. Is this situation, by arguing in the same way as in the case when $k \nmid t$, we get that $\partial(\gamma^t)$ counts only cylinders with negative end at $\gamma_i^{t/k}$.

Now we note that

$$\operatorname{ind}(u) = |\gamma^t| - |\gamma_i^{t/k}| \tag{4A.6}$$

for any pseudoholomorphic curve u in the moduli space $\mathcal{M}^J(\gamma^t; \gamma_i^{t/k})$, where $k \mid t \le N_\delta$ and J is an almost complex structure tailored to $((\mathbb{R} \times S^1 \times D^2)^*, \tilde{\alpha}_\delta^*)$. The index formula can be written in this way, since $H_2(S^1 \times D^2; \mathbb{Z}) = 0$ and hence $< c_1(\xi), A >= 0$ for all $A \in H_2(S^1 \times D^2, \mathbb{Z})$. We now use (4A.4) and get

$$|\gamma^t| - |\gamma_i^{t/k}| = -2m - (-2t/k - 1) = -2(m - t/k) + 1,$$

and $m = t/k$ for $i = 1, \dots, n$; $t \le N_\delta$. Hence, we can rewrite (4A.6) as

$$\operatorname{ind}(u) = |\gamma^t| - |\gamma_i^{t/k}| = -2(t/k - t/k) + 1 = 1 \tag{4A.7}$$

for $i = 1, \dots, n$ and $t \le N_\delta$. Therefore, Theorem 2.8 and Remark 2.11 imply that for every $u \in \mathcal{M}(\gamma^t, \gamma_i^{t/k})$ the linearization of the Cauchy–Riemann operator is surjective at u; here $k \mid t \le N_\delta$, J is any almost complex structure tailored to $((S^1 \times D^2)^*, \tilde{\alpha}_\delta^*)$ and $i = 1, \dots, n$.

Let $(S^1 \times D^2, \Gamma_{\text{long}}, U(\Gamma_{\text{long}}), \alpha_\delta^{\text{long}})$ be a sutured contact solid torus obtained from $([-1,1] \times D, \Gamma, U(\Gamma), \alpha)$ by using φ_{X_H} as a gluing map. Recall that we get $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}), \tilde{\alpha}_\delta)$ from $([-1,1] \times D, \Gamma, U(\Gamma), \alpha)$ by using $\varphi_{-k} = R_{-k} \circ \varphi_{X_H}$ as a gluing map. We now note that $(S^1 \times D^2, \Gamma_{\text{long}}, U(\Gamma_{\text{long}}), \alpha_\delta^{\text{long}})$ is a universally tight sutured contact solid torus with $2nk$ parallel longitudinal sutures, $k > 1$, and such that when one cuts it along the meridian disk the sutures on the disk are boundary-parallel. This follows from the gluing/classification result for universally tight contact structures on a sutured solid torus; see Section 2 in [Honda 2002] (more precisely, Corollary 2.3, Theorem 2.5 and Corollary 2.6). The cylindrical contact homology of this sutured contact manifold is computed in [Golovko 2011] and is given by

$$HC^{\mathrm{cyl},h}(S^1 \times D^2, \Gamma_{\mathrm{long}}, \xi_{\mathrm{long}}) \simeq \begin{cases} \mathbb{Q}^{nk-1} & \text{for } h \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4A.8}$$

Here $\xi_{\mathrm{long}} = \ker \alpha_{\mathrm{long}}$.

Note that $(S^1 \times D^2, \Gamma_{\mathrm{long}}, U(\Gamma_{\mathrm{long}}), \alpha_\delta^{\mathrm{long}})$ has nk hyperbolic orbits

$$\gamma_1^{\mathrm{long}}, \dots, \gamma_{nk}^{\mathrm{long}}$$

and one elliptic orbit γ^{long}. Here the γ_i^{long} correspond to the equally spaced saddle points of H and γ^{long} corresponds to the critical point of H at the center of $D(r_{\mathrm{sing}})$. In addition, observe that

$$[\gamma_i^{\mathrm{long}}] = [\gamma^{\mathrm{long}}] = 1 \in H_1(S^1 \times D^2; \mathbb{Z}). \tag{4A.9}$$

Finally, note that from Lemma 3.13 and from the construction of

$$\gamma^{\mathrm{long}} \quad \text{and} \quad \gamma_1^{\mathrm{long}}, \dots, \gamma_{nk}^{\mathrm{long}},$$

it follows that

$$\mathcal{A}(\gamma^{\mathrm{long}}) > \mathcal{A}(\gamma_i^{\mathrm{long}}), \quad \mathcal{A}(\gamma_i^{\mathrm{long}}) = \mathcal{A}(\gamma_j^{\mathrm{long}}) \tag{4A.10}$$

for $i, j = 1, \dots, nk$. Hence, Theorem 2.7, Remark 2.11 together with Fact 2.6, and (4A.8), (4A.9) and (4A.10) imply that $\partial(\gamma^{\mathrm{long}})^s \neq 0$ for $s > 0$; otherwise we arrive at a contradiction with (4A.8) (since $\partial(\gamma^{\mathrm{long}})^s = 0$ implies that the exponent of $\mathbb{Q}$ in (4A.8) must be $nk+1$). In addition, observe that $< \partial(\gamma^{\mathrm{long}})^s, (\gamma_i^{\mathrm{long}})^s > \neq 0$ for some i and all $s > 0$.

We now take an almost complex structure J^{long} tailored to $((S^1 \times D^2)^*, (\alpha_\delta^{\mathrm{long}})^*)$ such that as a map $\xi^{\mathrm{long}} \to \xi^{\mathrm{long}}$ it is obtained from some fixed $J^{\mathrm{cyl}} : \xi \to \xi$ which is defined on $([-1, 1] \times D, \Gamma, U(\Gamma), \alpha)$ and satisfies the following properties:

(1) $(J^{\mathrm{cyl}})^2 = -I$, $d\alpha(J^{\mathrm{cyl}}\cdot\,, J^{\mathrm{cyl}}\cdot\,) = d\alpha(\cdot\,,\cdot)$, $d\alpha(\cdot\,, J^{\mathrm{cyl}}\cdot\,) > 0$;

(2) $J^{\mathrm{cyl}}|_{\{1\}\times D} = \varphi_{X_H}^*(J^{\mathrm{cyl}}|_{\{-1\}\times D})$ and J^{cyl} is $\frac{2\pi}{nk}$-symmetric, i.e., it is invariant under $\frac{2\pi}{nk}$-rotation with respect to the center of D.

Here $\xi^{\mathrm{long}} = \ker \alpha_\delta^{\mathrm{long}}$ and $\xi = \ker \alpha$. By saying that J^{long} is obtained from J^{cyl} we simply mean that the gluing procedure with φ_{X_H} applied to $([-1,1]\times D, \Gamma, U(\Gamma), \alpha)$ transforms J^{cyl} to J^{long}. Since ξ is $\frac{2\pi}{nk}$-symmetric on $([-1, 1]\times D, \Gamma, U(\Gamma), \alpha)$, we claim that J^{cyl}, which satisfies Properties (1) and (2), exists and that Property (2) is not a serious restriction on J^{cyl}. The symmetry of ξ follows from the symmetry of β and X_H, and from the construction of α. From the symmetry of J^{long} it follows that $< \partial(\gamma^{\text{long}})^s, (\gamma_i^{\text{long}})^s > \neq 0$ for all $i = 1, \dots, nk$ and $s > 0$.

Now we take $\tilde{J}$ on $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}), \tilde{\alpha}_\delta)$, which is obtained from the same J^{cyl} defined on $([-1,1] \times D, \Gamma, U(\Gamma), \alpha)$ by applying the gluing procedure with

$$\varphi_{-k} = R_{-k} \circ \varphi_{X_H} \quad \text{to } ([-1,1] \times D, \Gamma, U(\Gamma), \alpha),$$

and possibly modify it near the boundary of $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}), \tilde{\alpha}_\delta)$ (far from the Reeb orbits) so that it becomes tailored to $((S^1 \times D^2)^*, (\tilde{\alpha}_\delta)^*)$. Observe that we can assume that $J^{\text{long}}{=}\tilde{J}$. From the symmetry of J^{cyl} and the form of the gluing maps for $(S^1 \times D^2, \tilde{\Gamma}, U(\tilde{\Gamma}), \tilde{\alpha}_\delta)$ and $(S^1 \times D^2, \Gamma_{\text{long}}, U(\Gamma_{\text{long}}), \alpha_\delta^{\text{long}})$ it follows that every J^{long}-holomorphic curve u which contributes to $< \partial(\gamma^{\text{long}})^{ks}, (\gamma_i^{\text{long}})^{ks} > \neq 0$ can be modified to a $\tilde{J}$-holomorphic curve $\tilde{u}$ from γ^{ks} to γ_i^s by modifying (composing) it with the rotation about the center of a meridian disk, and hence $< \partial\gamma^{ks}, \gamma_i^s > \neq 0$.

This choice of almost complex structures is possible since Theorem 2.8 and Remark 2.11 imply that we do not need to require almost complex structures to be generic. Finally, it follows from (4A.5) that

$$HC_m^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta) \simeq \begin{cases} \mathbb{Q} & \text{for } h > 0 \text{ and } m = 2\lfloor h(-1/k + \delta) \rfloor, \\ \mathbb{Q}^{n-1} & \text{for } k \mid h > 0 \text{ and } m = -2h/k - 1, \\ 0, & \text{otherwise.} \end{cases}$$

for $h \le N_\delta$.

We now note that $\tilde{\xi} = \ker \tilde{\alpha}_\delta$ is independent of δ. This follows from the gluing/classification result for universally tight contact structures on a sutured solid torus; see [Honda 2002, Corollary 2.3, Theorem 2.5 and Corollary 2.6]. Hence, from Theorem 2.7 and Remark 2.11 it follows that

$$HC^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\xi}) = HC^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta)$$

for all h and hence for $h \le N_\delta$, where δ is a small positive irrational number,

$$HC^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\xi}) := \bigoplus_m HC_m^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\xi}),$$

$$HC^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta) := \bigoplus_m HC_m^{\text{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta).$$

Now observe that $N_\delta \to \infty$ when $\delta \to 0$. In addition, we note that for fixed n, k and two small positive irrational numbers $\delta_1 \neq \delta_2$, the sets of closed orbits of $R_{\tilde{\alpha}_{\delta_1}}$ and $R_{\tilde{\alpha}_{\delta_2}}$ are the same, and the corresponding orbits with the same first homology class $h \le min\{N_{\delta_1}, N_{\delta_2}\}$ have the same Conley–Zehnder gradings in the corresponding complexes. Therefore, for every $0 < h \in \mathbb{Z} = H_1(S^1 \times D^2; \mathbb{Z})$, there exists δ such that

$$HC_m^{\mathrm{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\xi}) = HC_m^{\mathrm{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\alpha}_\delta)$$
$$\simeq \begin{cases} \mathbb{Q} & \text{for } h > 0 \text{ and } m = 2\lfloor h(-1/k+\delta)\rfloor, \\ \mathbb{Q}^{n-1} & \text{for } k \mid h > 0 \text{ and } m = -2h/k - 1, \\ 0, & \text{otherwise.} \end{cases}$$

for $h \leq N_\delta$ and hence

$$\text{(4A.11)}\quad HC_m^{\mathrm{cyl},h}(S^1 \times D^2, \tilde{\Gamma}, \tilde{\xi}) \simeq \begin{cases} \mathbb{Q} & \text{for } h > 0 \text{ and } m = 2\lfloor -h/k + \delta_k\rfloor, \\ \mathbb{Q}^{n-1} & \text{for } k \mid h > 0 \text{ and } m = -2h/k - 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \delta_k \ll 1/k$. Finally, (4A.11) implies that

$$\text{(4A.12)}\qquad HC^{\mathrm{cyl},h}(S^1 \times D^2, \Gamma, \xi) \simeq \begin{cases} \mathbb{Q} & \text{for } k \nmid h > 0, \\ \mathbb{Q}^{n-1} & \text{for } k \mid h > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This completes the proof of Theorem 1.1 when $l = 1$.

For $l > 1$, one can use the same observations as in the case when $l = 1$ and show that the only nonzero part of the cylindrical contact homology differential is given by $< \partial\gamma^t, \gamma_i^{t/k} > \neq 0$ for $k \mid t \leq N_\delta$. This will lead to (4A.12) for all l such that $(k, l) = 1$, $k > l > 0$.

Remark 4.2. Theorem 1.3 from [Golovko 2011] and Theorem 1.1 provide the formula for the sutured version of cylindrical contact homology of $(S^1 \times D^2, \Gamma, \xi)$, where Γ consists of $2n$ parallel sutures of arbitrary slope, ξ is a universally tight contact structure and such that if one cuts along the meridian disk, the sutures on the disk are ∂-parallel. In particular, this gives a complete calculation of the cylindrical contact homology of $(S^1 \times D^2, \Gamma, \xi)$, where Γ consists of 2 parallel sutures of arbitrary slope and ξ is a universally tight contact structure (observe that in this situation there are only two isomorphic (but not isotopic) universally tight contact structures; see Section 2 in [Honda 2002]). These are not all the universally tight contact structures on the solid torus, but all of them can be obtained from the $\#\Gamma = 2$ case by successively applying the folding operation.

Acknowledgements

The author is deeply grateful to Ko Honda for his guidance, help and support. He also thanks Dmytro Chebotarov, Oliver Fabert, Paolo Ghiggini, Jian He, Michael Hutchings and Mark McLean for helpful suggestions and interest in his work. In addition, the author is extremely grateful to Andrew Cotton-Clay for his critical comments on the first version of the paper. Also, the author is grateful to the referee of an earlier version of this paper for many valuable comments and suggestions. Finally, the author thanks the Mathematical Sciences Research Institute and the organizers of the "Symplectic and Contact Geometry and Topology" program for their hospitality.

References

[Bourgeois 2009] F. Bourgeois, "A survey of contact homology", pp. 45–72 in *New perspectives and challenges in symplectic field theory* (Stanford, CA, 2007), edited by M. Abreu et al., CRM Proc. Lecture Notes **49**, Amer. Math. Soc., Providence, RI, 2009. MR 2011a:53175 Zbl 1189.53082

[Bourgeois et al. 2003] F. Bourgeois, Y. Eliashberg, H. Hofer, K. Wysocki, and E. Zehnder, "Compactness results in symplectic field theory", *Geom. Topol.* **7** (2003), 799–888. MR 2004m:53152 Zbl 1131.53312

[Colin and Honda 2005] V. Colin and K. Honda, "Constructions contrôlées de champs de Reeb et applications", *Geom. Topol.* **9** (2005), 2193–2226. MR 2006m:53134 Zbl 1091.53057

[Colin et al. 2011] V. Colin, P. Ghiggini, K. Honda, and M. Hutchings, "Sutures and contact homology I", *Geom. Topol.* **15**:3 (2011), 1749–1842. MR 2851076 Zbl 1231.57026

[Cotton-Clay 2009] A. Cotton-Clay, "Symplectic Floer homology of area-preserving surface diffeomorphisms", *Geom. Topol.* **13**:5 (2009), 2619–2674. MR 2011a:53173 Zbl 1179.37077

[Eliashberg et al. 2000] Y. Eliashberg, A. Givental, and H. Hofer, "Introduction to symplectic field theory", pp. 560–673 in *Visions in mathematics: GAFA 2000 special volume, II* (Tel Aviv, 1999), edited by N. Alon et al., Birkhäuser, Basel, 2000. MR 2002e:53136 Zbl 0989.81114

[Gabai 1983] D. Gabai, "Foliations and the topology of 3-manifolds", *J. Differential Geom.* **18**:3 (1983), 445–503. MR 86a:57009 Zbl 0533.57013

[Golovko 2011] R. Golovko, "The embedded contact homology of sutured solid tori", *Algebr. Geom. Topol.* **11**:2 (2011), 1001–1031. MR 2012g:53186 Zbl 1233.57019

[Honda 2002] K. Honda, "Gluing tight contact structures", *Duke Math. J.* **115**:3 (2002), 435–478. MR 2003i:53125 Zbl 1026.53049

[Hutchings 2002] M. Hutchings, "An index inequality for embedded pseudoholomorphic curves in symplectizations", *J. Eur. Math. Soc.* **4**:4 (2002), 313–361. MR 2004b:53148 Zbl 1017.58005

[Momin 2011] A. Momin, "Contact homology of orbit complements and implied existence", *J. Mod. Dyn.* **5**:3 (2011), 409–472. Zbl 1248.37057

[Wendl 2010] C. Wendl, "Automatic transversality and orbifolds of punctured holomorphic curves in dimension four", *Comment. Math. Helv.* **85**:2 (2010), 347–407. MR 2011g:32037 Zbl 1207.32021

Received January 21, 2014. Revised June 15, 2014.

ROMAN GOLOVKO
DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ PARIS-SUD
BÂTIMENT 425
91405 ORSAY
FRANCE
roman.golovko@math.u-psud.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.97

UNIFORM BOUNDEDNESS OF S-UNITS IN ARITHMETIC DYNAMICS

HOLLY KRIEGER, AARON LEVIN, ZACHARY SCHERR, THOMAS TUCKER, YU YASUFUKU AND MICHAEL E. ZIEVE

Let K be a number field and let S be a finite set of places of K which contains all the archimedean places. For any $\phi(z) \in K(z)$ of degree $d \geq 2$ which is not a d-th power in $\bar{K}(z)$, Siegel's theorem implies that the image set $\phi(K)$ contains only finitely many S-units. We conjecture that the number of such S-units is bounded by a function of $|S|$ and d (independently of K, S and ϕ). We prove this conjecture for several classes of rational functions, and show that the full conjecture follows from the Bombieri–Lang conjecture.

1. Introduction

Let K be a number field, let S be a finite set of places of K which contains the set S_∞ of archimedean places of K, and write $\mathfrak{o}_S$ for the ring of S-integers of K and $\mathfrak{o}_S^*$ for the group of S-units of K. The genus-0 case of Siegel's theorem asserts that, for any $\phi(z) \in K(z)$ which has at least three poles in $\mathbb{P}^1(\bar{K})$, the image set $\phi(K)$ contains only finitely many S-integers. However, the number of S-integers in $\phi(K)$ cannot be bounded independently of $\phi(z)$, even if we restrict to functions $\phi(z)$ having a fixed degree, since $\psi(z) := \beta\phi(z)$ satisfies $\psi(K) = \beta\phi(K)$ for any $\beta \in K^*$.

Although the number of S-integers in $\phi(K)$ cannot be bounded in terms of only K, S, and $\deg \phi$, such a bound may be possible for the number of S-units in $\phi(K)$. In fact we conjecture that there is a bound depending only on $|S|$ and $\deg \phi$ (and not on K):

Conjecture 1.1. *For any integers $s \geq 1$ and $d \geq 2$, there is a constant $C = C(s, d)$ such that for any*

- *number field K,*

The authors were partially supported by NSF grants DMS-1303770 (H.K.), DMS-1102563 (A.L.), DMS-1200749 (T.T.), and DMS-1162181 (M.Z.). The fifth author was partially supported by JSPS Grants-in-Aid 23740033.
MSC2010: primary 37P05, 37P15; secondary 11G99, 11R99.
Keywords: arithmetic dynamics, S-units, uniform boundedness.

- *s-element set S of places of K with $S \supseteq S_\infty$,*
- *degree-d rational function $\phi(z) \in K(z)$ which is not a d-th power in $\overline{K}(z)$,*

we have

$$|\phi(K) \cap \mathfrak{o}_S^*| \leq C.$$

We will prove Conjecture 1.1 in case $\phi(z)$ is restricted to certain classes of rational functions, and we will also prove that the full conjecture is a consequence of a variant of the Caporaso–Harris–Mazur conjecture on uniform boundedness of rational points on curves of fixed genus.

We also consider a variant of Conjecture 1.1, which addresses S-units in an orbit of ϕ rather than in the image set $\phi(K)$. Here, for any $\alpha \in \mathbb{P}^1(K)$, the *orbit* of α under $\phi(z)$ is the set

$$\mathcal{O}_\phi(\alpha) := \{\phi^n(\alpha) : n \geq 1\},$$

where $\phi^n = \phi \circ \cdots \circ \phi$ denotes the n-fold composition of ϕ with itself. For any $\phi(z) \in K(z)$ of degree at least 2 such that $\phi^2(z) \notin K[z]$, Silverman [1993] showed that $\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S$ is finite. However, for any $\beta \in K^*$ the function $\psi(z) := \beta\phi(z/\beta)$ satisfies $\mathcal{O}_\psi(\alpha\beta) = \beta\mathcal{O}_\phi(\alpha)$, so the size of $\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S$ cannot be bounded independently of $\phi(z)$. We conjecture that there is a uniform bound on the number of S-units in an orbit:

Conjecture 1.2. *For any integers $s \geq 1$ and $d \geq 2$, there is a constant $C = C(s, d)$ such that for any*

- *number field K,*
- *s-element set S of places of K with $S \supseteq S_\infty$,*
- *degree-d rational function $\phi(z) \in K(z)$ which is not of the form $\beta z^{\pm d}$ with $\beta \in K^*$,*
- *$\alpha \in \mathbb{P}^1(K)$,*

we have

$$|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \leq C.$$

It turns out that this conjecture is a consequence of Conjecture 1.1:

Proposition 1.3. *If Conjecture 1.1 is true then Conjecture 1.2 is true.*

Remark 1.4. The hypotheses of Conjectures 1.1 and 1.2 imply that $[K : \mathbb{Q}] \leq 2s$, since $S_\infty \subseteq S$.

In Section 3 we prove the following preliminary results, which show that Conjectures 1.1 and 1.2 would be true if we allowed the constants C in those conjectures to depend on K, S, and ϕ, rather than just on s and d. We note that in the case of Conjecture 1.1 this simply says that $\phi(K) \cap \mathfrak{o}_S^*$ is finite. These results also indicate the special behavior of the functions excluded in the statements of these conjectures.

Proposition 1.5. *Let K be a number field, let S be a finite set of places of K with $S \supseteq S_\infty$, and let $\phi(z) \in K(z)$ be any rational function.*

(a) *If $|\phi^{-1}(\{0, \infty\})| \neq 2$ then $\phi(K) \cap \mathfrak{o}_S^*$ is finite.*

(b) *If $|\phi^{-1}(\{0, \infty\})| = 2$ then there is a finite set $S' \supseteq S$ for which $\phi(K) \cap \mathfrak{o}_{S'}^*$ is infinite.*

Proposition 1.6. *Let K be a number field, let S be a finite set of places of K with $S \supseteq S_\infty$, and let $\phi(z) \in K(z)$ have degree $d \geq 2$.*

(a) *If $\phi(z)$ does not have the form $\beta z^{\pm d}$ with $\beta \in K^*$, then there is a constant $C(K, S, \phi)$ such that every $\alpha \in \mathbb{P}^1(K)$ satisfies $|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \leq C(K, S, \phi)$.*

(b) *If $\phi(z) = \beta z^{\pm d}$ with $\beta \in K^*$, then there exist $\alpha \in \mathbb{P}^1(K)$ and a finite set $S' \supseteq S$ for which $\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_{S'}^*$ is infinite.*

We note that part (a) of each of these propositions follows from Siegel's theorem. For, if $|\phi^{-1}(\{0, \infty\})| > 2$ then $\psi(z) := \phi(z) + 1/\phi(z)$ has at least three poles so that $\psi(K) \cap \mathfrak{o}_S$ is finite; but $\psi(\beta)$ is in $\mathfrak{o}_S$ whenever $\phi(\beta)$ is in $\mathfrak{o}_S^*$, so also $\phi(K) \cap \mathfrak{o}_S^*$ is finite. Next, if $\phi^{-1}(\{0, \infty\})$ is a two-element set other than $\{0, \infty\}$, then Lemma 3.2 implies that $|\phi^{-2}(\{0, \infty\})| > 2$, so that $\phi^2(K) \cap \mathfrak{o}_S^*$ has size $N < \infty$, whence $|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \leq N + 1 = C(K, S, \phi)$.

In Section 2 we prove Conjectures 1.1 and 1.2 for some families of polynomial maps. The first family consists of monic polynomials in $\mathfrak{o}_S[z]$:

Theorem 1.7. *Let $s \geq 1$ and $d \geq 2$ be integers. There is a constant $C = C(s, d)$ such that for any*

- *number field K,*
- *s-element set S of places of K with $S \supseteq S_\infty$,*
- *degree-d monic polynomial $\phi(z) \in \mathfrak{o}_S[z]$ which does not equal $(z - \beta)^d$ for any $\beta \in K$,*

we have

$$|\phi(K) \cap \mathfrak{o}_S^*| \leq C.$$

Theorem 1.7 proves Conjecture 1.1 for monic polynomials in $\mathfrak{o}_S[z]$; for such polynomials, Conjecture 1.2 follows by applying Theorem 1.7 to $\phi^2(z)$.

We also prove Conjecture 1.2 for monic polynomials in $K[z]$ in which the coefficients of all but one term are in $\mathfrak{o}_S$, so long as this exceptional term does not have degree $d - 1$. We deduce this from the following more general result in v-adic dynamics.

Theorem 1.8. *Let K be a field with a nonarchimedean valuation v, and let*

$$\phi(z) = a_d z^d + \cdots + a_1 z + a_0 \in K[z]$$

be a polynomial satisfying

- $v(a_d) = 0$,
- *there is exactly one integer i for which $v(a_i) < 0$, and this exceptional i satisfies $i \neq d-1$.*

Then for each $\alpha \in K$, the set $\{n \geq 1 \mid v(\phi^n(\alpha)) = 0\}$ contains at most one element.

As an immediate corollary, we have the stated case of Conjecture 1.2:

Corollary 1.9. *Let K be a number field, and let S be a finite set of places of K with $S \supseteq S_\infty$. For any monic $\phi_0(z) \in \mathfrak{o}_S[z]$, any $\alpha, \beta \in K$ with $\beta \notin \mathfrak{o}_S$, and any integer i with $0 \leq i < \deg \phi_0 - 1$, the polynomial $\phi(z) := \phi_0(z) + \beta z^i$ satisfies*

$$|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \leq 1.$$

Remark 1.10. Conjecture 1.2 also follows from Theorem 2 of [Levin 2012] for rational functions of the form

$$\phi(z) := \frac{z^d + \beta_{d-1} z^{d-1} + \cdots + \beta_1 z}{\gamma_{d-1} z^{d-1} + \gamma_{d-2} z^{d-2} + \cdots + \gamma_1 z + 1}$$

with $\beta_1, \ldots, \beta_{d-1}, \gamma_1, \ldots, \gamma_{d-1} \in \mathfrak{o}_S$ and $\phi(z) \neq z^d$. For that theorem gives a uniform bound on the number of elements of K in the backwards orbit of any element of $\mathfrak{o}_S^*$. This also bounds the number of S-units in $\mathcal{O}_\phi(\alpha)$ for any $\alpha \in K$, since if $\phi^n(\alpha) \in \mathfrak{o}_S^*$ then $\alpha, \phi(\alpha), \ldots, \phi^{n-1}(\alpha)$ are elements of K in the backwards orbit of $\phi^n(\alpha)$.

We prove our conjectures for some further classes of rational functions in Section 4.

In Section 3 we show that our conjectures are consequences of the following variant of the deep conjecture of Caporaso, Harris and Mazur [Caporaso et al. 1997] concerning rational points on curves of a fixed genus.

Conjecture 1.11. *Fix integers $g \geq 2$ and $D \geq 1$. There is a constant $N = N(D, g)$ such that $|X(K)| \leq N$ for every smooth, projective, geometrically irreducible genus-g curve X defined over a degree-D number field K.*

Theorem 1.12. *If Conjecture 1.11 is true, then Conjecture 1.1 and Conjecture 1.2 are true.*

Remark 1.13. Conjecture 1.11 follows from the Bombieri–Lang conjecture [Pacelli 1997].

The referee provided the following geometric explanation of the difference between the questions of S-integers and S-units in the image set $\phi(K)$ of a rational function ϕ, indicating possible directions for future work. Writing $\phi(x/y) = \frac{f(x,y)}{g(x,y)}$ as the ratio of two coprime homogeneous polynomials, we see that the S-integral points of $\phi(K)$ correspond to the S-integral points of the quasi-affine variety cut out by

$$zg(x, y) = f(x, y) \ \text{ in } \ \mathbb{P}^1 \times \mathbb{A}^1.$$

Similarly, the S-unit points in $\phi(K)$ correspond to the S-integral points of the variety defined by

$$zg(x, y) = wf(x, y) \ \text{ and } \ zw = 1 \ \text{ in } \ \mathbb{P}^1 \times \mathbb{A}^2.$$

It would be interesting to seek generalizations of Conjecture 1.1 by considering more generally what sorts of families of varieties are likely to satisfy uniform boundedness statements for their S-integral points.

2. Special classes of rational functions

In this section we prove Theorems 1.7 and 1.8.

Proof of Theorem 1.7. Let K be a number field, let S be a finite set of places of K with $S \supseteq S_\infty$, and let $\phi(z) \in \mathfrak{o}_S[z]$ be monic of degree $d \geq 2$ with $\phi(z) \neq (z-\beta)^d$ for any $\beta \in K$. Then $\phi(z)$ has at least two distinct roots δ_1, δ_2 in $\bar{K}$. Let $K' = K(\delta_1, \delta_2)$ and let S' be the set of places of K' which lie over places in S, so that $|S'| \leq [K' : K]|S| \leq d(d-1)|S|$ and $\delta_i \in \mathfrak{o}_{S'}$. Then we can write

$$\phi(z) = (z - \delta_1)(z - \delta_2)\psi(z),$$

where $\psi(z)$ is a monic polynomial in $\mathfrak{o}_{S'}[z]$. Let $\gamma \in K$ satisfy $\phi(\gamma) \in \mathfrak{o}_S^*$. Then we must have $\gamma \in \mathfrak{o}_S$, so that both $u_i := \gamma - \delta_i$ and $\psi(\gamma)$ are in $\mathfrak{o}_{S'}$. Since $u_1 u_2 \psi(\gamma) = \phi(\gamma)$ is in $\mathfrak{o}_S^*$, it follows that $u_1, u_2 \in \mathfrak{o}_{S'}^*$. In addition we have

$$\frac{1}{\delta_2 - \delta_1} u_1 - \frac{1}{\delta_2 - \delta_1} u_2 = 1. \tag{2-1}$$

Moreover, γ is uniquely determined by u_1, so the number of elements $\gamma \in \mathfrak{o}_S$ for which $\phi(\gamma) \in \mathfrak{o}_S^*$ is at most the number of solutions to (2-1) in elements $u_1, u_2 \in \mathfrak{o}_{S'}^*$. Finally, by [Evertse 1984], the number of such solutions is at most $C_1 C_2^{|S'|-1}$ for some absolute constants C_1, C_2 (in fact, we can take $C_1 = C_2 = 256$ [Beukers and Schlickewei 1996]). Therefore $|\phi(K) \cap \mathfrak{o}_S^*|$ is bounded by a function of $|S'|$, and hence by a function of $|S|$ and d. □

Proof of Theorem 1.8. Suppose that $\mathcal{O}_\phi(\alpha)$ contains a unit of the valuation ring, and let m be the least positive integer for which $v(\phi^m(\alpha)) = 0$. Writing $\gamma := \phi^m(\alpha)$, we will show by induction that $|\phi^n(\gamma)|_v = |a_i|_v^{d^{n-1}}$ for every $n \geq 1$. The strong triangle inequality implies that $|\phi(\gamma)|_v = |a_i|_v$, proving the base case $n = 1$. If $\delta := \phi^n(\gamma)$ satisfies $|\delta|_v = |a_i|_v^{d^{n-1}}$ for some $n \geq 1$, then $|a_i \delta^i|_v = |a_i|_v^{1+id^{n-1}}$ and $|a_j \delta^j|_v \leq |a_i|_v^{jd^{n-1}}$ for $j \neq i$, with equality when $j = d$. Our hypothesis $i < d - 1$

implies that $d^n > 1 + id^{n-1}$, so that $|\phi^{n+1}(\gamma)|_v = |a_i|_v^{d^n}$, which completes the induction. It follows that $v(\phi^n(\gamma)) < 0$ for every $n > 0$, so that $\mathcal{O}_\phi(\alpha)$ contains exactly one unit, which concludes the proof. $\square$

3. Connection with rational points on curves

In this section we prove Theorem 1.12 and Propositions 1.3, 1.5, and 1.6. We begin by relating S-units in the image set $\phi(K)$ of a rational function to rational points on certain curves.

Lemma 3.1. *Let K be a number field, let S be a finite set of places of K with $S \supseteq S_\infty$, and let $\phi(z) \in K(z)$ be a nonconstant rational function. For any prime p with $p > \deg \phi$, there are elements $\gamma_1, \ldots, \gamma_t \in \mathfrak{o}_S^*$, where $t \le p^{|S|}$, with the following properties*:

- *For each i, the affine curve X_i defined by $y^p = \gamma_i \phi(z)$ is geometrically irreducible.*
- *We have $|\phi(K) \cap \mathfrak{o}_S^*| \le \sum_{i=1}^t N_i$ where N_i is the number of points in $X_i(K)$ having nonzero y-coordinate.*

Proof. First note that $y^p = \gamma\phi(z)$ is geometrically irreducible for any $\gamma \in K^*$, since $\gamma\phi(z)$ is not a p-th power in $\overline{K}(z)$. Dirichlet's S-unit theorem asserts that $\mathfrak{o}_S^* \cong \mu_K \times \mathbb{Z}^{|S|-1}$, where μ_K denotes the group of roots of unity in K. Since μ_K is cyclic, it follows that $\mathfrak{o}_S^*/(\mathfrak{o}_S^*)^p \cong (\mathbb{Z}/p\mathbb{Z})^r$ where $r \in \{|S|-1, |S|\}$. Let Γ be a set of p^r elements in $\mathfrak{o}_S^*$ whose images in $\mathfrak{o}_S^*/(\mathfrak{o}_S^*)^p$ are pairwise distinct. For any $\beta \in K$ such that $\phi(\beta) \in \mathfrak{o}_S^*$, we can write $\phi(\beta) = \gamma^{-1}\delta^p$ for some $\gamma \in \Gamma$ and $\delta \in \mathfrak{o}_S^*$. Then (δ, β) is a K-rational point on the curve $y^p = \gamma\phi(z)$ whose y-coordinate is nonzero. Since the z-coordinate of this point is β, the result follows. $\square$

Proof of Theorem 1.12. By Proposition 1.3, it suffices to show that Conjecture 1.11 implies Conjecture 1.1. Let K be a number field, let S be a finite set of places of K with $S \supseteq S_\infty$, and let $\phi(z) \in K(z)$ have degree $d \ge 2$. Assume that $\phi(z)$ is not a d-th power in $\overline{K}(z)$, so that $m := |\phi^{-1}(\{0, \infty\})|$ is at least 3. Let p be the smallest prime for which $p > d$ and $(p-1)(m-2) > 2$. Then $p = 5$ if $d = 2$ and $m = 3$, and in all other cases $p < 2d$ by Bertrand's postulate. Let $\gamma_1, \ldots, \gamma_t$ satisfy the conclusion of Lemma 3.1, so that $\gamma_i \in K^*$ and $t \le p^{|S|}$. Writing X_i for the curve $y^p = \gamma_i\phi(z)$, and N_i for the number of points in $X_i(K)$ having nonzero y-coordinate, it follows that $|\phi(K) \cap \mathfrak{o}_S^*| \le \sum_{i=1}^t N_i$. Since every point on X_i having nonzero y-coordinate is nonsingular, we see that N_i is bounded above by the number of K-rational points on the unique smooth projective curve Y_i over K which is birational to X_i. Since $p > d$, the classical genus formula for Kummer covers [Stichtenoth 2009, Proposition III.7.3] implies that the genus g of Y_i is

$(p-1)(m-2)/2$. Thus our choice of p ensures that

$$2 \le g \le \tfrac{1}{2}\big(\tfrac{5}{2}d-1\big)\big(2d-2\big).$$

If Conjecture 1.11 is true then $|Y_i(K)|$ is bounded by a constant which depends only on the genus of $Y_i(K)$ and the degree $[K:\mathbb{Q}]$. Since the genus is bounded by a function of d, and the degree $[K:\mathbb{Q}]$ is bounded by a function of $|S|$ (by Remark 1.4), it follows that $|Y_i(K)|$ is bounded by a constant depending on d and $|S|$. Since $t \le p^{|S|} \le (5d/2)^{|S|}$, this proves that Conjecture 1.11 implies Conjecture 1.1. □

Our proof of Proposition 1.3 relies on the following well-known lemma.

Lemma 3.2. *Let $\phi(z) \in \mathbb{C}(z)$ be any rational function of degree $d \ge 2$ which is not of the form $\beta z^{\pm d}$ with $\beta \in \mathbb{C}^*$. Then $|\phi^{-2}(\{0,\infty\})| \ge 3$.*

Proof. Write $m := |\phi^{-2}(\{0,\infty\})|$, so we must show that $m \ge 3$. Plainly $m \ge |\phi^{-1}(\{0,\infty\})| \ge 2$, so the conclusion holds unless $|\phi^{-1}(\{0,\infty\})| = 2$. In this case ϕ is totally ramified over both 0 and ∞, so the Riemann–Hurwitz formula (or writing down $\phi(z)$) implies that ϕ is unramified over all other points. Since $\phi(z)$ does not have the form $\beta z^{\pm d}$, we know that $\phi^{-1}(\{0,\infty\}) \ne \{0,\infty\}$, so that at least one point in $\phi^{-1}(\{0,\infty\})$ has d distinct ϕ-preimages. Since each point has at least one preimage, we conclude that $m \ge d+1 \ge 3$, as desired. □

Proof of Proposition 1.3. If $\phi(z) \ne \beta z^{\pm d}$ then $\phi^2(z)$ has a total of at least three zeroes and poles by Lemma 3.2, and hence is not a d^2-th power in $\bar{K}(z)$. Thus Conjecture 1.1 implies that $|\phi^2(K) \cap \mathfrak{o}_S^*| \le C(s,d)$, so that

$$|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \le C(s,d) + 1. \qquad \square$$

Part (a) of Proposition 1.5 follows from our proof of Theorem 1.12, by using Faltings' theorem [1983] instead of Conjecture 1.11. We now give a more elementary proof of Proposition 1.5.

Proof of Proposition 1.5. If $|\phi^{-1}(\{0,\infty\})| > 2$, the function $\psi(z) := \phi(z) + 1/\phi(z)$ satisfies $|\psi^{-1}(\{0,\infty\})| \ge 3$, so $\psi(K) \cap \mathfrak{o}_S$ is finite by Siegel's theorem; but $\psi(\beta)$ is in $\mathfrak{o}_S$ whenever $\phi(\beta)$ is in $\mathfrak{o}_S^*$, so it follows that $\phi(K) \cap \mathfrak{o}_S^*$ is finite. Now assume that $|\phi^{-1}(\{0,\infty\})| = 2$, so that $\phi(z) = \gamma\mu(z)^d$ for some $d \ge 1$, some $\gamma \in K^*$, and some degree-one $\mu(z) \in K(z)$. Let S' be a finite set of places of K such that $\gamma \in \mathfrak{o}_{S'}^*$, $S' \supseteq S$, and $|S'| > 1$. Since $\mu(K)$ contains all but at most one element of K, it follows that $\phi(K)$ contains all but at most one element of $\gamma(\mathfrak{o}_{S'}^*)^d$. Since $\gamma \in \mathfrak{o}_{S'}^*$ and $|S'| > 1$, this shows that $\phi(K) \cap \mathfrak{o}_{S'}^*$ is infinite. □

Proof of Proposition 1.6. If $\phi(z)$ does not have the form $\beta z^{\pm d}$ then $|\phi^{-2}(\{0,\infty\})| \ge 3$ by Lemma 3.2, so Proposition 1.5 implies that $\phi^2(K) \cap \mathfrak{o}_S^*$ has size $N < \infty$, whence

$$|\mathcal{O}_\phi(\alpha) \cap \mathfrak{o}_S^*| \le N + 1 = C(K,S,\phi).$$

Now consider $\phi(z)=\beta z^{\pm d}$ with $\beta\in K^*$ and $d\geq 2$. Any $\alpha\in K^*$ satisfies $\mathcal{O}_\phi(\alpha)\subseteq\mathfrak{o}_{S'}^*$ where S' is the union of S with the set of places v of K for which $|\alpha|_v\neq 1$ or $|\beta|_v\neq 1$. If $\alpha\in K^*$ is not a root of unity then $\mathcal{O}_\phi(\alpha)$ is infinite, so that $\mathcal{O}_\phi(\alpha)\cap\mathfrak{o}_{S'}^*$ is infinite. □

4. Additional remarks

We make two final remarks. First, the proofs of Theorems 1.7 and 1.8 can be modified to treat some classes of Laurent polynomials. For example, let d and d' be distinct positive integers, and let $\phi(z)=(\gamma_d z^d+\cdots+\gamma_1 z+\gamma_0)/z^{d'}$ where $\gamma_i\in\mathfrak{o}_S$ and $\gamma_d,\gamma_0\in\mathfrak{o}_S^*$. Suppose in addition that the numerator is not a d-th power in $\overline{K}[z]$. Then $|\phi(K)\cap\mathfrak{o}_S^*|\leq C(s,d)$ for any $\alpha\in\mathbb{P}^1(K)$. Indeed, since γ_0 and γ_d are assumed to be units, $\phi(\beta)$ cannot be in $\mathfrak{o}_S^*$ if $|\beta|_v\neq 1$ for some $v\notin S$. Thus we need only consider $\beta\in\mathfrak{o}_S^*$, and now the desired bound follows from the proof of Theorem 1.7.

As another example, consider $\phi(z)=(\gamma_d z^d+\cdots+\gamma_1 z+\gamma_0)/z^{d'}$ where $d>d'$, $\gamma_i\in K$, and there is some $v\notin S$ for which $|\gamma_d|_v>\max(1,|\gamma_i|_v)$ for each $i<d$. Then $|\mathcal{O}_\phi(\alpha)\cap\mathfrak{o}_S^*|\leq 1$ for any $\alpha\in\mathbb{P}^1(K)$, as the orbit of an S-unit cannot contain another S-integer by the proof of Theorem 1.8. Both this class of examples and the previous class are quite special, but they serve as further evidence for Conjectures 1.1 and 1.2.

We conclude this paper by noting that the constant C that appears in Conjectures 1.1 and 1.2 must depend on both s and d. The necessity of dependence on s is clear. Dependence on d is also required, since by Lagrange interpolation one can construct polynomials $\phi(z)\in K[z]$ in which the first several $\phi^i(\alpha)$ take on any prescribed distinct values in K while also $\phi(z)$ has at least two zeroes (and hence is not $\beta z^{\pm d}$).

Acknowledgments

We thank ICERM, where collaboration for this project began at the 2012 ICERM workshop on Global Arithmetic Dynamics. We also thank the referee for a thorough report which improved the exposition and content of this paper.

References

[Beukers and Schlickewei 1996] F. Beukers and H. P. Schlickewei, "The equation $x+y=1$ in finitely generated groups", *Acta Arith.* **78**:2 (1996), 189–199. MR 97k:11051 Zbl 0880.11034

[Caporaso et al. 1997] L. Caporaso, J. Harris, and B. Mazur, "Uniformity of rational points", *J. Amer. Math. Soc.* **10**:1 (1997), 1–35. MR 97d:14033 Zbl 0872.14017

[Evertse 1984] J.-H. Evertse, "On equations in S-units and the Thue–Mahler equation", *Invent. Math.* **75**:3 (1984), 561–584. MR 85f:11048 Zbl 0521.10015

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. MR 85g:11026a Zbl 0588.14026

[Levin 2012] A. Levin, "Rational preimages in families of dynamical systems", *Monatsh. Math.* **168**:3-4 (2012), 473–501. MR 2993960 Zbl 06111280

[Pacelli 1997] P. L. Pacelli, "Uniform boundedness for rational points", *Duke Math. J.* **88**:1 (1997), 77–102. MR 98b:14020 Zbl 0935.14016

[Silverman 1993] J. H. Silverman, "Integer points, Diophantine approximation, and iteration of rational maps", *Duke Math. J.* **71**:3 (1993), 793–829. MR 95e:11070 Zbl 0811.11052

[Stichtenoth 2009] H. Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics **254**, Springer, Berlin, 2009. MR 2010d:14034 Zbl 1155.14022

Received June 19, 2014. Revised August 27, 2014.

HOLLY KRIEGER
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MA 02139
UNITED STATES

hkrieger@math.mit.edu

AARON LEVIN
DEPARTMENT OF MATHEMATICS
MICHIGAN STATE UNIVERSITY
EAST LANSING, MI 48824
UNITED STATES

adlevin@math.msu.edu

ZACHARY SCHERR
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF PENNSYLVANIA
DAVID RITTENHOUSE LAB
PHILADELPHIA, PA 19104–6395
UNITED STATES

zscherr@math.upenn.edu

THOMAS TUCKER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14627
UNITED STATES

thomas.tucker@rochester.edu

YU YASUFUKU
DEPARTMENT OF MATHEMATICS
COLLEGE OF SCIENCE AND TECHNOLOGY
NIHON UNIVERSITY
CHIYODA-KU, TOKYO 101-8308
JAPAN

yasufuku@math.cst.nihon-u.ac.jp

MICHAEL E. ZIEVE
MATHEMATICAL SCIENCES CENTER
TSINGHUA UNIVERSITY
BEIJING, 100084
CHINA

and

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MI 48109–1043
UNITED STATES

zieve@umich.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.107

A COUNTEREXAMPLE TO THE ENERGY IDENTITY FOR SEQUENCES OF α-HARMONIC MAPS

YUXIANG LI AND YOUDE WANG

We construct a closed Riemannian manifold (N, h) and a sequence of α-harmonic maps from S^2 into N with uniformly bounded energy such that the energy identity for this sequence is not true.

1. Introduction

Let (Σ, g) be a Riemann surface and (N, h) be an n-dimensional smooth compact Riemannian manifold which is embedded in $\mathbb{R}^K$. Usually, we denote the space of Sobolev maps from Σ into N by $W^{k,p}(\Sigma, N)$, which is defined by

$$W^{k,p}(\Sigma, N) = \{u \in W^{k,p}(\Sigma, \mathbb{R}^K) : u(x) \in N \ \text{ for a.e. } x \in \Sigma\}.$$

For $u \in W^{1,2}(\Sigma, N)$, we define locally the energy density $e(u)$ of u at $x \in \Sigma$ by

$$e(u)(x) = |\nabla_g u|^2 = g^{ij}(x) h_{\alpha\beta}(u(x)) \frac{\partial u^\alpha}{\partial x^i} \frac{\partial u^\beta}{\partial x^j}.$$

The energy of u on Σ, denoted by $E(u)$ or $E(u, \Sigma)$, is defined by

$$E(u) = \frac{1}{2} \int_\Sigma e(u)\, dV_g,$$

and the critical points of E are called harmonic maps. We know that a harmonic map u satisfies

$$\tau(u) = \Delta u + A(u)(\nabla u, \nabla u) = 0,$$

where A is the second fundamental form of N in $\mathbb{R}^K$. Harmonic maps are related very closely to minimal surface. It is well known that a harmonic map from S^2 into N must be a branched conformal immersion in N.

Unfortunately, E does not satisfy the Palais–Smale condition. From the viewpoint of calculus of variation, it is difficult to show the existence of harmonic maps from a

MSC2010: primary 58E20; secondary 35Q60.
Keywords: α-harmonic map, energy identity.

surface. In order to obtain harmonic maps, Sacks and Uhlenbeck [1981] introduced the so-called α-energy E_α, instead of L^2 energy E, as

$$E_\alpha(u) = \frac{1}{2}\int_\Sigma \{(1+|\nabla u|^2)^\alpha - 1\}\, dV_g,$$

where we always assume that $\alpha > 1$. It is well known that this α-energy functional E_α satisfies the Palais–Smale condition. The critical points of E_α in $W^{1,2\alpha}(\Sigma, N)$, called α-harmonic maps, satisfy

$$\Delta_g u_\alpha + (\alpha-1)\frac{\nabla_g|\nabla_g u_\alpha|^2\nabla_g u_\alpha}{1+|\nabla_g u_\alpha|^2} + A(u_\alpha)(du_\alpha, du_\alpha) = 0. \tag{1-1}$$

The strategy of Sacks and Uhlenbeck is to employ a sequence of α-harmonic maps to approximate a harmonic map as α tends to 1. Hence, to show the existence of harmonic maps we need to study the convergence behavior of a sequence of α-harmonic maps u_α with $E_\alpha(u_\alpha) < C$ from a compact surface (Σ, g) into a compact Riemannian manifold (N, h) without boundary. Generally, such a sequence converges weakly to a harmonic map in $W^{1,2}(\Sigma, N)$ and strongly in C^∞ away from a finite set of points in Σ.

Concretely, let $\{u_{\alpha_k}\}$ be a sequence of α-harmonic maps from Σ into N with uniformly bounded α-energy, that is, $E_{\alpha_k}(u_{\alpha_k}) < \Lambda < \infty$. We assume that the sequence does not converge smoothly on Σ. By the theory of Sacks and Uhlenbeck, there exists a subsequence of $\{u_{\alpha_k}\}$, still denoted by $\{u_{\alpha_k}\}$, and a finite set $\mathcal{S} \subset \Sigma$ such that the subsequence converges to a harmonic map u_0 in $C^\infty_{\mathrm{loc}}(\Sigma\setminus\mathcal{S})$. We know that, at each point $p_i \in \mathcal{S}$, the energy of the subsequence concentrates and the blowup phenomena occurs. Moreover, there exist point sequences $\{x^l_{i_k}\}$ in Σ with $\lim_{k\to+\infty} x^l_{i_k} = p_i$ and scaling constant number sequences $\{\lambda^l_{i_k}\}$ with $\lim_{k\to+\infty}\lambda^l_{i_k} \to 0$, $l = 1, \dots, n_0$, such that

$$u_{\alpha_k}(x^l_{i_k} + \lambda^l_{i_k}x) \to v^l \quad \text{in } C^j_{\mathrm{loc}}(\mathbb{R}^2\setminus\mathcal{A}^i),$$

where all v^i are nontrivial harmonic maps from S^2 into N, and $\mathcal{A}^i \subset \mathbb{R}^2$ is a finite set.

In order to explore and describe the asymptotic behavior of $\{u_{\alpha_k}\}$ at each blowup point, the following two problems arise naturally. The first is whether or not the energy identity holds true:

$$\lim_{\alpha_k\to 1} E_{\alpha_k}(u_{\alpha_k}, B^\Sigma_{r_0}(p_i)) = E(u_0, B^\Sigma_{r_0}(p_i)) + \sum_{l=1}^{n_0} E(v^l).$$

Here, $B^\Sigma_{r_0}(p_i)$ is a geodesic ball in Σ which contains only one blowup point p_i. The other is whether or not the necks connecting bubbles are some geodesics of finite length?

Considerable progress has been made regarding these problems; let us now recall some main results on them. Chen and Tian [1999] considered a special sequence $\{u_{\alpha_k}\}$ with uniformly bounded α-energy, for which every u_{α_k} is a minimizing α_k-harmonic map and all maps u_{α_k} belong to a fixed homotopy class. They studied the convergence behavior of such a special sequence and provided a proof on the above energy identity. Later, for the same sequence, Li and Wang [2010a] gave another constructing proof on the energy identity, which is completely different from that given in [Chen and Tian 1999].

The energy identity for a minimax sequence of α-harmonic maps has also been considered. Suppose that A is a parameter manifold. Let $h_0 : \Sigma \times A \to N$ be a continuous map, and H be such a set of continuous maps $h : \Sigma \times A \to N$ that every $h \in H$ is homotopic to h_0 and satisfies $h(t) \in W^{1,2\alpha}(\Sigma, N)$ for any fixed $t \in A$. Set

$$\beta_\alpha(H) = \inf_{h \in H} \sup_{t \in A} E_\alpha(h(\cdot\,, t)).$$

It is known that there is at least a sequence $\{u_{\alpha_k}\}$, each u_{α_k} of which attains $\beta_{\alpha_k}(H)$, satisfies the energy identity as $\alpha_k \to 1$. For more details, we refer to [Jost 1991; Lamm 2010].

On the other hand, it should be pointed out that some effective methods have been established to successfully prove the energy identity and give a detailed description of the connecting necks for the heat flow of harmonic maps from a Riemann surface, or more generally, a sequence of maps from a Riemann surface with tension fields τ bounded in the sense of L^2 [Ding 1998; Ding and Tian 1995; Qing 1995; Qing and Tian 1997].

Recently, Li and Wang [2010b] studied the above problems on the sequences of α-harmonic maps and obtained some results which can be summarized as follows. If the energy concentration phenomena appears for $\{u_{\alpha_k}\}$, one can prove a weak energy identity and a direct convergence relation between the blowup radius and the parameter α, which ensures the energy identity and no-neck property. Li and Wang also showed that the necks converge to some geodesics and gave a length formula for the neck in the case where only one bubble appears.

Motivated by an example given by Duzaar and Kuwert [1998], Li and Wang [2010b] also constructed an α-harmonic map sequence with uniformly bounded energy, for which the blowup phenomenon occurs and there exists at least a neck (geodesic) of infinite length. This answers negatively the second problem on α-harmonic map sequence.

Although some mathematicians think that the energy identity for the sequence of α-harmonic maps should also be true, up to now it has been unclear in general whether the energy identity for an α-harmonic map sequence with bounded energy holds true or not. In this short paper, we will modify the construction in [Li and Wang 2010b] to show that the energy identity is also not true.

On the other hand, a natural problem is whether the set of the values of energy for harmonic spheres in any given Riemannian manifold (N, h) is discrete or not, since the bubbles produced in the convergence of a sequence of α-harmonic maps from (Σ, g) are always harmonic spheres.

We denote this set by

$$\mathscr{E}(N, h) = \{E(u) : u \text{ is a harmonic map from } S^2 \text{ into } (N, h)\}.$$

It is well known that if (N, h) is the standard sphere S^2, we have

$$\mathscr{E}(N, h) = \{4k\pi : k = 0, 1, \dots, n, \dots\}.$$

We also know from [Valli 1988] that if (N, h) is the unitary group $U(n)$ with the standard metric, then the energy of harmonic maps $S^2 \to U(n)$ can take as values only integral multiples of 8π. Some other energy gap phenomena on unitons were discussed in [Anand 1995; Dong 2002; Uhlenbeck 1989]. Some mathematicians conjectured that $\mathscr{E}(N, h)$ is a discrete set. Here, we will also give a counterexample to show that $\mathscr{E}(N, h)$ is not discrete.

2. α-harmonic maps

Later, we will discuss the convergence behavior of some α-harmonic map sequences with uniformly bounded α-energy or L^2 energy. In fact, by discussing the convergence of α-harmonic map sequences, Sacks and Uhlenbeck developed an existence theory on minimal surfaces in [Sacks and Uhlenbeck 1981; 1982]. In particular, they established the well-known ϵ-regularity theorem on α-harmonic maps and removal singularity theorem on harmonic maps [1981], which will be used repeatedly in the present paper.

Theorem 2.1. *Let $D = D_1(0) = \{z : |z| < 1\} \subset \mathbb{C}$ be a disk with radius 1 and N be a Riemannian manifold. Assume that $u : D \to N$ satisfies Equation (1-1). Then there exists $\epsilon_0 > 0$ and $\alpha_0 > 1$ such that if $E(u, D) < \epsilon_0$ and $1 \le \alpha \le \alpha_0$, then we have*

$$\|\nabla^k u\|_{L^\infty(D_{1/2})} \le C(k) E(u, D).$$

Theorem 2.2. *Assume that $u : D \setminus \{0\} \to N$ is a harmonic map with $E(u) < +\infty$. Then u is a harmonic map from D into N.*

The above theorem tells us that, if u is a harmonic map from $\mathbb{C} \setminus \{p_i \in \mathbb{C} : i = 1, 2, \dots, l < \infty\}$ into N with $E(u) < +\infty$, then u can be viewed as a harmonic map from S^2 into N.

Now, we can state more precisely the energy concentration of $\{u_{\alpha_k}\}$. Let $B_t^\Sigma(x)$ denote the geodesic ball of Σ which is centered at x and has geodesic radius t. By Theorem 2.1, the finite singular set of $\{u_{\alpha_k}\}$ can be defined precisely by

$$\mathcal{S} = \left\{ x \in \Sigma : \lim_{t\to 0} \varliminf_{k\to+\infty} \int_{B_t^\Sigma(x)} |\nabla u_{\alpha_k}|^2 \geq \frac{\epsilon_0}{2} \right\}.$$

For any $\tilde{x}_0 \notin \mathcal{S}$, there exists $\delta > 0$ such that $E(u_{\alpha_k}, B_\delta^\Sigma(\tilde{x}_0)) < \epsilon_0$. Applying Theorem 2.1, $\{u_{\alpha_k}\}$ converges smoothly on any $\Omega \Subset \Sigma \setminus \mathcal{S}$. The limit map is a harmonic map from $\Sigma\setminus\mathcal{S}$ into N. Theorem 2.2 tells us that the singular points of the limit map can be removed, in other words, it is a harmonic map from Σ into N.

If $x_0 \in \mathcal{S}$, it is easy to check that

$$\|\nabla u_{\alpha_k}\|_{C^0(B_t^\Sigma(x_0))} \to +\infty$$

for any t. Choose $x_{\alpha_k} \in B_\delta^\Sigma(x_0)$ such that

$$|\nabla u_{\alpha_k}(x_{\alpha_k})| = \max_{B_\delta^\Sigma(x_0)} |\nabla u_{\alpha_k}|,$$

and let

$$\lambda_{\alpha_k} = \frac{1}{\max_{B_\delta^\Sigma(x_0)} |\nabla u_{\alpha_k}|}.$$

It is easy to see that $x_{\alpha_k} \to x_0$ as $k \to \infty$. Then, in an isothermal coordinate system around x_0, we may define

$$v_k(x) = u_{\alpha_k}(x_{\alpha_k} + \lambda_{\alpha_k} x).$$

It is well known that v_k converges in $C^\infty(D_R)$ to a harmonic map $v^1 : \mathbb{C} \to N$ for any fixed R, where $D_R = D_R(0) = \{z : |z| < R\} \subset \mathbb{C}$ is a disk with radius $R > 0$. We can regard v^1 as a harmonic map from S^2 into N. Usually, v^1 is called the first bubble. For the details on getting all the bubbles we refer to the appendix of [Li and Wang 2010b]. Moreover, in [Li and Wang 2010a] (see also [Chen and Tian 1999; Hong and Yin 2010]) we prove the following theorem which will be used later.

Theorem 2.3. *Let (Σ, g) be a closed Riemann surface and N a compact Riemannian manifold. Suppose that H is a fixed homotopy class of maps from Σ into N and u_α is a minimizer of E_α in the set $W^{1,2\alpha}(\Sigma, N) \cap H$. Then when $\alpha \to 1$ there exists a subsequence $\{u_\alpha\}$ and harmonic map u_0 such that $\{u_\alpha\}$ converges to u_0 weakly in $W^{1,2}(\Sigma, N)$ and blows up at finitely many points $\{p_i : i = 1, 2, \dots, m\}$. Moreover, associated with each $\{p_i\}$ there exist finitely many harmonic maps w_{i_j} from S^2 into N, $j = 1, 2, \dots, i_0$, such that*

$$\lim_{\alpha\to 1} E_\alpha(u_\alpha) = E(u_0) + \sum_{i=1}^{m} \sum_{j=1}^{i_0} E(w_{i_j}).$$

3. Construction of the counterexample

3A. ***Constructing the manifold (N, h).*** Let h_1 be the standard metric on

$$Y_1 = \mathbb{T}^3 = S^1 \times S^1 \times S^1 = \mathbb{R}^3/2\pi\mathbb{Z} \oplus 2\pi\mathbb{Z} \oplus 2\pi\mathbb{Z}.$$

Let $B_r(p)$ denote a geodesic ball in $\mathbb{T}^3$ with radius r and center p. Fix a point $p \in Y_1$, and set

$$X_1 = \mathbb{T}^3 \setminus B_r(p),$$

where $r < \pi/(4\sqrt{3}+2)$. It is easy to see that the injective radius of Y_1 at p is π and $B_\pi(p) \setminus B_r(p)$ is isometric to

$$\mathbb{T}_0 = \big(S^2 \times (-\log \pi, -\log r], e^{-2t}(d\mathfrak{s}^2 + dt^2)\big),$$

where $g_{\mathfrak{s}} = d\mathfrak{s}^2$ is the standard metric over S^2. It is also easy to check that $\mathbb{T}_0$ is isometric to

$$\mathbb{T}_0' = \left(S^2 \times \left[0, \log \frac{\pi}{r}\right), e^{2t+2\log r}(d\mathfrak{s}^2 + dt^2)\right)$$

and

$$\mathbb{T}_0'' = \left(S^2 \times \left(-\log \frac{\pi}{r}, 0\right], e^{-2t+2\log r}(d\mathfrak{s}^2 + dt^2)\right).$$

Let $(X_2, h_2) = (X_1, h_1)$. We consider the quotient space of $X_1 \cup X_2$, obtained by gluing every point $x \in \partial X_1$ with the same point $x \in \partial X_2$ together. In this way, we get a closed compact manifold N and a projection map $\phi : X_1 \cup X_2 \to N$. We set

$$M = \phi(\partial B_r(p)).$$

On $N \setminus M$, the metric $h_0 = (\phi^{-1})^*(h_1) \cup (\phi^{-1})^*(h_2)$ is well defined and can be extended to a metric g_0 over N. However, g_0 is not smooth and need to be modified. Obviously, M has a neighborhood which is isometric to

$$T = \left(S^2 \times \left(-\log \frac{\pi}{r}, \log \frac{\pi}{r}\right), e^{2|t|+2\log r}(d\mathfrak{s}^2 + dt^2)\right).$$

In fact, T is obtained by gluing $\mathbb{T}_0'$ and $\mathbb{T}_0''$ along $S^2 \times \{0\}$.

We let ψ be a smooth function defined on $\big(-\log \frac{\pi}{r}, \log \frac{\pi}{r}\big)$ which satisfies

(1) $\psi = e^{2|t|+2\log r}$ when $|t| \geq \log 2$;

(2) $\psi' < 0$ on $(-\log 2, 0)$ and $\psi' > 0$ on $(0, \log 2)$.

Note that (2) implies that 0 is the only critical point of ψ on $(-\log 2, \log 2)$.

We define a new metric h on N which is h_0 on $N \setminus T$, and $\psi(t)(d\mathfrak{s}^2 + dt^2)$ on T. It is easy to see that h is smooth on N. For convenience, we set

$$Q(a) = S^2 \times \left(-\log \frac{a}{r}, \log \frac{a}{r}\right) \subset T.$$

Obviously, we have

$$\phi^{-1}(Q(a)) \cap X_1 = B_a(p) \setminus B_r(p) \subset Y_1.$$

Lemma 3.1. *Let (N,h), T and $Q(a)$ be defined as above. Assume that $u : S^2 \to (N,h)$ is a nontrivial harmonic map with $u(S^2) \subset Q(\pi) = T$. Then u is a harmonic map from S^2 into M.*

Proof. Let $u = (v, f) : S^2 \to Q(\pi)$ be a harmonic map, where $v \in C^\infty(S^2, S^2)$ and $f \in C^\infty(S^2)$. The energy can be written as

$$E(u) = \frac{1}{2}\int_{S^2} |\nabla u|^2 \, dV = \frac{1}{2}\int_{S^2} \big(|\nabla v|^2 + |\nabla f|^2\big)\psi(f)\, dV.$$

Here $dV = dV_{g_{\mathfrak{s}}}$ is the standard volume form of S^2. By a direct calculation, it is easy to see that u satisfies

$$\begin{aligned} -\nabla(\psi(f)\nabla v) + \psi(f)|\nabla v|^2 v = 0, \\ -\nabla(\psi(f)\nabla f) + \tfrac{1}{2}\big(|\nabla v|^2 + |\nabla f|^2\big)\psi'(f) = 0. \end{aligned} \tag{3-1}$$

Multiplying both sides of the second equation of (3-1) by f and then integrating the obtained identity over S^2, we get the identity

$$\int_{S^2} \Big(|\nabla f|^2 \psi(f) + \tfrac{1}{2}\big(|\nabla v|^2 + |\nabla f|^2\big)\psi'(f) f\Big)\, dV = 0.$$

Noting that $\psi'(f) f \geq 0$ always holds true, we infer from the above identity

$$\int_{S^2} |\nabla f|^2 \psi(f)\, d\mathfrak{s} = \frac{1}{2}\int_{S^2} \big(|\nabla v|^2 + |\nabla f|^2\big)\psi'(f) f \, dV = 0.$$

This implies that $\nabla f = 0$ and f is a constant. Moreover, from the above identity we also have

$$|\nabla v|^2 \psi' f \equiv 0.$$

Since u is nontrivial by assumption, there always exists a point $x_1 \in S^2$ such that $|\nabla v|(x_1) \neq 0$. Hence we conclude that $\psi'(f) f \equiv 0$ which implies $f \equiv 0$. It follows that v is a harmonic map from S^2 into M. □

Lemma 3.2. *Let (N,h) and Q be the same as in Lemma 3.1. Assume that u is a harmonic map from S^2 into (N,h) such that $u(S^2) \cap Q(2r) \neq \varnothing$ and $u(S^2) \cap \partial Q(\pi) \neq \varnothing$. Then we have*

$$E(u) \geq \pi(\pi - 2r)^2.$$

Proof. Without loss of generality, we assume $p_1 \in X_1$ is such that p_1 is in $\partial B_\pi(p)$ in Y_1 and $\phi(p_1)$ is in $u(S^2)$. First, u is a branched minimal surface since u is a harmonic map from S^2 into N. On the other hand, as h is flat on $\phi(B_{\pi-2r}(p_1))$, it is easy to check that $u(S^2)\cap\phi(B_{\pi-2r}(p_1))$ is a stationary varifold. Denote by $\mu(u(S^2)\cap B_{\pi-2r}(p_1))$ the area of $u(S^2)\cap B_{\pi-2r}(p_1)$. By the monotonicity inequality for stationary varifolds (see [Simon 1983]), we have

$$\frac{\mu(u(S^2)\cap B_{\pi-2r}(p_1))}{\pi(\pi-2r)^2}\geq 1.$$

In light of this inequality and the fact $E(u)\geq\mu(u(S^2)\cap B_{\pi-2r}(p_1))$, we derive the desired inequality

$$E(u)\geq\pi(\pi-2r)^2;$$

and the proof is complete. □

Since h is flat on $N\setminus Q(2r)$, we have the following lemma.

Lemma 3.3. *Let (N,h) and Q be the same as in Lemma 3.1. Then there is no nontrivial harmonic map $u: S^2\to(N,h)$ such that $u(S^2)\cap\overline{Q(2r)}=\varnothing$.*

By the definition of ψ, it is easy to check that

$$4\pi\psi(0)\leq 16\pi r^2<\tfrac{1}{3}\pi(\pi-2r)^2$$

when r is small enough. Using Lemma 3.2 and Lemma 3.3, we get the following result.

Corollary 3.4. *Let (N,h) and Q be the same as in Lemma 3.1. Assume that u is a nontrivial harmonic map with $E(u)<\pi(\pi-2r)^2$; then*

$$E(u)=4m\pi\psi(0)$$

where m is a positive integer.

It is easy to check that

$$12\pi\psi(0)<48\pi r^2<\pi(\pi-2r)^2,$$

if $r<\frac{\pi}{4\sqrt{3}+2}$. Therefore we know that if $E(u)<12\pi\psi(0)$ and u is a nontrivial harmonic map, then $E(u)=4\pi\psi(0)$ or $8\pi\psi(0)$.

3B. ***The homotopy class* $[u_k]$.** We have $\pi_1(Y_1)=\pi_1(\mathbb{T}^3)=\mathbb{Z}^3$. Let $\beta\in\pi_1(Y_1)$ which represents $(1,0,0)$. Let $x_1,x_2\in M$, and γ_0 be a curve in M such that $\gamma_0(0)=x_2$, and $\gamma_0(1)=x_1$. Let $\gamma_k:[0,1]\to X$ be a curve with $\gamma_k(0)=x_1,\gamma_k(1)=x_2$ and $[\gamma_k+\gamma_0]=k\beta$. Let w_0 be a diffeomorphism from S^2 onto M satisfying $w_0(0,0,1)=x_1$ and $w_0(0,0,-1)=x_2$, where $(0,0,1)$ and $(0,0,-1)$ are the north and the south poles of $S^2\subset\mathbb{R}^3$, respectively.

For the sake of convenience, we introduce the stereographic projection coordinates on S^2 with the south pole corresponding to ∞. Thus, $w_0 : S^2 \to N$ can be viewed as a map from $\mathbb{C}\cup\{\infty\}$ into N. For simplicity, we neglect the stereographic projection map $\mathfrak{S} : S^2 \to \mathbb{C}\cup\{\infty\}$ and still denote $w_0 \circ \mathfrak{S}^{-1}$ by w_0.

By the continuity of w_0, there exists a small $\delta_0 > 0$ such that $w_0(D_{\delta_0})$ is contained in a small neighborhood of x_1, where $D_{\delta_0} = \{z \in \mathbb{C} : |z| < \delta_0\}$, and a large $R_0 > 0$ such that $w_0(\mathbb{C}\setminus D_{R_0})$ is contained in a small neighborhood of x_2, where $D_{R_0} = \{z \in \mathbb{C} : |z| < R_0\}$.

In order to construct a sequence of maps, we need to define the following two smooth nonnegative functions λ and ν on $[0,\infty)$:

(1) $\lambda(s) : [0,\infty] \to [0,1]$ with $\lambda(s) \equiv 0$ as $s \in [0,\delta_0]$ and $\lambda(s) \equiv 1$ as $s \in [2\delta_0,\infty)$.

(2) $\nu(s) : [0,\infty) \to [0,1]$ with $\nu(s) \equiv 1$ as $s \in [0, R_0 - R_0^c]$ and $\nu(s) \equiv 0$ as $s > R_0$, where R_0^c is a small positive constant number.

Now we define a sequence of maps $u_k : S^2 \to N$ by

$$u_k = \begin{cases} w_0(\lambda(|z|)z) & |z| \geq \delta_0, \\ \gamma_k\left(\dfrac{\log|z| - \log R_0\epsilon_0}{\log\delta_0 - \log R_0\epsilon_0}\right) & R_0\epsilon_0 < |z| < \delta_0, \\ w_0\left(\dfrac{z}{\nu(|z|/\epsilon_0)\epsilon_0}\right) & |z| \leq \epsilon_0 R_0. \end{cases}$$

Here $\epsilon_0 > 0$ is a fixed constant number such that $R_0\epsilon_0 < \delta_0$. By the arguments in [Li and Wang 2010b], for any $i \neq j$, u_i is not homotopic to u_j. For the sequence $\{u_i\}$ constructed above, we have the following lemma:

Lemma 3.5. *Let u_k be the maps from S^2 into (N,h) constructed above and $[u_k]$ denote the class of maps in $W^{1,2}(S^2,N)\cap C(S^2,N)$, each map of which is homotopic to u_k. For any fixed k, we have*

$$\inf_{u\in[u_k]} E(u) = 8\pi\psi(0).$$

Moreover, $\inf E(u)$ *cannot be attained by a harmonic map belonging to* $[u_k]$.

Proof. First of all, we prove that for every fixed k

$$\inf_{u\in[u_k]} E(u) \leq 8\pi\psi(0). \tag{3-2}$$

Denote $z_1 = (0,0,1)$ and $z_2 = (0,0,-1) \in S^2$. Without loss of generality, we assume w_0 is a harmonic map from S^2 into M with $E(w_0) = 4\pi\psi(0)$ with $w_0(z_1) = x_1$ and $w_0(z_2) = x_2$. Let $\mathfrak{S}$ be the stereographic projection from $S^2\setminus\{z_2\}$ to $\mathbb{C}$ and

$$\hat{u}_0(z) = w_0(\mathfrak{S}^{-1}(z)) : \mathbb{C}\cup\{\infty\} \to N.$$

Choose a coordinate system (y_1, y_2, y_3) in a geodesic ball $B_\rho(x_1)$ around $x_1 \in N$ with $x_1 = (0,0,0)$ and $\{(y_1, y_2, 0) : (y_1, y_2, 0) \in B_\rho(x_1)\} \subset M$. By the continuity of w_0, there exists a small $\delta > 0$ such that $w_0(z) \in B_\rho(x_1)$ when $|z| < \delta$. We define

$$u_0' = \eta_1 \hat{u}_0,$$

where η_1 is a smooth nonnegative function which equals 1 outside $D_{2\delta}$, 0 on D_δ, and satisfies $|\nabla \eta_1| < \frac{C}{\delta}$. Here $D_{2\delta} \subset \mathbb{C}$ denotes the disk centered at the origin. Then we have

$$\int_{D_{2\delta}} |\nabla u_0'|^2 \, dx^2 \le 2 \int_{D_{2\delta}} \left(|\nabla \eta_1|^2 |\hat{u}_0|^2 + |\nabla \hat{u}_0|^2 \right) dx^2 \le C\delta.$$

Thus u_0' satisfies

$$\operatorname{dist}^M(u_0', \hat{u}_0) < C\delta, \quad E(u_0') < 4\pi\psi(0) + C\delta, \quad \text{and} \quad u_0'(D_\delta) = x_1.$$

Since E is conformally invariant, $\hat{u}_0(1/z)$ is also a harmonic map from $\mathbb{C}\setminus\{0\}$ into N with

$$E(\hat{u}_0(1/z), \mathbb{C}) = E(\hat{u}_0(z), \mathbb{C}).$$

Thus, $\hat{u}_0(1/z)$ can be extended smoothly to $\{0\}$. Choose a coordinate system (y_1, y_2, y_3) in a geodesic ball $B_\rho(x_2)$ around $x_2 \in N$ with $x_2 = (0,0,0)$ and $\{(y_1, y_2, 0) : (y_1, y_2, 0) \in B_\rho(x_2)\} \subset M$. By the continuity of w_0, there exists a large $R > 0$ such that $\hat{u}_0(z) \in B_\rho(x_2)$ as $|z| > R$. Then we have

$$\hat{u}_0(1/z) = O(z) \quad \text{and} \quad |\nabla \hat{u}_0(1/z)| = O(1), \quad \text{as } z \to 0.$$

Hence, we have

$$\hat{u}_0(z) = O(1/z) \quad \text{and} \quad |z^2 \nabla \hat{u}_0(z)| = O(1), \quad \text{as } z \to \infty.$$

Let

$$u_0''(z) = \eta_2(|z|)\hat{u}_0(z),$$

where $\eta_2(|z|)$ is a smooth nonnegative function which equals 0 outside D_R, 1 on $D_{R/2}$, and satisfies $|\nabla \eta_1| < \frac{C}{R}$. Then we have

$$\int_{\mathbb{C}\setminus D_{R/2}} |\nabla u_0''|^2 \, dx^2 \le 2 \int_{D_R \setminus D_{R/2}} \left(|\nabla \eta_2|^2 |\hat{u}_0|^2 + |\nabla \hat{u}_0|^2 \right) dx^2 \le \frac{C}{R}.$$

Thus

$$\operatorname{dist}^M(u_0'', u_0) < \frac{C}{R}, \quad E(u_0'') < 4\pi\psi(0) + \frac{C}{R}, \quad \text{and} \quad u_0''(\mathbb{C}\setminus D_R) = x_2.$$

We define

$$\phi_k = \begin{cases} u_0'(z), & |z| \geq \delta, \\ \gamma_k\left(\dfrac{\log|z| - \log R\epsilon}{\log\delta - \log R\epsilon}\right), & R\epsilon < |z| < \delta, \\ u_0''\left(\dfrac{z}{\epsilon}\right), & |z| \leq \epsilon R. \end{cases}$$

By a direct calculation, we obtain

$$\begin{aligned} \int_{D_\delta \setminus D_{R\epsilon}} |\nabla\phi_k|^2 &= 2\pi \int_{R\epsilon}^{\delta} \left|\frac{\partial\gamma_k}{\partial r}\right|^2 r\,dr \\ &< \frac{c\|\dot{\gamma}_k\|_{L^\infty}^2}{(-\log R\epsilon + \log\delta)^2} \int_{R\epsilon}^{\delta} \frac{dr}{r} = \frac{c\|\dot{\gamma}_k\|_{L^\infty}^2}{\log\delta - \log R\epsilon}. \end{aligned}$$

Thus, for any $\epsilon_1 > 0$, we can choose suitable δ, R and ϵ such that

$$E(\phi_k) < 8\pi\psi(0) + \epsilon_1.$$

Obviously, $\varphi_k = \phi_k(\mathfrak{S}^{-1})$ is homotopic to u_k, denoted by $\varphi_k \sim u_k$. Thus, we get (3-2).

Next, we prove that $\inf_{u\in[u_k]} E(u)$ cannot be attained by a harmonic map. Assume it is attained by a harmonic map v_0. Recall that

$$8\pi\psi(0) < 12\pi\psi(0) < 48\pi r^2 < \pi(\pi - 2r)^2,$$

where $r > 0$ is small enough. By Lemma 3.2, $v_0(S^2) \subset Q(\pi)$. Thus v_0 is a harmonic map from S^2 into M. This contradicts the fact $v_0 \sim u_k$. Hence $\inf_{u\in[u_k]} E(u)$ cannot be attained by a harmonic map.

Let u_α be the α-harmonic map such that, for fixed k,

$$E_\alpha(u_\alpha) = \inf_{u\in[u_k]\cap W^{1,2\alpha}(S^2,N)} E_\alpha(u).$$

Then each map of $\{u_\alpha\}$ is minimizing and belongs to $[u_k]$. We claim that $\{u_\alpha\}$ does not converge smoothly. Otherwise, the limit map is a harmonic map from S^2 into N, which is homotopic to u_k. This contradicts the above fact that $\inf_{u\in[u_k]} E(u)$ cannot be attained by a harmonic map. Hence, the bubbles must appear in the convergence of u_α. If we denote the weak limit of $\{u_\alpha\}$ as u_0 and the bubbles as $v^1, \dots, v^m$, then, by Theorem 2.3, we have

$$\inf_{u\in[u_k]} E(u) = \lim_{\alpha\to 1} E_\alpha(u_\alpha) = E(u_0) + \sum_{i=1}^{m} E(v^i).$$

Since $E(u_0)$ and $E(v^i)$ are smaller than $\pi(\pi - 2r)^2$, $E(u_0) + \sum_{i=1}^{m} E(v^i)$ can only equal $8\pi\psi(0)$ or $4\pi\psi(0)$.

Next, we will show that the following identity does not hold true:

$$E(u_0)+\sum_{i=1}^{m} E(v^i)=4\pi\psi(0).$$

If we assume this is true, then u_0 is trivial and u_α has only one bubble v^1. To derive a contradiction, we only need to prove $u_\alpha \sim v^1$.

Let $x_0 \in S^2$ be a blowup point. Take an isothermal coordinate system around x_0 with $x_0=(0,0)$ on $S^2=\mathbb{C}\cup\{\infty\}$. Let v^1 be the limit map of $u_\alpha(z_\alpha+\lambda_\alpha^1 z)$, where $z_\alpha \to 0$, $\lambda_\alpha^1 \to 0$. Then

$$v_\alpha^1(z)=u_\alpha(z_\alpha+\lambda_\alpha^1 z)$$

converges smoothly to v^1 on any $D_R=D_R(0)\subset\mathbb{C}$. Moreover, u_α converges smoothly in $\mathbb{C}\cup\{\infty\}\setminus D_{1/R}$ to a point $y_0\in N$. For us to prove $u_\alpha \sim v^1$, it is enough to check that for any $\epsilon>0$, there exists an $R>0$, such that

$$\sup_{t\in[R\lambda_\alpha^1,1/R]} \operatorname{osc}_{\partial D_t(z_\alpha)} u_\alpha<\epsilon.$$

Indeed, if this is not true then there exists a sequence of λ_α^2 with $\lambda_\alpha^2\to 0$ and $\lambda_\alpha^2/\lambda_\alpha^1\to+\infty$, such that

$$\operatorname{osc}_{\partial D_{\lambda_\alpha^2}(z_\alpha)} u_\alpha \to \epsilon_1 \neq 0.$$

Let

$$v_\alpha^2(z)=u_\alpha(z_\alpha+\lambda_\alpha^2 z).$$

If the sequence $\{v_\alpha^2\}$ has blowup points, then at each blowup point there exists at least a bubble of $\{v_\alpha^2\}$ which is also a bubble of u_α and is different from the previous bubble v^1. However, this is impossible since there only exists one bubble for $\{u_\alpha\}$. Hence, we infer that as $\alpha\to 1$, $\{v_\alpha^2\}$ converges smoothly on $D_{R'}\setminus D_{1/R'}\subset\mathbb{C}$ for any R'. It follows that

$$\operatorname{osc}_{\partial D_1} v_\alpha^2 \to \epsilon_1\neq 0.$$

This means that the limit map of $\{v_\alpha^2\}$ is not trivial and the limit map is also a bubble of $\{u_\alpha\}$ which is different from v^1. This is a contradiction. Thus, we conclude

$$\inf_{u\in[u_k]} E(u)=E(u_0)+\sum_{i=1}^{m}E(v^i)=8\pi\psi(0).$$

This completes the proof of the lemma. □

By the Sobolev embedding theorem, we know that for $\alpha>1$,

$$W^{1,2\alpha}(S^2,N)\subset C(S^2,N).$$

For simplicity, let $[u_k]^\alpha$ denote the class of maps belonging to $W^{1,2\alpha}(S^2, N)$, each map of which is homotopic to u_k. In fact, it is easy to see that

$$[u_k]^\alpha = [u_k] \cap W^{1,2\alpha}(S^2, N).$$

From now on, we will always denote the smooth map which attains $\inf_{u\in[u_k]^\alpha} E_\alpha(u)$ by $u_{\alpha,k}$:

$$E_\alpha(u_{\alpha,k}) = \inf_{u\in[u_k]^\alpha} E_\alpha(u).$$

Lemma 3.6. *For any $\lambda_0 > 8\pi\psi(0)$, there exists a sequence $\{\alpha_k\}$ with $\alpha_k \to 1$ and a sequence $\{i_k\}$ such that $E_{\alpha_k}(u_{\alpha_k,i_k}) = \lambda_0$ for every k.*

Proof. For $\alpha \in [1, \alpha_0)$ where $\alpha_0 - 1 > 0$ is small enough, we define the following function

$$\varphi_k(\alpha) = \inf_{u\in[u_k]^\alpha} E_\alpha(u).$$

Firstly, we need to show that for any fixed $\alpha \in (1, \alpha_0)$,

$$\lim_{k\to+\infty} \varphi_k(\alpha) = +\infty. \tag{3-3}$$

If this is false, then there exists a constant C such that $\varphi_k(\alpha) \le C$ as k is large enough. We note that for any small δ and $x \in S^2$,

$$E(u_{\alpha,k}, B_\delta(x)) = \frac{1}{2}\int_{B_\delta(x)} |\nabla u_{\alpha,k}|^2 \le \frac{1}{2}\left(\int_{B_\delta(x)} |\nabla u_{\alpha,k}|^{2\alpha}\right)^{1/\alpha} |B_\delta(x)|^{(\alpha-1)/\alpha}. \tag{3-4}$$

Hence, we can pick a fixed δ, which is small enough, such that

$$E(u_{\alpha,k}, B_\delta(x)) < \epsilon_0.$$

Thus, by Theorem 2.1, there exists a subsequence of $u_{\alpha,k}$ which converges smoothly to a smooth map u_0 as k tends to ∞. Hence, we know that $u_{\alpha,k}$ are homotopic to u_0 for any k. This contradicts the fact that $u_{\alpha,i}$ is not homotopic to $u_{\alpha,j}$ as $i \ne j$.

Next, we want to prove φ_k is continuous on $[1, \alpha_0)$. Using (3-4) again, we can prove that, for a fixed small $\epsilon > 0$,

$$\|\nabla u_{\alpha,k}\|_{C^0(S^2)} < \Lambda(\epsilon)$$

for any $\alpha \in (1+\epsilon, \alpha_0)$. For any $\alpha, \alpha' \in (1+\epsilon, \alpha_0)$, we have

$$\begin{aligned}\varphi_k(\alpha) &\ge \tfrac{1}{2}(1+C_1^2)^{\alpha-\alpha'}\int_{S^2}(1+|\nabla u_{\alpha,k}|^2)^{\alpha'} - \tfrac{1}{2}\\ &\ge (1+C_1^2)^{\alpha-\alpha'}\varphi_k(\alpha') + \tfrac{1}{2}(1+C_1^2)^{\alpha-\alpha'} - \tfrac{1}{2},\end{aligned}$$

where

$$C_1 = \begin{cases} 0 & \text{when } \alpha > \alpha', \\ \Lambda(\epsilon) & \text{when } \alpha < \alpha'. \end{cases}$$

It follows that

$$\varliminf_{\alpha \to \alpha'} \varphi_k(\alpha) \geq \varphi_k(\alpha').$$

On the other hand, we also have

$$\varphi_k(\alpha') \geq \tfrac{1}{2}(1 + C_2^2)^{\alpha' - \alpha} \int_{S^2} (1 + |\nabla u_{\alpha',k}|^2)^\alpha - \tfrac{1}{2},$$

where

$$C_2 = \begin{cases} 0 & \text{when } \alpha' > \alpha, \\ \|\nabla u_{\alpha' k}\|_{L^\infty} & \text{when } \alpha' < \alpha. \end{cases}$$

It follows that

$$\varphi_k(\alpha') \geq (1 + C_2^2)^{\alpha' - \alpha} \varphi_k(\alpha) + \tfrac{1}{2}(1 + C_2^2)^{\alpha' - \alpha} - \tfrac{1}{2},$$

and

$$\varlimsup_{\alpha \to \alpha'} \varphi_k(\alpha) \leq \varphi_k(\alpha').$$

Therefore, we have

$$\lim_{\alpha \to \alpha'} \varphi_k(\alpha) = \varphi_k(\alpha'),$$

and we have shown the continuity of $\varphi_k(\alpha)$ on $(1, \alpha_0)$.

Next, we want to prove that $\varphi_k(\alpha)$ is left continuous at 1. Equivalently, we need to show

$$\lim_{\alpha \searrow 1} \varphi_k(\alpha) = \varphi_k(1). \tag{3-5}$$

Obviously, for any fixed $u \in W^{1,2}(S^2, N)$ and $\alpha_1 > \alpha_2 > 1$,

$$E_{\alpha_1}(u) \geq E_{\alpha_2}(u) \geq E(u).$$

It follows that

$$\varphi_k(\alpha_1) \geq \varphi_k(\alpha_2) \geq \varphi_k(1).$$

Hence, $\lim_{\alpha \searrow 1} \varphi_k$ exists and

$$\lim_{\alpha \searrow 1} \varphi_k(\alpha) \geq \varphi_k(1).$$

On the other hand, note that u_k is a smooth map. Then for any $\epsilon > 0$, there exists a smooth map $u_k' \in C^\infty(S^2, N)$ which is homotopic to u_k (i.e., $u_k' \sim u_k$), and satisfies

$$E(u_k') \leq \varphi_k(1) + \epsilon.$$

Since

$$\lim_{\alpha\searrow 1} E_\alpha(u_k') = E(u_k') \quad \text{and} \quad \varphi_k(\alpha) \le E_\alpha(u_k'),$$

we have

$$\lim_{\alpha\searrow 1} \varphi_k(\alpha) \le \varphi_k(1) + \epsilon,$$

which implies (3-5), and shows that $\varphi_k(\alpha)$ is continuous on $[1, \alpha_0)$ for any fixed k.

By (3-3), for any given sequence $\{\alpha_k'\}$ with $\alpha_k' \to 1$, there exists a sequence $\{i_k\}$ such that $E_{\alpha_k'}(u_{\alpha_k', i_k}) > \lambda_0$, or equivalently, $\varphi_{i_k}(\alpha_k') > \lambda_0$. Lemma 3.5 tells us that $\varphi_{i_k}(1) = 8\pi\psi(0)$ for any i_k. By the assumption $\lambda_0 > 8\pi\psi(0)$ we have

$$\varphi_{i_k}(\alpha_k') > \lambda_0 > \varphi_{i_k}(1).$$

Since $\varphi_k(\alpha)$ is continuous on $[1, \alpha_0)$, we conclude that for any fixed i_k there exists $\alpha_k \in (1, \alpha_k')$ such that

$$\varphi_{i_k}(\alpha_k) = E_{\alpha_k}(u_{\alpha_k, i_k}) = \lambda_0.$$

This completes the proof. □

3C. ***The counterexample.*** By Lemma 3.6, for given $\tau \in (8\pi\psi(0), 12\pi\psi(0))$ there exist a sequence $\{\alpha_k : \alpha_k > 1, k \in \mathbb{N}\}$ with $\alpha_k \to 1$ and a sequence of minimizing α_k-harmonic maps $v_k \in W^{1,2\alpha_k}(S^2, N)$ with $v_k \sim u_{i_k}$ such that

$$\tau = E_{\alpha_k}(v_k) = \inf_{u \in [u_{i_k}]^{\alpha_k}} E_{\alpha_k}(u) \;\text{ for all } k \in \mathbb{N}.$$

Since v_i and v_j are not in the same homotopy class for any $i \neq j$, v_k must blow up as $k \to +\infty$. Let v^0 be the weak limit of $\{v_k\}$ in $W^{1,2}(S^2, N)$, and $v^1, \dots, v^m$ be all the bubbles produced in the convergence of $\{v_k\}$. Since $E(v^i) < 12\pi\psi(0)$, it follows from Corollary 3.4 that $E(v^i) = 4\pi\psi(0)$ or $8\pi\psi(0)$. Hence,

$$\frac{1}{4\pi\psi(0)}\left(E(v^0) + \sum_{i=1}^m E(v^i)\right)$$

is always an integer. However, certainly $\frac{\tau}{4\pi\psi(0)}$ is not an integer by the previous assumption. So the energy identity is not true for the sequence $\{v_k\}$:

$$\lim_{k\to\infty} E_{\alpha_k}(v_k) \neq E(v^0) + \sum_{i=1}^m E(v^i).$$

Remark 3.7. By an argument in [Li and Wang 2010b], we also have

$$\lim_{k\to\infty} E(v_k) \neq E(v^0) + \sum_{i=1}^m E(v^i).$$

4. An example of a manifold whose energy set is nondiscrete for harmonic 2-spheres

In this section, we will construct a Riemannian manifold (N, h) for which $\mathscr{E}(N, h)$ is not discrete. In other words, $\mathscr{E}(N, h)$ admits limit points.

Let $\psi(t)$ be a smooth positive function defined on $(-1, 1)$ satisfying

$$\psi(t) = e^{-1/t^2} \sin \tfrac{1}{t} + 1, \quad t \in (-\tfrac{1}{2}, \tfrac{1}{2}).$$

It is easy to check that the critical point of $\psi(t)$ satisfies the equation

$$\tan \tfrac{1}{t} = \tfrac{t}{2}.$$

Thus, we can find $t_k \to 0$, such that $\psi'(t_k) = 0$, $\psi(t_k) \neq 1$ and $\psi(t_k) \to 1$.

Let

$$h = \psi(t)(d\mathfrak{s}^2 + dt^2),$$

which is a metric over $S^2 \times (-1, 1)$. Let v be the identity map from S^2 to S^2 and

$$u_k = (v, t_k) : S^2 \to (N, h) \equiv (S^2 \times (-1, 1), h).$$

By (3-1), it is easy to see that u_k is a harmonic map from S^2 into $(S^2 \times (-1, 1), h)$ with

$$E(u_k) = 4\pi \psi(t_k).$$

Thus, 4π is not a discrete number in $\mathscr{E}(S^2 \times (-1, 1), h)$.

Acknowledgement

The authors are grateful to thank Professor Weiyue Ding for his encouragement and beneficial suggestions. The first author is supported by NSFC, Grant No. 11131007; the second author is supported by NSFC, Grant No. 10990013.

References

[Anand 1995] C. K. Anand, "Uniton bundles", *Comm. Anal. Geom.* **3**:3-4 (1995), 371–419. MR 96k:58052 Zbl 0848.58013

[Chen and Tian 1999] J. Chen and G. Tian, "Compactification of moduli space of harmonic mappings", *Comment. Math. Helv.* **74**:2 (1999), 201–237. MR 2001k:58024 Zbl 0958.53047

[Ding 1998] W.-Y. Ding, "Lectures on the heat flow of harmonic maps", preprint, National Tsing Hua University, Taiwan, 1998, Available at http://www.math.pku.edu.cn:8000/var/teacher_writings/20080327073216.pdf.

[Ding and Tian 1995] W.-Y. Ding and G. Tian, "Energy identity for a class of approximate harmonic maps from surfaces", *Comm. Anal. Geom.* **3**:3-4 (1995), 543–554. MR 97e:58055 Zbl 0855.58016

[Dong 2002] Y. Dong, "On energy gap of unitons", *Math. Z.* **240**:4 (2002), 677–688. MR 2003e:53081 Zbl 1015.58002

[Duzaar and Kuwert 1998] F. Duzaar and E. Kuwert, "Minimization of conformally invariant energies in homotopy classes", *Calc. Var. Partial Differential Equations* **6**:4 (1998), 285–313. MR 99d:58045 Zbl 0909.49008

[Hong and Yin 2010] M. Hong and H. Yin, "On the Sacks–Uhlenbeck flow of Riemannian surfaces", preprint, 2010. arXiv 1007.3052

[Jost 1991] J. Jost, *Two-dimensional geometric variational problems*, Wiley, Chichester, 1991. MR 92h:58045 Zbl 0729.49001

[Lamm 2010] T. Lamm, "Energy identity for approximations of harmonic maps from surfaces", *Trans. Amer. Math. Soc.* **362**:8 (2010), 4077–4097. MR 2011d:58036 Zbl 1200.58016

[Li and Wang 2010a] Y. Li and Y. Wang, "Bubbling location for sequences of approximate f-harmonic maps from surfaces", *Internat. J. Math.* **21**:4 (2010), 475–495. MR 2011g:58028 Zbl 1190.35067

[Li and Wang 2010b] Y. Li and Y. Wang, "A weak energy identity and the length of necks for a sequence of Sacks–Uhlenbeck α-harmonic maps", *Adv. Math.* **225**:3 (2010), 1134–1184. MR 2011j:58022 Zbl 1203.58003

[Qing 1995] J. Qing, "On singularities of the heat flow for harmonic maps from surfaces into spheres", *Comm. Anal. Geom.* **3**:1-2 (1995), 297–315. MR 97c:58154 Zbl 0868.58021

[Qing and Tian 1997] J. Qing and G. Tian, "Bubbling of the heat flows for harmonic maps from surfaces", *Comm. Pure Appl. Math.* **50**:4 (1997), 295–310. MR 98k:58070 Zbl 0879.58017

[Sacks and Uhlenbeck 1981] J. Sacks and K. Uhlenbeck, "The existence of minimal immersions of 2-spheres", *Ann. of Math.* (2) **113**:1 (1981), 1–24. MR 82f:58035 Zbl 0462.58014

[Sacks and Uhlenbeck 1982] J. Sacks and K. Uhlenbeck, "Minimal immersions of closed Riemann surfaces", *Trans. Amer. Math. Soc.* **271**:2 (1982), 639–652. MR 83i:58030 Zbl 0527.58008

[Simon 1983] L. Simon, *Lectures on geometric measure theory*, Proceedings of the Centre for Mathematical Analysis, Australian National University **3**, Australian National University, Centre for Mathematical Analysis, Canberra, 1983. MR 87a:49001 Zbl 0546.49019

[Uhlenbeck 1989] K. Uhlenbeck, "Harmonic maps into Lie groups: classical solutions of the chiral model", *J. Differential Geom.* **30**:1 (1989), 1–50. MR 90g:58028 Zbl 0677.58020

[Valli 1988] G. Valli, "On the energy spectrum of harmonic 2-spheres in unitary groups", *Topology* **27**:2 (1988), 129–136. MR 90f:58042 Zbl 0744.53027

Received September 29, 2013. Revised January 4, 2014.

YUXIANG LI
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING, 100084
CHINA

yxli@math.tsinghua.edu.cn

YOUDE WANG
ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCES
CHINESE ACADEMY OF SCIENCES
BEIJING, 100080
CHINA

wyd@math.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.125

THEORY OF NEWFORMS OF HALF-INTEGRAL WEIGHT

MURUGESAN MANICKAM, JABAN MEHER
AND BALAKRISHNAN RAMAKRISHNAN

We set up the theory of newforms of half-integral weight on $\Gamma_0(8N)$ and $\Gamma_0(16N)$, where N is odd and squarefree. Further, we extend the definition of the Kohnen plus space in general for trivial character and also study the theory of newforms in the plus spaces on $\Gamma_0(8N)$, $\Gamma_0(16N)$, where N is odd and squarefree. Finally, we show that the Atkin–Lehner W-operator W_4 acts as the identity operator on $S_{2k}^{\mathrm{new}}(4N)$, where N is odd and squarefree. This proves that $S_{2k}^{-}(4) = S_{2k}(4)$.

1. Introduction

Let k, M be positive integers, $k \geq 2$. Write $M = 2^{\alpha}N$, $\alpha \geq 0$, $N \geq 1$, N odd. Let χ_0 be a Dirichlet character modulo N with $\epsilon = \chi_0(-1)$ and let χ_1 be an even Dirichlet character modulo $2^{\alpha+2}$. Let $\chi = \left(\frac{4\epsilon}{\cdot}\right)\chi_1\chi_0$. Let $S_{k+1/2}(4M, \chi)$ be the space of cusp forms of half-integral weight $k + \frac{1}{2}$ for $\Gamma_0(4M)$ with character χ, and let $S_{2k}(2M, \chi^2)$ be the space of cusp forms of weight $2k$, level $2M$ with character χ^2. By the work of G. Shimura [1973] and S. Niwa [1975], there exist linear operators $\mathcal{S}_{t,4M,\chi}$ indexed by squarefree integers t, $\epsilon(-1)^k t > 0$, which commute with the action of Hecke operators $T(n^2)$, $(n, 2M) = 1$, and map the space $S_{k+1/2}(4M, \chi)$ into the space $S_{2k}(2M, \chi^2)$. If M is an odd integer, W. Kohnen [1980; 1982] introduced a canonical subspace $S^{+}_{k+1/2}(4M, \chi)$, called the Kohnen plus space, in the full space $S_{k+1/2}(4M, \chi)$. He defined modified Shimura lifts $\mathcal{S}^{+}_{D,4M,\chi}$, called Shimura–Kohnen lifts, indexed by fundamental discriminants D, $\epsilon(-1)^k D > 0$, which commute with the action of Hecke operators $T(n^2)$, $(n, M) = 1$, where $T(4) = T^{+}(4)$ is the Hecke operator introduced by Kohnen on the plus space. He proved that the linear operator $\mathcal{S}^{+}_{D,4M,\chi}$ maps the space $S^{+}_{k+1/2}(4M, \chi)$ into the space $S_{2k}(M, \chi^2)$. The idea of characterising the spaces of half-integral weight forms Hecke-equivalent to a fixed integral weight newform is important, and establishing Hecke equivariant isomorphisms via trace identities is certainly a powerful tool. These isomorphisms often give hints as to how to further decompose these eigenspaces to obtain multiplicity-one results. The first such work

MSC2010: primary 11F37; secondary 11F11, 11F32.
Keywords: modular forms of half-integral weight, theory of newforms, Kohnen plus space.

was by Kohnen, who achieved that goal by introducing the plus space, which we now wish to generalise. Kohnen [1980; 1982] initiated the study of the theory of newforms for the plus space $S^{+}_{k+1/2}(4M,\chi)$ along the lines of Atkin and Lehner [1970], where M is odd and squarefree and $\chi^2=1$. Using the trace identities proved by Niwa [1977], M. Manickam, B. Ramakrishnan and T. C. Vasudevan [Manickam et al. 1990] set up the theory of newforms for the full space $S_{k+1/2}(4M,\chi)$, where M is odd and squarefree and $\chi^2=1$. If M is even and squarefree, this theory is known on the full space $S_{k+1/2}(4M,\chi)$ by the work of Manickam [1980; 2011]. For similar theories we refer to [Serre and Stark 1977; Shemanske 1996; Ueda 1988; 1991; 1993; 1998; 2001].

Kohnen introduced the plus space in $S_{k+1/2}(4M,\chi)$ when M is odd by letting

$$S^{+}_{k+1/2}(4M,\chi)=\{f\in S_{k+1/2}(4M,\chi): a_f(n)=0 \text{ unless } \epsilon(-1)^k n\equiv 0,1 \pmod 4\}.$$

Ueda and Yamana [2010] extended the definition of the plus space for $S_{k+1/2}(4M)$ (M is even and squarefree) by using the same condition on the Fourier coefficients. If M is even, let

$$S^{+}_{k+1/2}(4M)=\{f\in S_{k+1/2}(4M): a_f(n)=0 \text{ unless } (-1)^k n\equiv 0,1 \pmod 4\}.$$

In the case where M is odd, the Kohnen plus space $S^{+}_{k+1/2}(4M)$ is an eigensubspace of $S_{k+1/2}(4M)$ under a hermitian operator $U(4)W(4)$ [Kohnen 1982; Manickam et al. 1990], whereas in all other cases it is the image of the projection operator $\mathcal{P}_+$ on $S_{k+1/2}(4M)$ (M even) given by

$$\mathcal{P}_+ : \sum_{n\geq 0} a(n)q^n \longrightarrow \sum_{\substack{n\geq 0\\ (-1)^k n\equiv 0,1 \ (\mathrm{mod}\ 4)}} a(n)q^n.$$

This operator $\mathcal{P}_+$ was introduced by Kohnen and considered by Ueda and Yamana [2010]. If M is even, $\mathcal{P}_+$ preserves the space $S_{k+1/2}(4M)$. This phenomenon is striking and it allows us to define the plus space for an even integer M by

$$S^{+}_{k+1/2}(4M)=S_{k+1/2}(4M)|\mathcal{P}_+.$$

In this paper we generalise the theory of newforms for the Kohnen plus space and the full space whenever the traces of Hecke operators acting on the spaces of integral and half-integral weight modular forms are equal. We also consider the space $S_{k+1/2}(16N)$, N odd and squarefree, and develop the theory of newforms by computing the dimension, since Ueda's trace formula is known for the case where the character of the space is nontrivial. In this case, we prove that the newform spaces $S^{\mathrm{new}}_{k+1/2}(16N)$ and $S^{+,\mathrm{new}}_{k+1/2}(16N)$ contain only the zero function.

Let us now explain the results of this paper. Let $M=2^\alpha N$, $\alpha=1,2$, N odd and squarefree, $\chi^2=1$ and $\chi=\chi_8$ when $\alpha=2$, where χ_8 is the real quadratic primitive even character modulo 8 defined by $\chi_8(n) = \left(\frac{2}{n}\right)$. Then there is a Hecke-equivariant isomorphism [Ueda 1988]

$$\psi : S_{k+1/2}(4M, \chi) \longrightarrow S_{2k}(2M).$$

We define the space of newforms in the full space as

$$S^{\text{new}}_{k+1/2}(4M, \chi) = \bigoplus_F S^{\text{new}}_{k+1/2}(4M, \chi; F),$$

where the sum varies over an orthogonal basis of normalised Hecke eigenforms in $S^{\text{new}}_{2k}(2M)$, and for each such F let

$$S^{\text{new}}_{k+1/2}(4M; F) = \{f \in S_{k+1/2}(4M, \chi) : f|T(n^2) = a_F(n)f, \forall n \geq 1, (n, 2M) = 1\}.$$

Then, $S^{\text{new}}_{k+1/2}(4M, \chi)$ is the inverse image of $S^{\text{new}}_{2k}(2M)$ under the isomorphism ψ, so the "multiplicity-one" result is valid for $S^{\text{new}}_{k+1/2}(4M, \chi)$.

Consider the plus space $S^{+}_{k+1/2}(8N)$. Since $\mathcal{P}_+$ preserves the space $S_{k+1/2}(8N)$ and $\mathcal{P}_+ T(n^2) = T(n^2)\mathcal{P}_+$, $(n, 2N) = 1$, we define $S^{+,\text{new}}_{k+1/2}(8N) = S^{\text{new}}_{k+1/2}(8N)|\mathcal{P}_+$, and as such the plus space $S^{+,\text{new}}_{k+1/2}(8N)$ is a subspace of $S^{\text{new}}_{k+1/2}(8N)$. For a nonzero Hecke eigenform $f \in S^{\text{new}}_{k+1/2}(8N; F)$, the form $f|\mathcal{P}_+$ is also a nonzero Hecke eigenform belonging to the same space having the same eigenvalues (for almost all Hecke operators) as that of f. Since N is odd and squarefree, a multiplicity-one result holds for the space $S^{\text{new}}_{k+1/2}(8N)$ and hence $f|\mathcal{P}_+ = f$. This proves the equality $S^{+,\text{new}}_{k+1/2}(8N) = S^{\text{new}}_{k+1/2}(8N)$. To get $f|\mathcal{P}_+ \neq 0$, we use the multiplicity-one result along with the fact that $F|\mathcal{S}^*_t \neq 0$ for some squarefree integer $t \equiv 1 \pmod 4$, $(-1)^k t > 0$. Here $\mathcal{S}^*_t$ is the Shintani lifting, which is the adjoint of the Shimura map $\mathcal{S}_t$ with respect to the Petersson scalar product ($\mathcal{S}_t$ maps $S_{k+1/2}(8N)$ into $S_{2k}(4N)$) — see [Manickam et al. 1989; Shintani 1975]. The nonvanishing of $F|\mathcal{S}^*_t$ follows from the fact that the $|t|$-th Fourier coefficient of $F|\mathcal{S}^*_t$ is (up to a nonzero constant) equal to the special value $L(F, t, k)$ and, for some choice of squarefree integer t, $(t, 2N) = 1$, this special value is nonzero — see [Murty and Murty 1997]. Thus, we get $F|\mathcal{S}^*_t|\mathcal{P}_+ \neq 0$, since $t \equiv 1 \pmod 4$.

Now, we let $M = 4N$ and χ be trivial. Through the dimension formula we observe that $S^{\text{new}}_{k+1/2}(16N) = S^{+,\text{new}}_{k+1/2}(16N) = \{0\}$. Further, we develop the theory of newforms on $S_{k+1/2}(16N, \chi)$, where χ is trivial or $\chi = \chi_8$. Thus, in this paper we consider the above assumptions on M:

$$M = \begin{cases} 2N & \chi \text{ trivial}, \\ 4N & \chi \text{ trivial or } \chi = \chi_8, \end{cases}$$

with N odd and squarefree, and set up the theory of newforms. We observe that the Shimura–Kohnen lifts map the space $S^{+,\text{new}}_{k+1/2}(8N)$ into the space $S^{\text{new}}_{2k}(4N)$ instead of $S^{\text{new}}_{2k}(2N)$.

Finally, as an application of the theory of newforms of half-integral weight, we get explicit eigenvalues for the W-operators on $S_{2k}(2M)$ (see [Gun et al. 2010], for example). More precisely, if $M=2N$ or $4N$ (N odd and squarefree), and if $F \in S_{2k}^{\text{new}}(2M)$ is a normalised newform with associated newform $f \in S_{k+1/2}^{\text{new}}(4M, \chi)$ ($8|\operatorname{cond}\chi$ if $M=4N$), then we have

$$f|w_p = \left(\frac{D}{p}\right) f$$

for all $p|N$, where D is a fundamental discriminant, $(-1)^k D > 0$, $(D, M) = 1$ with $a_f(|D|) \neq 0$. To get this, we use $f|w_p = \lambda_p f$ and the explicit Fourier expansion of $f|w_p$ (see [Kohnen 1982]). Thus, for $p|N$, $F|W_p = \left(\frac{D}{p}\right) F$. Now,

$$L^*(F, D, s) := \left(\frac{2\pi}{\sqrt{2M}|D|}\right)^{-s} \Gamma(s) L(F, D, s)$$

satisfies

$$L^*(F, D, 2k-s) = \left(\frac{D}{2M}\right) \lambda_{2M} L^*(F, D, s), \quad \left(\frac{D}{-1}\right) = (-1)^k,$$

where λ_{2M} is the product of eigenvalues of the various W-operators W_{p^β},

$$\beta = \begin{cases} \alpha+1 & \text{if } p=2, \\ 1 & \text{otherwise.} \end{cases}$$

Using $\lambda_p = \left(\frac{D}{p}\right)$ for all primes $p|N$ in the above functional equation, we get $\left(\frac{D}{2^\beta}\right) \cdot \lambda_{2^\beta} = 1$, since $L(F, D, k)$ is nonzero for some fundamental discriminant D, $(D, 2N) = 1$. From this we conclude that the eigenvalue of the W-operator W_{2^β} on $S_{2k}^{\text{new}}(2M)$ is equal to 1 when β is even. This proves that $S_{2k}(4) = S_{2k}^{-}(4)$, where

$$S_{2k}^{-}(m) = \left\{ f \in S_{2k}(m) : f \left| \begin{pmatrix} 0 & -1 \\ m & 0 \end{pmatrix} \right. = f \right\}.$$

The above subspace was introduced by Skoruppa and Zagier [1988] in connection with the theory of newforms for the space of Jacobi cusp forms.

2. Preliminaries

We begin by recalling some basic facts regarding modular forms of half-integral weight. Let $\mathcal{H}$ denote the upper half-plane consisting of complex numbers $\tau \in \mathbb{C}$ with $\operatorname{Im}(\tau) > 0$. For complex numbers $z \neq 0$, x, we let $z^x = e^{x \log z}$, $\log z = \log|z| + i \arg z$, $-\pi < \arg z \leq \pi$. Let ζ be a fourth root of unity. Let G denote the four-sheeted covering of $GL_2^+(\mathbb{Q})$ defined as the set of all ordered pairs $(\alpha, \phi(\tau))$, where $\phi(\tau)$ is a holomorphic function on $\mathcal{H}$ such that $\phi^2(\tau) = \zeta^2(c\tau+d)/\sqrt{\det\alpha}$ and $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in GL_2^+(\mathbb{Q})$. Then G is a group with multiplication $(\alpha, \phi(\tau))(\beta, \psi(\tau)) = (\alpha\beta, \phi(\beta\tau)\psi(\tau))$. Let $k \geq 2$ be a natural number. For a complex valued function f defined on the upper half-plane $\mathcal{H}$ and an element $(\alpha, \phi(\tau)) \in G$, define the stroke operator by $f|_{k+1/2}(\alpha,\phi(\tau))(\tau)=\phi(\tau)^{-2k-1}f(\alpha\tau)$. We omit the subscript $k+\frac{1}{2}$ wherever there is no ambiguity. For $\Gamma_0(4)$ and its subgroups, we take the lifting $\Gamma_0(4)\to G$ as the collection $\{(\alpha, j(\alpha,\tau))\}$, where

$$\alpha=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\Gamma_0(4)\quad\text{and}\quad j(\alpha,\tau)=\Big(\frac{c}{d}\Big)\Big(\frac{-4}{d}\Big)^{-1/2}(c\tau+d)^{1/2}.$$

Here $\big(\frac{c}{d}\big)$ denotes the generalised quadratic residue symbol and $\big(\frac{-4}{d}\big)^{1/2}$ is equal to 1 or i according as d is 1 or 3 modulo 4. Let M be a natural number. A holomorphic function $f:\mathcal{H}\to\mathbb{C}$ is called a modular form of weight $k+\frac{1}{2}$ for $\Gamma_0(4M)$ with character χ (modulo $4M$) if $f|_{k+1/2}(\gamma, j(\gamma,\tau))(\tau)=\chi(d)f(\tau)$ for all $\gamma=\big(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\big)\in\Gamma_0(4M)$ and f is holomorphic at all the cusps of $\Gamma_0(4M)$. If, further, it vanishes at all the cusps, then it is called a cusp form. The set of cusp forms defined as above forms a complex vector space denoted by $S_{k+1/2}(4M,\chi)$. If χ is the trivial character, then the space is denoted by $S_{k+1/2}(4M)$. We also denote by $S_k(M)$ the space of cusp forms of weight k on $\Gamma_0(M)$ with trivial character. The Fourier expansion of a cusp form f at the cusp infinity is usually written as $f(\tau)=\sum_{n\ge 1}a_f(n)q^n$, where $q=e^{2\pi i\tau}$. For a prime p, the p-th Hecke operator on $S_{k+1/2}(4M)$ is denoted by $T(p^2)$ if $p\nmid 2M$ and $U(p^2)$ if $p\,|\,2M$; and on $S_{2k}(M)$ is denoted by $T(p)$ if $p\nmid M$ and $U(p)$ if $p|M$. By a Hecke eigenform in $S_{k+1/2}(4M,\chi)$, we mean a nonzero form in the space which is a simultaneous eigenform for all Hecke operators $T(n^2)$, $(n,2M)=1$. For any positive integer n, the operators $U(n)$ and $B(n)$ are defined on formal sums by $U(n):\sum_{m\ge 1}a(m)q^m\mapsto\sum_{m\ge 1}a(mn)q^m$, $B(n):\sum_{m\ge 1}a(m)q^m\mapsto\sum_{m\ge 1}a(m)q^{nm}$. The Petersson inner product for forms $f,g\in S_{k+1/2}(4M)$ is defined by

$$\langle f,g\rangle=\frac{1}{i_{4M}}\int_{\mathcal{F}}f(\tau)\overline{g(\tau)}v^{k-3/2}\,du\,dv,\tag{1}$$

where $\mathcal{F}$ is a fundamental domain for the action of $\Gamma_0(4M)$ on $\mathcal{H}$, i_{4M} is the index of $\Gamma_0(4M)$ in $\mathrm{SL}_2(\mathbb{Z})$ and $\tau=u+iv$.

2.1. ***Shimura and Shintani liftings.*** Let t be a squarefree integer with $(-1)^k t>0$. Then the t-th Shimura map on the space $S_{k+1/2}(4M)$ is defined by

$$f|\mathcal{S}_t=\sum_{n\ge 1}\bigg(\sum_{\substack{d|n\\(d,2M)=1}}\Big(\frac{4t}{d}\Big)d^{k-1}a_f(|t|n^2/d^2)\bigg)q^n.\tag{2}$$

We summarise the Shintani lifting [Manickam et al. 1989] when $M=2^\alpha N$, N is odd and $\alpha\ge 1$. If t is a squarefree integer, $(-1)^k t>0$, then for $F\in S_{2k}(2M)$ we have $F|\mathcal{S}_t^*\in S_{k+1/2}(4M)$ and it is given by

(3)

$$F|\mathcal{S}_t^*=(-1)^{[k/2]}2^{k-1+(a+1)(-k+1/2)}\sum_{m\ge 1}\bigg(\sum_{r|N}\mu(r)\Big(\frac{t}{r}\Big)r^{-k}r_{k,2Mr}(F;\Delta mr^2)\bigg)q^m,$$

where $r_{k,2M}(F;\,\Delta m)$ is a certain cycle integral given by

$$r_{k,2M}(F;\,\Delta m) = \sum \omega_t(Q) \int_{C_Q} F(z)(az^2+bz+c)^{k-1}\,dz. \tag{4}$$

In the above, the sum is over all $\Gamma_0(2M)$-equivalent quadratic forms $Q=[a,b,c]$ with discriminant $b^2-4ac=\Delta m$, $\Delta = 4^{\alpha+1}|t|$ and $a \equiv 0 \pmod{2^{2\alpha+1}N}$; C_Q is the image in $\Gamma_0(2M)\backslash\mathscr{H}$ of the semicircle $a|z|^2+b\,\Re(z)+c=0$ oriented from $(-b-\sqrt{\Delta}m)/2a$ to $(-b+\sqrt{\Delta}m)/2a$ if $a\neq 0$, or of the vertical line $b\,\Re(z)+c=0$ oriented from $-c/b$ to $i\infty$ if $b>0$ and from $i\infty$ to $-c/b$ if $b<0$, $a=0$.

Let us compute $r_{k,2M}(F;\,\Delta|t|)$. Since $\Delta|t| = 4^{\alpha+1}t^2$, we take the representatives $\{[0,\,2^{\alpha+1}|t|,\,\mu]\circ W_r : \mu \pmod{2^{\alpha+1}|t|},\, r|2M,\, r>0\}$, where W_r is the Atkin–Lehner W-operator. Note that $\omega_t(Q_\mu\circ W_r) = \left(\frac{t}{r}\right)\omega_t(Q_\mu) = \left(\frac{t}{r}\right)\left(\frac{t}{\mu}\right)$. Now, following the arguments in [Kohnen 1985, p. 243] we get

$$r_{k,2M}(F;\,4^{\alpha+1}t^2) = 2^{\nu(2M)}(-1)^{[k/2]}(2\pi)^{-k}\Gamma(k)(2^{\alpha+1}|t|)^{k-1/2}L(F,t,k), \tag{5}$$

where $\nu(2M)$ is the number of prime factors of $2M$. From this we get that, when F is a newform, the $|t|$-th Fourier coefficient of $F|\mathscr{S}_t^*$ is (up to a nonzero constant) the special value $L(F,t,k)$.

2.2. ***W-operators and the projection operator*** $\mathscr{P}_+$**.** For $p|2N$, let W_p denote the Atkin–Lehner W-operator on $S_{2k}(2N)$. For $p=2$, we define the analogous Atkin–Lehner W-operators $W(4)$ on $S_{k+1/2}(4N)$ and $W(8)$ on $S_{k+1/2}(8N)$ as follows:

$$W(4) = \left(\begin{pmatrix} 4a & b \\ 4Nc & 4 \end{pmatrix},\, 2^{1/2}e^{i\pi/4}(Nc\tau+1)^{1/2}\right),$$

where a, b, c are integers satisfying $4a-Nbc=1$ and $b\equiv 1 \pmod 4$;

$$W(8) = \left(\begin{pmatrix} 8x & y \\ 8Nw & 8 \end{pmatrix},\, 8^{1/4}e^{i\pi/4}(Nw\tau+1)^{1/2}\right), \tag{6}$$

where x, y, w are integers such that $y\equiv 1 \pmod 8$, $8x-Nwy=1$. We also let

$$W_*(4) = \left(\begin{pmatrix} 4u & v \\ 4Nr & 8 \end{pmatrix},\, 2^{1/2}e^{i\pi/4}(Nr\tau+2)^{1/2}\right),$$

where r, u, v are integers satisfying $8u-Nrv=1$ and $v\equiv 1 \pmod 8$.

Remark 2.1. The W-operators defined above are independent of the choice of the integers a, b, c, x, y, w, r, u, v with the given conditions. We note that $W_*(4)=W(4)$ on $S_{k+1/2}(4N)$; see [Manickam 1980; 2011] for details. The operator $W(8)$ maps $S_{k+1/2}(8N)$ into $S_{k+1/2}(8N,\,\chi_8)$, and $W(8)^2=I$ on $S_{k+1/2}(8N,\,\chi)$, where χ is the principal character or $\chi=\chi_8$ and I denotes the identity operator.

We now define the projection operator $\mathcal{P}_+$ on $S_{k+1/2}(4M)$ when M is even. Let $\xi = \left(\left(\begin{smallmatrix}4 & 1\\ 0 & 4\end{smallmatrix}\right), e^{\pi i/4}\right)$ and $\xi' = \left(\left(\begin{smallmatrix}4 & -1\\ 0 & 4\end{smallmatrix}\right), e^{-\pi i/4}\right)$. Then a formal computation shows that ξ (and hence ξ') preserves the space $S_{k+1/2}(4M)$ if $4|M$. Hence, if $4|M$, we have

$$\xi + \xi' : S_{k+1/2}(4M) \longrightarrow S_{k+1/2}(4M). \tag{7}$$

However, in the following we prove the above property for any even integer M. Let $M = 2N$, where N is an odd positive integer. We write

$$\begin{aligned}\xi + \xi' &= \xi + \begin{pmatrix}1-2N & (N-1)/2\\ 8N & 1-2N\end{pmatrix}^* \xi \begin{pmatrix}1 & 0\\ -8N & 1\end{pmatrix}^*\\ &= \xi + \xi \begin{pmatrix}1 & 0\\ -8N & 1\end{pmatrix}^*\\ &= \xi \,\mathrm{Tr} \quad \text{on } S_{k+1/2}(8N),\end{aligned}$$

where $\mathrm{Tr} = \sum_{\nu=0,1}\left(\begin{smallmatrix}1 & 0\\ -4N\nu & 1\end{smallmatrix}\right)^*$ is adjoint to the inclusion $S_{k+1/2}(8N) \hookrightarrow S_{k+1/2}(16N)$ with respect to the Petersson scalar product. On formal Fourier series $\sum a_n q^n$, we have

$$\sum a_n q^n |(\xi + \xi') = \chi_8(2k+1)\sqrt{2}\Biggl(\sum_{(-1)^k n \equiv 0,1 \ (\mathrm{mod}\ 4)} a_n q^n - \sum_{(-1)^k n \equiv 2,3 \ (\mathrm{mod}\ 4)} a_n q^n\Biggr). \tag{8}$$

We define

$$\mathcal{P}_+ := \frac{1}{2}\left(\frac{\chi_8(2k+1)}{\sqrt{2}}(\xi + \xi') + I\right). \tag{9}$$

Then

$$f|\mathcal{P}_+ = \sum_{(-1)^k n \equiv 0,1 \ (\mathrm{mod}\ 4)} a_f(n) q^n \in S_{k+1/2}(4M),$$

where $f = \sum_{n \geq 1} a_f(n) q^n \in S_{k+1/2}(4M)$.

3. Newforms on the plus space $S^+_{k+1/2}(8N)$

In the recent work of Ueda and Yamana [2010], the plus space for $S_{k+1/2}(8N)$ has been introduced and they studied the theory of newforms. In this case each newform in the full space $S^{\mathrm{new}}_{k+1/2}(8N)$ (see [Manickam 1980; 2011]) satisfies $f|\mathcal{P}_+ = f$. This follows by using that $\mathcal{P}_+$ maps $S^{\mathrm{new}}_{k+1/2}(8N)$ into itself and the multiplicity-one result obtained from Ueda's trace formula, together with the nonvanishing of $F|\mathcal{S}^*_t$ for some squarefree $t \equiv 1 \pmod 4$, where $F \in S^{\mathrm{new}}_{2k}(4N)$ is a normalised newform equivalent to f. Hence, the elements of $S^{\mathrm{new}}_{k+1/2}(8N)$ also satisfies the same plus space condition. Therefore, we consider the development of the theory of newforms on $S^+_{k+1/2}(8N)$ and present the results in this section.

Let us first state the results for the full space $S_{k+1/2}(8N)$, where N is odd and squarefree. The following orthogonal decomposition of $S_{k+1/2}(8N)$ has been obtained in [Manickam 1980; 2011]:

$$S_{k+1/2}(8N) = S^{\text{new}}_{k+1/2}(8N) \oplus S^{\text{old}}_{k+1/2}(8N), \tag{10}$$

where $S^{\text{new}}_{k+1/2}(8N) = S^{+,\text{new}}_{k+1/2}(8N)$ and the space of oldforms $S^{\text{old}}_{k+1/2}(8N)$ has the decomposition

$$\begin{aligned} S^{\text{old}}_{k+1/2}(8N) = \bigoplus_{rd|N,d<N} S^{+,\text{new}}_{k+1/2}(8d)|U(r^2) \oplus \bigoplus_{rd|N} S^{\text{new}}_{k+1/2}(4d)|U(r^2) \\ \oplus \bigoplus_{rd|N} S^{\text{new}}_{k+1/2}(4d)U(r^2)|\mathcal{P}_+ \oplus \bigoplus_{\substack{d|N \\ rd|2N}} S^{+,\text{new}}_{k+1/2}(4d)|U(r^2) \\ \oplus \bigoplus_{rd|N} S^{+,\text{new}}_{k+1/2}(4d)|U(4r^2)|\mathcal{P}_+. \end{aligned} \tag{11}$$

We need to show only that, for a fixed divisor $d|N$, the sum

$$S^{+,\text{new}}_{k+1/2}(4d) + S^{+,\text{new}}_{k+1/2}(4d)|U(4) + S^{+,\text{new}}_{k+1/2}(4d)|U(4)\mathcal{P}_+$$

is direct. For some constants α, β, γ and a newform $f \in S^{+,\text{new}}_{k+1/2}(4d)$, if we have

$$\alpha f + \beta f|U(4) + \gamma f|U(4)\mathcal{P}_+ = 0,$$

then, applying the operator $U(4)$ we get

$$\alpha f|U(4) = -(\beta+\gamma) f|U(16),$$

from which we conclude that $\alpha = 0$. Since $S^{+,\text{new}}_{k+1/2}(4d)|U(4) \oplus S^{+,\text{new}}_{k+1/2}(4d)|U(4)\mathcal{P}_+$ is a direct sum, it follows that $\beta = \gamma = 0$. This proves the required direct sum.

Thus, we get the following theorem regarding the plus space $S^{+}_{k+1/2}(8N)$:

Theorem 3.1. *The plus space $S^{+}_{k+1/2}(8N)$ has the orthogonal decomposition*

$$S^{+}_{k+1/2}(8N) = S^{+,\text{new}}_{k+1/2}(8N) \oplus S^{+,\text{old}}_{k+1/2}(8N),$$

where

$$\begin{aligned} S^{+,\text{old}}_{k+1/2}(8N) = \bigoplus_{rd|N,d<N} S^{+,\text{new}}_{k+1/2}(8d)|U(r^2) \oplus \bigoplus_{rd|N} S^{\text{new}}_{k+1/2}(4d)U(r^2)|\mathcal{P}_+ \\ \oplus \bigoplus_{rd|N} S^{+,\text{new}}_{k+1/2}(4d)|U(4r^2)|\mathcal{P}_+. \end{aligned} \tag{12}$$

The spaces $S^{+,\text{new}}_{k+1/2}(8N)$ and $S^{+,\text{old}}_{k+1/2}(8N)$ are mapped into the spaces $S^{\text{new}}_{2k}(4N)$ and $S^{\text{old}}_{2k}(4N)$ respectively under the Shimura lifting. Moreover, the spaces of newforms $S^{+,\text{new}}_{k+1/2}(8N)$ and $S^{\text{new}}_{2k}(4N)$ are isomorphic under a linear combination of Shimura lifts indexed by squarefree integers $t \equiv 1 \pmod 4$, $(-1)^k t > 0$.

Remark 3.2. If $f \in S^{+,\mathrm{new}}_{k+1/2}(8N) = S^{\mathrm{new}}_{k+1/2}(8N)$, then $a_f(n) = 0$ whenever $(-1)^k n$ is not congruent to 1 modulo 4. Hence, the Shimura maps $\mathcal{S}_{t,8N}$ annihilate $S^{\mathrm{new}}_{k+1/2}(8N)$ whenever $t \not\equiv 1 \pmod 4$, $(-1)^k t > 0$.

4. Newform theory on $S_{k+1/2}(16N)$

In this section, we extend the theory of newforms to the space $S_{k+1/2}(16N)$, where N is odd and squarefree. In this case, Ueda's trace formula is not valid as cond $\chi = 1$. Also from the work of Manickam, Ramakrishnan and Vasudevan [Manickam et al. 1989] on the Shintani lifting, it seems that there exists no Shintani lift from $S^{\mathrm{new}}_{2k}(8N)$ to $S_{k+1/2}(16N)$. But, such a lifting exists if we replace the trivial character by a primitive character modulo 8 or 16 (see [Manickam et al. 1989]). This indicates the nonexistence of a nontrivial space of newforms in $S_{k+1/2}(16N)$, which is mapped to $S^{\mathrm{new}}_{2k}(8N)$ under the Shimura lifting. To realise this, we compute the dimension of the space $S_{k+1/2}(16N)$ and give a decomposition of the space of oldforms (which turns out to be the full space).

Let us now compute the dimensions of the spaces $S_{2k}(4N)$ and $S_{k+1/2}(16N)$. Using [Martin 2005], we have

$$\dim S_{2k}(4N) = \frac{2k-1}{12} 4N \prod_{p \mid 2N}\Big(1+\frac{1}{p}\Big) - \frac{3}{2}2^{\nu(N)} = \frac{(2k-1)}{2}\prod_{p\mid N}(p+1) - 3\cdot 2^{\nu(N)-1}, \tag{13}$$

where $\nu(N)$ is the number of prime factors of N. Now, using [Cohen and Oesterlé 1977], we get

$$\dim S_{k+1/2}(16N) = \frac{2k-1}{24}16N\prod_{p\mid 2N}\Big(1+\frac{1}{p}\Big) - \frac{\zeta(k,16N,1)}{2}\prod_{p\mid N}\lambda(r_p,s_p,p) = (2k-1)\prod_{p\mid N}(p+1) - 3\cdot 2^{\nu(N)}. \tag{14}$$

(In the above we have used the dimension formula as given in [Ono 2004, Theorem 1.56, p. 16].) Equations (13), (14) imply that $\dim S_{k+1/2}(16N) = 2\dim S_{2k}(4N)$.

We now state the main theorem of this section.

Theorem 4.1. *We have*

$$S^{\mathrm{new}}_{k+1/2}(16N) = \{0\} \tag{15}$$

and

$$
\begin{aligned}
(16)\quad & S_{k+1/2}(16N) \\
&= \bigoplus_{rd|N} \big(S^{+,\mathrm{new}}_{k+1/2}(4d) \oplus S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4) \oplus S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4)\mathcal{P}_+ \\
&\qquad\qquad \oplus S^{+,\mathrm{new}}_{k+1/2}(4d)|U(8)B(2) \oplus S^{+,\mathrm{new}}_{k+1/2}(4d)|B(4) \\
&\qquad\qquad\qquad \oplus S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4)B(4)\big)|U(r^2) \\
&\quad \oplus \bigoplus_{rd|N} \big(S^{\mathrm{new}}_{k+1/2}(4d) \oplus S^{\mathrm{new}}_{k+1/2}(4d)|\mathcal{P}_+ \oplus S^{\mathrm{new}}_{k+1/2}(4d)|U(2)B(2) \\
&\qquad\qquad\qquad \oplus S^{\mathrm{new}}_{k+1/2}(4d)|B(4)\big)|U(r^2) \\
&\quad \oplus \bigoplus_{rd|N} \big(S^{\mathrm{new}}_{k+1/2}(8d) \oplus S^{\mathrm{new}}_{k+1/2}(8d)|W(16)\big)|U(r^2),
\end{aligned}
$$

where $W(16)$ is the W-operator corresponding to the prime $p=2$ in $S_{k+1/2}(16N)$.

Proof. It is enough to show the direct sum in the respective eigensubspaces. First consider the eigensubspace generated by $S^{+,\mathrm{new}}_{k+1/2}(4d)$. By Theorem 3.1, the sum $S^{+,\mathrm{new}}_{k+1/2}(4d) + S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4) + S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4)\mathcal{P}_+$ is direct and, assuming the rest of the sum in the eigensubspace is not direct, then we have $f \in S^{+,\mathrm{new}}_{k+1/2}(4d)$ which is nonzero and such that all odd coefficients of $f|U(8)$ are zero, by assuming $S^{+,\mathrm{new}}_{k+1/2}(4d)|U(8) \cap (S^{+,\mathrm{new}}_{k+1/2}(4d)|B(2) + S^{+,\mathrm{new}}_{k+1/2}(4d)|U(4)B(2))$ is nonzero. That is, $f|U(4) \in S_{k+1/2}(4d)$ has the property that its n-th Fourier coefficient is zero whenever $n \equiv 2 \pmod 4$. This means that $f|U(4) \in S^{+}_{k+1/2}(4d)$, a contradiction since $0 \neq f \in S^{+}_{k+1/2}(4d)$. Hence all the sums in the eigensubspace generated by $S^{+,\mathrm{new}}_{k+1/2}(4d)$ are direct. Next, consider the eigensubspace generated by $S^{\mathrm{new}}_{k+1/2}(4d)$. Clearly $S^{\mathrm{new}}_{k+1/2}(4d) \oplus S^{\mathrm{new}}_{k+1/2}(4d)|\mathcal{P}_+$ is a direct sum in $S_{k+1/2}(8N)$. If there is a nonzero element in the intersection of $S^{\mathrm{new}}_{k+1/2}(4d)|U(2)B(2)$ and $S^{\mathrm{new}}_{k+1/2}(4d)|B(4)$, then the n-th Fourier coefficient of a nonzero form $f \in S^{\mathrm{new}}_{k+1/2}(4d)$ vanishes whenever $n \equiv 2 \pmod 4$ and hence, by [Kohnen 1982, Lemma], $0 \neq f \in S^{+}_{k+1/2}(4d)$, a contradiction. So, the subspace $S_{k+1/2}|U(2)B(2) \oplus S^{\mathrm{new}}_{k+1/2}(4d)|B(4)$ is a direct sum in $S_{k+1/2}(16N)$. In order to prove that all the sum as above generated by $S^{\mathrm{new}}_{k+1/2}(4d)$ is direct, we use the following fact. If $f \in S_{k+1/2}(8N, \chi_8)$ and $f|B(2) \in S_{k+1/2}(8N)$, then $f = 0$, by [Serre and Stark 1977, Lemma 7]. Finally, applying $U(2)$ on the eigensubspace of $S^{\mathrm{new}}_{k+1/2}(8d)$, one component is mapped to zero and the other component is $S^{\mathrm{new}}_{k+1/2}(8d)|W(8)$, which is nonzero. Hence, we get that the sum in this eigensubspace is direct. This completes the proof for the direct sum decomposition of $S^{\mathrm{old}}_{k+1/2}(16N)$.

Since the spaces $S^{+,\mathrm{new}}_{k+1/2}(4d)$, $S^{\mathrm{new}}_{k+1/2}(4d)$ and $S^{\mathrm{new}}_{k+1/2}(8d)$ are isomorphic (under the Shimura correspondence) to the spaces $S^{\mathrm{new}}_{2k}(d)$, $S^{\mathrm{new}}_{2k}(2d)$ and $S^{\mathrm{new}}_{2k}(4d)$ respectively, we see that

(17)

$$
\dim S^{\mathrm{old}}_{k+1/2}(16N) = \sum_{rd|N} \big(6 \dim S^{\mathrm{new}}_{2k}(d) + 4 \dim S^{\mathrm{new}}_{2k}(2d) + 2 \dim S^{\mathrm{new}}_{2k}(4d)\big)
$$

$$= 2\sum_{rd|N}(3\dim S_{2k}^{\text{new}}(d) + 2\dim S_{2k}^{\text{new}}(2d) + \dim S_{2k}^{\text{new}}(4d))$$
$$= 2\dim S_{2k}(4N) = \dim S_{k+1/2}(16N)$$

from the above computation. Therefore, it follows that $S_{k+1/2}^{\text{new}}(16N) = \{0\}$. □

5. Newform theory on $S_{k+1/2}(16N, \chi_8)$

In this section, we study the theory of newforms on $S_{k+1/2}(16N, \chi_8)$, where χ_8 is the even quadratic character modulo 8 defined in the introduction and N is odd and squarefree. Since cond $\chi_8 = 8$, by Ueda's result [1988] there exists a Hecke equivariant isomorphism between the spaces $S_{k+1/2}(16N, \chi_8)$ and $S_{2k}(8N)$. Define the space of oldforms in $S_{k+1/2}(16N, \chi_8)$ as follows:

$$\begin{aligned}(18)\quad & S_{k+1/2}^{\text{old}}(16N, \chi_8)\\ &= \sum_{rd|N}(S_{k+1/2}^{+,\text{new}}(4d)|B(2) + S_{k+1/2}^{+,\text{new}}(4d)|U(2))U(r^2)\\ &\quad + \sum_{rd|N}(S_{k+1/2}^{+,\text{new}}(4d)|U(8) + S_{k+1/2}^{+,\text{new}}(4d)|U(8)W(8)B(2))U(r^2)\\ &\quad + \sum_{rd|N}(S_{k+1/2}^{\text{new}}(4d)|U(2) + S_{k+1/2}^{\text{new}}(4d)|B(2))U(r^2)\\ &\quad + \sum_{rd|N} S_{k+1/2}^{\text{new}}(4d)|U(2)W(8)B(2)U(r^2) + \sum_{rd|N} S_{k+1/2}^{\text{new}}(8d)|B(2)U(r^2)\\ &\quad + \sum_{rd|N} S_{k+1/2}^{\text{new}}(8d)|W(8)U(r^2) + \sum_{rd|N, d<N} S_{k+1/2}^{\text{new}}(16d, \chi_8)|U(r^2).\end{aligned}$$

First consider the sum in the eigensubspace generated by $S_{k+1/2}^{+,\text{new}}(4d)$. Suppose $(f_1|U(4) + f_2)|U(2) = f_3|B(2)$, where $f_i \in S_{k+1/2}^{+,\text{new}}(4d)$, $i = 1, 2, 3$. This implies that $f_1|U(4) + f_2 \in S_{k+1/2}(4d)$ is such that all its Fourier coefficients which are congruent to 2 modulo 4 are zero. Hence, by [Kohnen 1982, Lemma], we conclude that $f_1|U(4) + f_2 \in S_{k+1/2}^{+}(4d)$. Thus, $f_1 = 0$. Therefore, $f_2|U(2) = f_3|B(2)$, i.e., f_2 and $f_2|U(4)$ belong to $S_{k+1/2}^{+}(4d)$, which implies that f_2 and hence $f_3 = 0$. Now, among the four components, the first three direct sums belong to $S_{k+1/2}(8d, \chi_8)$. But, the fourth one is in $S_{k+1/2}(4d)|B(2) \in S_{k+1/2}(16, \chi_8)$. This shows that all the four components form a direct sum. Next, consider the eigensubspaces generated by $S_{k+1/2}^{\text{new}}(4d)$ and $S_{k+1/2}^{\text{new}}(8d)$. A similar argument as above together with the following lemma shows that the respective sums are direct.

Lemma 5.1. *The operator $U(2)W(8)$ has the following mapping property*:

$$U(2)W(8) : S_{k+1/2}(4N) \longrightarrow S_{k+1/2}(8N).$$

Moreover, if $f \in S_{k+1/2}(4N)$, then $f|U(2)W(8) \in S_{k+1/2}(4N)$ if and only if $f \in S^+_{k+1/2}(4N)$.

Proof. The mapping property follows from a straightforward verification. Suppose $f|U(2)W(8) = g$, where $f, g \in S_{k+1/2}(4N)$. Using

$$W(8)W_*(4) = \chi_8(2k+1)\left(\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}, 2^{1/4}\right) \quad \text{on } S_{k+1/2}(8N, \chi_8) \tag{19}$$

and

$$W_*(4) = W(4) \quad \text{on } S_{k+1/2}(4N), \tag{20}$$

we get

$$f|U(2)\Bigg|\left(\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}, 2^{1/4}\right) = \chi_8(2k+1)\, g|W(4).$$

Now, $g|W(4)$ is invariant under $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)^*$. Hence, $a_{f|U(2)}(n) = 0$ if n is odd and, therefore, $a_f(n) = 0$ whenever $n \equiv 2 \pmod 4$. This proves that $f \in S^+_{k+1/2}(4N)$, a contradiction. For a detailed proof of the identities (19) and (20), we refer to [Manickam 2011]. □

Define the space of newforms in $S_{k+1/2}(16N, \chi_8)$ to be the orthogonal complement (with respect to the Petersson scalar product) of $S^{\text{old}}_{k+1/2}(16N, \chi_8)$ in $S_{k+1/2}(16N, \chi_8)$. It is already known that the spaces $S^{+,\text{new}}_{k+1/2}(4d)$, $S^{\text{new}}_{k+1/2}(4d)$ and $S^{\text{new}}_{k+1/2}(8d)$ are isomorphic (respectively) to $S^{\text{new}}_{2k}(d)$, $S^{\text{new}}_{2k}(2d)$ and $S^{\text{new}}_{2k}(4d)$. Using induction on the number of prime factors of N, it follows that the space $S^{\text{new}}_{k+1/2}(16d, \chi_8)$ is isomorphic to $S^{\text{new}}_{2k}(8d)$ if $d|N$ and $d < N$. Now, comparing the dimension of the space $S^{\text{old}}_{2k}(8N)$, we see that the spaces $S^{\text{old}}_{k+1/2}(16N, \chi_8)$ and $S^{\text{old}}_{2k}(8N)$ have equal dimension. As mentioned at the beginning of this section, Ueda [1988] has shown that the spaces $S_{k+1/2}(16N, \chi_8)$ and $S_{2k}(8N)$ are Hecke-equivariantly isomorphic when N is odd and squarefree. Therefore, combining all these facts, it follows that the space $S^{\text{new}}_{k+1/2}(16N, \chi_8)$ is isomorphic to $S^{\text{new}}_{2k}(8N)$.

We summarise the results of this section in the following.

Theorem 5.2. *Let N be an odd and squarefree natural number and let χ_8 be the primitive even quadratic Dirichlet character modulo 8. Then $S_{k+1/2}(16N, \chi_8)$ has an orthogonal decomposition*

$$S_{k+1/2}(16N, \chi_8) = S^{\text{new}}_{k+1/2}(16N, \chi_8) \oplus S^{\text{old}}_{k+1/2}(16N, \chi_8),$$

and

$$
\begin{aligned}
(21)\quad & S^{\text{old}}_{k+1/2}(16N,\chi_8)\\
&= \bigoplus_{rd|N}\big(S^{+,\text{new}}_{k+1/2}(4d)|B(2)\oplus S^{+,\text{new}}_{k+1/2}(4d)|U(2)\oplus S^{+,\text{new}}_{k+1/2}(4d)|U(8)\\
&\qquad\qquad\oplus S^{+,\text{new}}_{k+1/2}(4d)|U(8)W(8)B(2)\big)U(r^2)\\
&\quad\oplus \bigoplus_{rd|N}\big(S^{\text{new}}_{k+1/2}(4d)|U(2)\oplus S^{\text{new}}_{k+1/2}(4d)|B(2)\\
&\qquad\qquad\oplus_{rd|N} S^{\text{new}}_{k+1/2}(4d)|U(2)W(8)B(2)\big)U(r^2)\\
&\quad\oplus \bigoplus_{rd|N}\big(S^{\text{new}}_{k+1/2}(8d)|B(2)\oplus S^{\text{new}}_{k+1/2}(8d)|W(8)\big)U(r^2)\\
&\quad\oplus \bigoplus_{rd|N,d<N} S^{\text{new}}_{k+1/2}(16d,\chi_8)|U(r^2).
\end{aligned}
$$

The spaces $S^{\text{new}}_{k+1/2}(16N,\chi_8)$ *and* $S^{\text{old}}_{k+1/2}(16N,\chi_8)$ *are mapped, respectively, into the spaces* $S^{\text{new}}_{2k}(8N)$ *and* $S^{\text{old}}_{2k}(8N)$ *under the Shimura lifting. Moreover, the spaces of newforms* $S^{\text{new}}_{k+1/2}(16N,\chi_8)$ *and* $S^{\text{new}}_{2k}(8N)$ *are isomorphic under a linear combination of Shimura maps indexed by squarefree integers* $t\equiv 1 \pmod 4$, $(-1)^k t>0$.

Acknowledgements

Some part of this work was done during Manickam and Ramakrishnan's visit to the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy during May 2012. They thank the centre for the hospitality provided during their visit. Manickam also thanks the Harish-Chandra Research Institute, Allahabad for the hospitality during his visits to the institute. Finally, we thank the referee for a careful reading of the manuscript and for making useful suggestions.

References

[Atkin and Lehner 1970] A. O. L. Atkin and J. Lehner, "Hecke operators on $\Gamma_0(m)$", *Math. Ann.* **185** (1970), 134–160. MR 42 #3022 Zbl 0177.34901

[Cohen and Oesterlé 1977] H. Cohen and J. Oesterlé, "Dimensions des espaces de formes modulaires", pp. 69–78 in *Modular functions of one variable, VI* (Bonn, 1976), edited by J.-P. Serre and D. B. Zagier, Lecture Notes in Math. **627**, Springer, Berlin, 1977. MR 57 #12396 Zbl 0371.10020

[Gun et al. 2010] S. Gun, M. Manickam, and B. Ramakrishnan, "A canonical subspace of modular forms of half-integral weight", *Math. Ann.* **347**:4 (2010), 899–916. MR 2011g:11088 Zbl 1219.11070

[Kohnen 1980] W. Kohnen, "Modular forms of half-integral weight on $\Gamma_0(4)$", *Math. Ann.* **248**:3 (1980), 249–266. MR 81j:10030 Zbl 0416.10023

[Kohnen 1982] W. Kohnen, "Newforms of half-integral weight", *J. Reine Angew. Math.* **333** (1982), 32–72. MR 84b:10038 Zbl 0475.10025

[Kohnen 1985] W. Kohnen, "Fourier coefficients of modular forms of half-integral weight", *Math. Ann.* **271**:2 (1985), 237–268. MR 86i:11018 Zbl 0542.10018

[Manickam 1980] M. Manickam, *Newforms of half-integral weight and some problems on modular forms*, thesis, University of Madras, 1980.

[Manickam 2011] M. Manickam, "Newforms of half-integral weight on $\Gamma_0(8N)$", pp. 63–71 in *Number theory*, edited by M. Manickam and B. Ramakrishnan, Ramanujan Math. Soc. Lect. Notes Ser. **15**, Ramanujan Mathematical Society, Mysore, 2011. MR 2905488 Zbl 06103450

[Manickam et al. 1989] M. Manickam, B. Ramakrishnan, and T. C. Vasudevan, "On Shintani correspondence", *Proc. Indian Acad. Sci. Math. Sci.* **99**:3 (1989), 235–247. MR 90m:11061 Zbl 0689.10033

[Manickam et al. 1990] M. Manickam, B. Ramakrishnan, and T. C. Vasudevan, "On the theory of newforms of half-integral weight", *J. Number Theory* **34**:2 (1990), 210–224. MR 91b:11060 Zbl 0704.11013

[Martin 2005] G. Martin, "Dimensions of the spaces of cusp forms and newforms on $\Gamma_0(N)$ and $\Gamma_1(N)$", *J. Number Theory* **112**:2 (2005), 298–331. MR 2005m:11069 Zbl 1095.11026

[Murty and Murty 1997] M. R. Murty and V. K. Murty, *Non-vanishing of L-functions and applications*, Progress in Mathematics **157**, Birkhäuser, Basel, 1997. MR 98h:11106 Zbl 0916.11001

[Niwa 1975] S. Niwa, "Modular forms of half integral weight and the integral of certain theta-functions", *Nagoya Math. J.* **56** (1975), 147–161. MR 51 #361 Zbl 0303.10027

[Niwa 1977] S. Niwa, "On Shimura's trace formula", *Nagoya Math. J.* **66** (1977), 183–202. MR 58 #27781 Zbl 0351.10018

[Ono 2004] K. Ono, *The web of modularity: arithmetic of the coefficients of modular forms and q-series*, CBMS Regional Conference Series in Mathematics **102**, American Mathematical Society, Providence, RI, 2004. MR 2005c:11053 Zbl 1119.11026

[Serre and Stark 1977] J.-P. Serre and H. M. Stark, "Modular forms of weight 1/2", pp. 27–67 in *Modular functions of one variable, VI* (Bonn, 1976), edited by J.-P. Serre and D. B. Zagier, Lecture Notes in Math. **627**, Springer, Berlin, 1977. MR 57 #12400 Zbl 0371.10019

[Shemanske 1996] T. R. Shemanske, "Newforms of half-integral weight", *Nagoya Math. J.* **143** (1996), 147–169. MR 98k:11053 Zbl 0863.11032

[Shimura 1973] G. Shimura, "On modular forms of half integral weight", *Ann. of Math.* (2) **97** (1973), 440–481. MR 48 #10989 Zbl 0266.10022

[Shintani 1975] T. Shintani, "On construction of holomorphic cusp forms of half integral weight", *Nagoya Math. J.* **58** (1975), 83–126. MR 52 #10603 Zbl 0316.10016

[Skoruppa and Zagier 1988] N.-P. Skoruppa and D. B. Zagier, "Jacobi forms and a certain space of modular forms", *Invent. Math.* **94**:1 (1988), 113–146. MR 89k:11029 Zbl 0651.10020

[Ueda 1988] M. Ueda, "The decomposition of the spaces of cusp forms of half-integral weight and trace formula of Hecke operators", *J. Math. Kyoto Univ.* **28**:3 (1988), 505–555. MR 90a:11054 Zbl 0673.10021

[Ueda 1991] M. Ueda, "The trace formulae of twisting operators on the spaces of cusp forms of half-integral weight and some trace relations", *Japan. J. Math.* (*N.S.*) **17**:1 (1991), 83–135. MR 92g:11045 Zbl 0742.11031

[Ueda 1993] M. Ueda, "On twisting operators and newforms of half-integral weight", *Nagoya Math. J.* **131** (1993), 135–205. MR 95a:11040 Zbl 0778.11027

[Ueda 1998] M. Ueda, "On twisting operators and newforms of half-integral weight, II: Complete theory of newforms for Kohnen space", *Nagoya Math. J.* **149** (1998), 117–171. MR 99c:11050 Zbl 1016.11505

[Ueda 2001] M. Ueda, "On twisting operators and newforms of half-integral weight, III: Subspace corresponding to very-newforms", *Comment. Math. Univ. St. Paul.* **50**:1 (2001), 1–27. MR 2002d:11048 Zbl 0995.11034

[Ueda and Yamana 2010] M. Ueda and S. Yamana, "On newforms for Kohnen plus spaces", *Math. Z.* **264**:1 (2010), 1–13. MR 2011a:11088 Zbl 1277.11045

Received April 24, 2014. Revised August 22, 2014.

MURUGESAN MANICKAM
KERALA SCHOOL OF MATHEMATICS
KUNNAMANGALAM P.O.
KOZHIKODE 673 571
INDIA
murugumanick@ksom.res.in

JABAN MEHER
INSPIRE FELLOW, DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560 012
INDIA
jaban@math.iisc.ernet.in

BALAKRISHNAN RAMAKRISHNAN
HARISH-CHANDRA RESEARCH INSTITUTE
CHHATNAG ROAD, JHUNSI
ALLAHABAD 211 019
INDIA
ramki@hri.res.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.141

ALGEBRAIC FAMILIES OF HYPERELLIPTIC CURVES VIOLATING THE HASSE PRINCIPLE

NGUYEN NGOC DONG QUAN

In 2000, Colliot-Thélène and Poonen showed how to construct algebraic families of genus-one curves violating the Hasse principle. Poonen explicitly constructed such a family of cubic curves using the general method developed by Colliot-Thélène and himself. The main result in this paper generalizes the result of Colliot-Thélène and Poonen to arbitrarily high genus hyperelliptic curves. More precisely, for $n > 5$ and $n \not\equiv 0 \pmod 4$, we show that there is an explicit algebraic family of hyperelliptic curves of genus n that are counterexamples to the Hasse principle explained by the Brauer–Manin obstruction.

1. Introduction

The aim of this article is to prove the following result.

Theorem 1.1 (see Theorem 6.8). *Let $n > 5$ be an integer such that $n \not\equiv 0 \pmod 4$. Then there is an algebraic family $\mathcal{C}_t$ of hyperelliptic curves of genus n such that $\mathcal{C}_t$ is a counterexample to the Hasse principle explained by the Brauer–Manin obstruction for all $t \in \mathbb{Q}$. Furthermore, $\mathcal{C}_t$ contains no zero-cycles of odd degree over $\mathbb{Q}$ for all $t \in \mathbb{Q}$.*

MSC2010: 14G05, 11G35, 11G30.
Keywords: Azumaya algebras, Brauer groups, Brauer–Manin obstruction, Hasse principle, hyperelliptic curves.

We will shortly relate this theorem to existing results in literature, and sketch the ideas of the proof of Theorem 1.1. Let us begin by briefly recalling some terminology which appears in many places in this paper. For a basic introduction to the Brauer–Manin obstruction, see [Skorobogatov 2001; Poonen 2008].

Recall from [Poonen 2001] that an *algebraic* family of curves is a family of curves depending on a parameter T such that substituting any rational number for T results in a smooth curve over $\mathbb{Q}$.

A smooth geometrically irreducible curve $\mathcal{C}$ over $\mathbb{Q}$ is said to *satisfy the Hasse principle* if the everywhere local solvability of $\mathcal{C}$ is equivalent to the global solvability of $\mathcal{C}$. In more concrete terms, this means that

$$\mathcal{C}(\mathbb{Q}) \neq \varnothing \text{ if and only if } \mathcal{C}(\mathbb{Q}_p) \neq \varnothing \text{ for every prime } p \text{ including } p = \infty.$$

If $\mathcal{C}$ has points locally everywhere but has no rational points, we say that $\mathcal{C}$ is a *counterexample to the Hasse principle*. Furthermore, if we also have $\mathcal{C}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}} = \varnothing$ (see [Poonen 2008], or [Skorobogatov 2001] for the definition of $\mathcal{C}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}}$), we say that $\mathcal{C}$ is a *counterexample to the Hasse principle explained by the Brauer–Manin obstruction*. The Hasse principle fails in general. The first counterexamples of genus-one curves to the Hasse principle were discovered by Lind [1940] and independently shortly thereafter by Reichardt [1942].

Let us relate Theorem 1.1 to existing results in literature. For $n = 1$, Colliot-Thélène and Poonen [2000] showed how to produce one-parameter families of curves of genus one violating the Hasse principle. Poonen [2001] explicitly constructed an algebraic family of genus-one cubic curves violating the Hasse principle using the general method developed in [Colliot-Thélène and Poonen 2000]. It is not known whether there exists an algebraic family of curves of genus n violating the Hasse principle for all $n \geq 2$.

Here, as throughout the article, we say that a smooth geometrically irreducible variety $\mathcal{V}$ over $\mathbb{Q}$ *satisfies* CHP if it is a counterexample to the Hasse principle explained by the Brauer–Manin obstruction. A smooth geometrically irreducible variety $\mathcal{V}$ over $\mathbb{Q}$ is said to *satisfy* NZC if it contains no zero-cycles of odd degree over $\mathbb{Q}$.

Coray and Manoil [1996] showed that for each positive integer $n \geq 2$, the smooth projective model of the affine curve defined by

$$z^2 = 605 \cdot 10^6 x^{2n+2} + (18x^2 - 4400)(45x^2 - 8800) \tag{1}$$

satisfies CHP and NZC. The Coray–Manoil family of curves is the first family of hyperelliptic curves of varying genus that satisfies CHP and NZC. Although the authors restricted themselves to constructing only one hyperelliptic curve of genus n satisfying CHP and NZC for each integer $n \geq 2$, it seems plausible that their approach can be modified to produce algebraic families of hyperelliptic curves of arbitrary genus satisfying CHP and NZC. Since we will follow the approach of Coray and Manoil with some modifications to prove Theorem 1.1, we briefly recall their main ideas for constructing the family (1).

Colliot-Thélène, Coray and Sansuc [Colliot-Thélène et al. 1980] proved that the threefold $\mathcal{Y}_{(5,1,1)}$ in $\mathbb{P}^5_{\mathbb{Q}}$, defined by

$$\mathcal{Y}_{(5,1,1)} : \begin{cases} u_1^2 - 5v_1^2 = 2xy, \\ u_2^2 - 5v_2^2 = 2(x+20y)(x+25y), \end{cases}$$

satisfies CHP and NZC. Building on this result, Coray and Manoil [1996] introduced a geometric construction of hyperelliptic curves that allows to smoothly embed the family of curves defined by (1) into the threefold $\mathcal{Y}_{(5,1,1)}$. It follows immediately from functoriality that the Coray–Manoil family of curves satisfies CHP and NZC.

In order to generalize the result of Coray and Manoil, we first construct a family of threefolds in $\mathbb{P}^5_{\mathbb{Q}}$ that satisfies CHP and NZC and has the threefold $\mathcal{Y}_{(5,1,1)}$ as a member. The construction of such threefolds is achieved by building on that of the threefold $\mathcal{Y}_{(5,1,1)}$. In order to show that the Brauer–Manin obstruction for these threefolds is nonempty, we also need to show the existence of infinitely many primes p and q satisfying certain quadratic equations. We do this by calling on the result of [Iwaniec 1974] that a quadratic polynomial in two variables represents infinitely many primes. Since the existence of certain threefolds in $\mathbb{P}^5_{\mathbb{Q}}$ satisfying CHP and NZC is of interest in its own right, we state this result here.

Theorem 1.2. *Let p be a prime such that $p \equiv 5 \pmod 8$ and 3 is quadratic nonresidue in $\mathbb{F}_p^\times$. Then there exist infinitely many pairs $(b, d) \in \mathbb{Z}^2$ such that any smooth and proper $\mathbb{Q}$-model $\mathcal{Z}$ of the smooth $\mathbb{Q}$-variety $\mathcal{X}$ in $\mathbb{A}^5_{\mathbb{Q}}$, defined by*

$$\begin{cases} 0 \neq u_1^2 - pv_1^2 = 2x, \\ 0 \neq u_2^2 - pv_2^2 = 2(x+4pb^2)(x+p^2d^2), \end{cases}$$

satisfies CHP *and* NZC.

The next step is to choose a family of hyperelliptic curves of arbitrary genus that can be smoothly embedded into the family of threefolds in Theorem 1.2 using the geometric construction of Coray and Manoil. For each $n \geq 2$, we define a family of hyperelliptic curves of genus n of the shape

$$z^2 = p\alpha^2 Q^2 x^{2n+2} + (2b^2 Px^2 + \beta Q)(d^2 pPx^2 + 2\beta Q), \tag{2}$$

where α, β, γ are certain rational numbers, and P, Q depend on α, β, γ, p, b, d. In order to apply the geometric construction of hyperelliptic curves of Coray and Manoil, the polynomials on the right-hand side of (2) are required to be separable.

In order to smoothly embed these hyperelliptic curves into the threefolds in Theorem 1.2, we impose certain conditions on α, β, γ such that these rational numbers satisfy certain local congruences and certain conics in $\mathbb{P}^2_{\mathbb{Q}}$ constructed from sextuples $(p, b, d, \alpha, \beta, \gamma)$ possess at least one nontrivial rational point. Lemmas 5.1 and 5.4 show that there are infinitely many sextuples $(p, b, d, \alpha, \beta, \gamma)$ satisfying these conditions. For any such sextuple, it follows from functoriality and Theorem 1.2 that the family of hyperelliptic curves of genus n defined by (2) satisfies CHP and NZC for each $n \geq 2$.

In the last step, the main difficulty is to show the existence of rational functions in $\mathbb{Q}(T)$ that parametrize rational numbers α, β, γ such that for each integer $n \geq 2$, substituting any rational number for T in the polynomials on the right-hand side of (2) results in a separable polynomial of degree $2n+2$ over $\mathbb{Q}$. We do this by calling on a *separability criterion* from [Dong Quan 2014], which will be reviewed in Section 6.

After this article was finished, the author learned that Bhargava, Gross, and Wang [Bhargava et al. 2013] showed that for any integer $n \geq 1$, there is a positive proportion of everywhere locally solvable hyperelliptic curves over $\mathbb{Q}$ of genus n that have no points over any number field of odd degree over $\mathbb{Q}$. Despite this remarkable result, it cannot determine whether an explicit hyperelliptic curve over Q satisfies CHP and NZC. The main theorem of this article describes an explicit algebraic family of such curves of genus n with $\gcd(n, 4) = 1$ and $n > 5$.

2. The Hasse principle for certain threefolds in $\mathbb{P}^5_{\mathbb{Q}}$

In this section, we will construct families of threefolds satisfying CHP and NZC. We begin by stating some lemmas that we will need in the proof of the main results throughout the paper.

Lemma 2.1 (see [Coray and Manoil 1996, Lemma 4.8]). *Let k be a number field, and let $\mathcal{V}_1$ and $\mathcal{V}_2$ be (proper) k-varieties. Assume that there is a k-morphism $\alpha : \mathcal{V}_1 \to \mathcal{V}_2$ and $\mathcal{V}_2(\mathbb{A}_k)^{\mathrm{Br}} = \varnothing$. Then $\mathcal{V}_1(\mathbb{A}_k)^{\mathrm{Br}} = \varnothing$.*

Lemma 2.2 (see, for example, [Corn 2007, Proposition 6.4]). *Let $\mathcal{X}$ be a smooth F-variety. Let L/F be a cyclic extension, and let $F(\mathcal{X})$ be the function field of $\mathcal{X}$. Let f be an element of $F(\mathcal{X})$, and let $\mathcal{X}_L = \mathcal{X} \times_F L$. Then the class of the cyclic algebra $(L/F, f) \in \mathrm{Br}(F(\mathcal{X}))$ lies in the image of the inclusion $\mathrm{Br}(\mathcal{X}) \hookrightarrow \mathrm{Br}(F(\mathcal{X}))$ if and only if $\mathrm{div}(f) = \mathrm{Norm}_{L/F}(D)$ for some $D \in \mathrm{Div}(\mathcal{X}_L)$.*

Lemma 2.3 (Lang–Nishimura, [Colliot-Thélène et al. 1980, p. 164, Lemme 3.1.1]). *Let F be a field, and let $\mathcal{X}$ be an integral F-variety. Let $\mathcal{Y}$ be a proper F-variety, and let $f : \mathcal{X} \longrightarrow \mathcal{Y}$ be an F-rational map. If $\mathcal{X}(F)$ contains a regular F-point, then $\mathcal{Y}(F)$ is nonempty. In particular, the condition $\mathcal{X}(F) \neq \varnothing$ is an F-birational invariant in the category of smooth, proper and integral F-varieties $\mathcal{X}$.*

We now describe a construction of certain Azumaya algebras on certain threefolds.

Lemma 2.4. *Let p be a prime such that $p \equiv 5 \pmod 8$. Assume that:*

(A1) 3 *is a quadratic nonresidue in* $\mathbb{F}_p^{\times}$.

(B) *There exists a pair* $(b, c) \in \mathbb{Z}^2$ *such that* $\gcd(b, c) = 1$, $b \not\equiv 0 \pmod p$, *and* $q := |pc - 4b^2|$ *is either* 1 *or an odd power of an odd prime. Here* $|\cdot|$ *denotes the absolute value in* $\mathbb{Q}$. *Furthermore, if* $b \equiv 0 \pmod 3$, *then* $c \equiv 2 \pmod 3$.

Let $\mathcal{V}$ *be a smooth, proper* $\mathbb{Q}$*-model of the smooth* $\mathbb{Q}$*-variety* $\mathcal{U}$ *in* $\mathbb{A}^5_{\mathbb{Q}}$ *defined by*

$$\mathcal{U} : \begin{cases} 0 \neq u_1^2 - pv_1^2 = 2x, \\ 0 \neq u_2^2 - pv_2^2 = 2(x + 4pb^2)(x + p^2c). \end{cases} \tag{3}$$

Let $\mathbb{Q}(\mathcal{V})$ *be the function field of* $\mathcal{V}$, *and let* $\mathcal{A}$ *be the class of the quaternion algebra* $(p, x + 4pb^2)$. *Then* $\mathcal{A}$ *is an Azumaya algebra of* $\mathcal{V}$, *that is,* $\mathcal{A}$ *belongs to the subgroup* $\mathrm{Br}(\mathcal{V})$ *of* $\mathrm{Br}(\mathbb{Q}(\mathcal{V}))$.

Proof. Let $K = \mathbb{Q}(\sqrt{p})$. Let Γ be the divisor defined over $\mathbb{Q}(\sqrt{p})$ and lying on $\mathcal{V}$ defined by

$$\Gamma : \quad f := x + 4pb^2 = 0, \quad u_2 - \sqrt{p}v_2 = 0, \quad u_1^2 - pv_1^2 = -8pb^2.$$

Let σ be a generator of $\mathrm{Gal}(K/\mathbb{Q})$. We see that $\mathrm{div}(f) = \Gamma + \sigma\Gamma$, and it thus follows from Lemma 2.2 that $\mathcal{A}$ is in the image of $\mathrm{Br}(\mathcal{V}) \hookrightarrow \mathrm{Br}(\mathbb{Q}(\mathcal{V}))$. □

Lemma 2.5. *Let* p *be a prime such that* $p \equiv 5 \pmod 8$. *Assume that conditions (A1) and (B) in Lemma 2.4 are true. Then there exists a nonzero integer* a *such that*

$$\gcd\big((a^2 + 2pb^2)(2a^2 + p^2c), 3(2b^2 + pc)\big) = 1. \tag{4}$$

Proof. Assume that $H_1 := 2b^2 + pc = \pm\prod_{i=1}^m l_i^{\alpha_i}$, where l_i are distinct primes and $\alpha_i \in \mathbb{Z}_{>0}$. Note that since $q = |pc - 4b^2|$ is either 1 or an odd power of an odd prime, c is odd. Thus H_1 is odd, and therefore $l_i \neq 2$ for each $1 \leq i \leq m$. We also have that $l_i \neq p$ for each $1 \leq i \leq m$; otherwise, $l_i = p$ for some integer $1 \leq i \leq m$. Since $2b^2 + pc \equiv 0 \pmod{l_i}$ and $l_i = p$, it follows that $b \equiv 0 \pmod p$, which is a contradiction. We consider the following cases:

Case 1. $b \equiv 0 \pmod 3$.

By assumption (B), one knows that $c \equiv 2 \pmod 3$. Define $a := \prod_{i=1}^m l_i$. We contend that a satisfies (4). Indeed, we have that $l_i \neq 3$ for each $1 \leq i \leq m$; otherwise, $l_i = 3$ for some integer $1 \leq i \leq m$. Since $b \equiv 0 \pmod 3$ and $p \neq 3$, it follows that $c \equiv 0 \pmod 3$, which is a contradiction.

Let $H_2 := a^2 + 2pb^2$ and $H_3 := 2a^2 + p^2c$. We see that $a^2 = \prod_{i=1}^m l_i^2 \equiv 1 \pmod 3$. Since $p \neq 3$, we deduce that $H_2 \equiv 1 \pmod 3$ and $H_3 \equiv 2 + c \equiv 1 \pmod 3$, and thus $H_2H_3 \equiv 1 \pmod 3$.

Suppose that l_j divides H_2 for some integer $1 \leq j \leq m$. Since $a = \prod_{i=1}^m l_i \equiv 0 \pmod{l_j}$, it follows that $b \equiv 0 \pmod{l_j}$. Thus $c \equiv 0 \pmod{l_j}$, which is a contradiction to (B).

Suppose that l_j divides H_3 for some integer $1 \leq j \leq m$. Since $a = \prod_{i=1}^{m} l_i \equiv 0 \pmod{l_j}$ and $l_j \neq p$, it follows that $c \equiv 0 \pmod{l_j}$. Hence $b \equiv 0 \pmod{l_j}$, which is a contradiction to (B). Therefore, in any event, (4) holds.

Case 2. $b \not\equiv 0 \pmod 3$ *and* $c \equiv 0 \pmod 3$.

Let $a := \prod_{i=1}^{m} l_i$. By (A1), we know that $p \equiv 2 \pmod 3$. Hence repeating in the same manner as in Case 1, we deduce that (4) holds.

Case 3. $b \not\equiv 0 \pmod 3$ *and* $c \not\equiv 0 \pmod 3$.

Let $a := 3\prod_{i=1}^{m} l_i$. The same arguments as in Case 1 show that (4) holds. □

Following the techniques in the proof of [Colliot-Thélène et al. 1980, Proposition 7.1], we now prove the main theorem in this section.

Theorem 2.6. *We maintain the same notation as in Lemma 2.4. Let p be a prime such that $p \equiv 5 \pmod 8$. Assume further that (A1) and (B) are true. Let $\mathcal{U}$ and $\mathcal{V}$ be the $\mathbb{Q}$-varieties defined in Lemma 2.4. Let $\mathcal{T}$ be the singular $\mathbb{Q}$-variety in $\mathbb{P}^5_{\mathbb{Q}}$ defined by*

$$\mathcal{T}: \begin{cases} u_1^2 - pv_1^2 = 2xy, \\ u_2^2 - pv_2^2 = 2(x + 4pb^2y)(x + p^2cy). \end{cases} \tag{5}$$

Then $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{T}$ satisfy CHP *and* NZC.

Proof. The proof of Theorem 2.6 is divided into several steps.

Step 1. $\mathcal{U}(\mathbb{Q}) = \mathcal{T}(\mathbb{Q})$.

It is clear that $\mathcal{U}(\mathbb{Q}) \subseteq \mathcal{T}(\mathbb{Q})$. Assume that there is a point

$$P := (x : y : u_1 : v_1 : u_2 : v_2) \in \mathcal{T}(\mathbb{Q}).$$

Suppose first that $y = 0$. Then $u_1 = v_1 = 0$. If furthermore $x = 0$, then $u_2 = v_2 = 0$, which is a contradiction. Hence $x \neq 0$, and thus $2 = (u_2/x)^2 - p(v_2/x)^2$. Hence 2 is the norm of an element in $\mathbb{Q}(\sqrt{p})^\times$, and therefore 2 is the norm of an element in $\mathbb{Q}_p(\sqrt{p})^\times$. Thus the local Hilbert symbol $(2, p)_p$ is 1. On the other hand, using [Cohen 2007, p. 296, Theorem 5.2.7] and $p \equiv 5 \pmod 8$, we deduce that

$$(2, p)_p = \left(\frac{2}{p}\right) = -1,$$

which is a contradiction.

Now we assume that $y \neq 0$, and with no loss of generality, assume further that $y = 1$. We consider the following cases:

Case 1. $x = 0$.

The second equation of (5) implies that $u_2^2 - pv_2^2 = 8p^3b^2c$. Thus $8p^3b^2c$ is the norm of an element in $\mathbb{Q}_2(\sqrt{p})^\times$, and hence the local Hilbert symbol $(8p^3b^2c, p)_2$ is 1. Since $q = |pc - 4b^2|$ is either 1 or an odd power of an odd prime, c is odd.

Hence $v_2(8p^3b^2c) = 3 + 2v_2(b)$, which is an odd integer. Using [Cohen 2007, loc. cit.], we deduce that

$$(8p^3b^2c, p)_2 = \left(\frac{p}{2}\right) = -1,$$

which is a contradiction.

Case 2. $x = -4pb^2$.

It follows from (5) that $u_1^2 - pv_1^2 = -8pb^2$. Using the same arguments as in Case 1, we deduce that $-8pb^2$ is not the norm of any element in $\mathbb{Q}_2(\sqrt{p})^\times$, which is a contradiction to the last identity.

Case 3. $x = -p^2c$.

We see from (5) that $u_1^2 - pv_1^2 = -2p^2c$. Using the same arguments as in Case 1, we deduce that $-2p^2c$ is not the norm of any element in $\mathbb{Q}_2(\sqrt{p})^\times$, which is a contradiction to the last identity.

Therefore, in any event, we have shown that if the point $P := (x : y : u_1 : v_1 : u_2 : v_2)$ belongs to $\mathcal{T}(\mathbb{Q})$, then $y = 1$, $x \neq 0$, $x + 4pb^2 \neq 0$ and $x + p^2c \neq 0$. In other words, the point P satisfies

$$\begin{cases} 0 \neq u_1^2 - pv_1^2 = 2x, \\ 0 \neq u_2^2 - pv_2^2 = 2(x + 4pb^2)(x + p^2c), \end{cases}$$

and thus $P \in \mathcal{U}(\mathbb{Q})$. Therefore $\mathcal{U}(\mathbb{Q}) = \mathcal{T}(\mathbb{Q})$.

Step 2. *$\mathcal{U}$, $\mathcal{V}$, and $\mathcal{T}$ are everywhere locally solvable.*

We now prove that $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{T}$ are everywhere locally solvable. By Lemma 2.3, it suffices to prove that $\mathcal{U}$ is everywhere locally solvable. Recall that by Lemma 2.5, there is a nonzero integer a such that

$$\gcd\big((a^2 + 2pb^2)(2a^2 + p^2c), 3(2b^2 + pc)\big) = 1.$$

Hence it suffices to consider the following cases:

Case I. l is a prime such that $l \neq p$ and $\gcd(l, (a^2 + 2pb^2)(2a^2 + p^2c)) = 1$.

Let $x = 2a^2$. Since $2x = 4a^2$ is a square in $\mathbb{Z}$, we see that the local Hilbert symbol $(2x, p)_l$ satisfies

$$(2x, p)_l = (4a^2, p)_l = 1.$$

Thus $2x$ is the norm of an element in $\mathbb{Q}_l(\sqrt{p})^\times$.

We see that

$$\begin{aligned} v_l(2(x + 4pb^2)(x + p^2c)) &= v_l(4(a^2 + 2pb^2)(2a^2 + p^2c)) \\ &= 2v_l(2) + v_l((a^2 + 2pb^2)(2a^2 + p^2c)) = 2v_l(2). \end{aligned}$$

Hence, using [Cohen 2007, loc. cit.], we deduce that the local Hilbert symbol $(2(x+4pb^2)(x+p^2c), p)_l$ equals 1. Thus $2(x+4pb^2)(x+p^2c)$ is the norm of an element in $\mathbb{Q}_l(\sqrt{p})^\times$. Therefore $\mathcal{U}$ is locally solvable at l.

Case II. l is a prime such that $\gcd(l, 3(2b^2+pc)) = 1$. Note that p is among these primes.

Assume first that $l = p$, and set $x = 2pb^2$. We see that $2x = p(2b)^2$, and $2(x+4pb^2)(x+p^2c) = p^2(12b^2)(2b^2+pc)$. Note that $(2b)^2 \not\equiv 0 \pmod{p}$ and $(12b^2)(2b^2+pc) \equiv 6(2b^2)^2 \not\equiv 0 \pmod{p}$. Hence, using [Cohen 2007, loc. cit.], we deduce that the local Hilbert symbol $(2x, p)_p$ satisfies

$$(2x, p)_p = (-1)^{(p-1)/2}\left(\frac{(2b)^2}{p}\right) = 1.$$

Hence $2x$ is the norm of an element in $\mathbb{Q}_p(\sqrt{p})$.

By (A1), we know that 6 is quadratic residue in $\mathbb{F}_p^\times$. Since $(12b^2)(2b^2+pc) \equiv 6(2b^2)^2 \pmod{p}$, we see that $(12b^2)(2b^2+pc)$ is a quadratic residue in $\mathbb{F}_p^\times$. Thus using the same arguments as above, we deduce that

$$(2(x+4pb^2)(x+p^2c), p)_p = (p^2(12b^2)(2b^2+pc), p)_p = 1.$$

Therefore $2(x+4pb^2)(x+p^2c)$ is the norm of an element in $\mathbb{Q}_p(\sqrt{p})$. Hence $\mathcal{U}$ is locally solvable at p.

Suppose that $l \neq p$, and set $x = 2pb^2$. We see that

$$\begin{aligned} v_l(2x) &= v_l(4pb^2) = v_l(p) + 2v_l(2b) = 2v_l(2b), \\ v_l(2(x+4pb^2)(x+p^2c)) &= v_l(p^2(12b^2)(2b^2+pc)) \\ &= 2v_l(2b) + v_l(3(2b^2+pc)) = 2v_l(2b). \end{aligned}$$

Using the same arguments as in Case I, we deduce that $\mathcal{U}$ is locally solvable at l.

It is not difficult to see that $\mathcal{U}(\mathbb{R}) \neq \varnothing$. It follows from Cases I and II that $\mathcal{U}$ is everywhere locally solvable, and thus $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{T}$ are everywhere locally solvable.

Step 3. *$\mathcal{V}$ satisfies* CHP.

We will prove that $\mathcal{V}(\mathbb{A}_\mathbb{Q})^{\mathrm{Br}} = \varnothing$. Let $\mathbb{Q}(\mathcal{V})$ be the function field of $\mathcal{V}$, and let $\mathcal{A}$ be the class of quaternion algebra $(p, x+4pb^2)$ in $\mathrm{Br}(\mathbb{Q}(\mathcal{V}))$. It follows from Lemma 2.4 that $\mathcal{A}$ is an Azumaya algebra of $\mathcal{V}$. We will prove that for any $P_l \in \mathcal{V}(\mathbb{Q}_l)$,

$$\mathrm{inv}_l(\mathcal{A}(P_l)) = \begin{cases} 0 & \text{if } l \neq 2, \\ \frac{1}{2} & \text{if } l = 2. \end{cases} \tag{6}$$

Since $\mathcal{V}$ is smooth, we know that $\mathcal{U}(\mathbb{Q}_l)$ is l-adically dense in $\mathcal{V}(\mathbb{Q}_l)$. It is well-known (see, for example, [Viray 2012, Lemma 3.2]) that $\mathrm{inv}_l(\mathcal{A}(P_l))$ is a continuous function on $\mathcal{V}(\mathbb{Q}_l)$ with the l-adic topology. Hence it suffices to prove (6) for $P_l \in \mathcal{U}(\mathbb{Q}_l)$.

Suppose that $l = \infty$, or l is an odd prime such that $l \neq p$ and p is a square in $\mathbb{Q}_l^\times$. We see that $p \in \mathbb{Q}_l^{2,\times}$, and hence the local Hilbert symbol $(p, t)_l$ is 1 for any $t \in \mathbb{Q}_l^\times$. Thus $\mathrm{inv}_l(\mathcal{A}(P_l))$ is 0.

Suppose that l is an odd prime such that $l \neq p$ and p is not a square in $\mathbb{Q}_l^\times$. Let $P_l \in \mathcal{U}(\mathbb{Q}_l)$, and let $x = x(P_l)$. It follows from (3) and [Cohen 2007, loc. cit.] that $v_l(x)$ and $v_l((x+4pb^2)(x+p^2c))$ are even, and hence the sum $v_l(x+4pb^2)+v_l(x+p^2c)$ is even. Assume first that $v_l(x) < 0$. We deduce that $v_l(x+4pb^2) = v_l(x)$, and hence it is even. Suppose now that $v_l(x) \geq 0$. We then see that $v_l(x+4pb^2) \geq 0$ and $v_l(x+p^2c) \geq 0$. We contend that at least one of the last two numbers is zero. Otherwise, since $x \in \mathbb{Z}_l$, one sees that $x+4pb^2 \equiv 0 \pmod{l}$ and $x+p^2c \equiv 0 \pmod{l}$. Hence l divides $p(pc-4b^2)$, and thus by condition (B), we deduce that l divides pq.

If q is 1, then $l = p$, which is a contradiction. If q is an odd power of an odd prime, say q_1^{2m+1} for some odd prime q_1 and $m \in \mathbb{Z}_{\geq 0}$, then $l = q_1$. By condition (B), we know that $q = q_1^{2m+1} \equiv \pm 4b^2 \not\equiv 0 \pmod{p}$. Hence

$$l = q_1 \equiv \pm\left(\frac{2b}{q_1^m}\right)^2 \pmod{p}.$$

Since -1 is a square in $\mathbb{F}_p^\times$, it follows from the congruence above that l is a square in $\mathbb{F}_p^\times$. By the quadratic reciprocity law, p is a square in $\mathbb{Q}_l^\times$, which is a contradiction. Since the sum $v_l(x+4pb^2)+v_l(x+p^2c)$ is even and at least one of the two summands is even, we deduce that each of them is even. Hence, using [Cohen 2007, loc. cit.], we deduce that the local Hilbert symbol $(p, x+4pb^2)_l$ is 1. Therefore $\mathrm{inv}_l(\mathcal{A}(P_l))$ is 0.

Suppose that $l = p$. Let $P_p \in \mathcal{U}(\mathbb{Q}_p)$ and $x = x(P_p)$. Since the local Hilbert symbol $(p, 2)_p$ is -1, we deduce from (3) and [Cohen 2007, loc. cit.] that

$$(7)\qquad \begin{cases} x = p^n\alpha & \text{with } n \in \mathbb{Z}, \alpha \in \mathbb{Z}_p^\times \text{ and } \left(\frac{\alpha}{p}\right) = -1, \\ (x+4pb^2)(x+p^2c) = p^m\beta & \text{with } m \in \mathbb{Z}, \beta \in \mathbb{Z}_p^\times \text{ and } \left(\frac{\beta}{p}\right) = -1. \end{cases}$$

Assume that $n \leq 0$. We see that $p^{-n}x \equiv \alpha \pmod{p}$. Hence $p^{-n}(x+4pb^2) \equiv \alpha \pmod{p}$ and $p^{-n}(x+p^2c) \equiv \alpha \pmod{p}$. Thus the product of the two last congruences contradicts the second equation of (7). Hence, with no loss of generality, we may assume that $n \geq 1$. Assume first that $n = 1$. We deduce that $p^{-1}x \equiv \alpha \pmod{p}$, and hence $p^{-1}(x+p^2c) = p^{-1}x + pc \equiv \alpha \pmod{p}$. Thus, by (7), there exists an integer $k \in \mathbb{Z}$ such that $p^k(x+4pb^2) \equiv \beta\alpha^{-1} \pmod{p}$. We see that $\left(\frac{\beta\alpha^{-1}}{p}\right) = 1$. Hence, using [Cohen 2007, loc. cit.], we deduce that the local Hilbert symbol $(p, x+4pb^2)_p$ satisfies

$$(p, x+4pb^2)_p = \left(\frac{\beta\alpha^{-1}}{p}\right) = 1.$$

Therefore $\mathrm{inv}_p(\mathcal{A}(P_p))$ is 0.

Suppose now that $n \geq 2$. We see that

$$p^{-1}(x+4pb^2) = p^{n-1}\alpha + 4b^2 \equiv 4b^2 \pmod p.$$

Hence, using the same arguments as above, we deduce that the local Hilbert symbol $(p, x+4pb^2)_p$ is 1, and thus $\mathrm{inv}_p(\mathcal{A}(P_p))$ equals 0.

Therefore, in any event, we see that $\mathrm{inv}_p(\mathcal{A}(P_p)) = 0$.

Suppose that $l = 2$. Let $P_2 \in \mathcal{U}(\mathbb{Q}_2)$, and let $x = x(P_2)$. Since the local Hilbert symbol $(p, 2)_2$ satisfies

$$(p, 2)_2 = \left(\frac{p}{2}\right) = -1,$$

we deduce from (3) and [Cohen 2007, loc. cit.] that

$$(p, x)_2 = (p, (x+4pb^2)(x+p^2c))_2 = -1.$$

Hence $v_2(x)$ and $v_2((x+4pb^2)(x+p^2c))$ are odd. Thus $v_2(x+4pb^2)+v_2(x+p^2c)$ is odd. We contend that $v_2(x) \geq 0$. Otherwise, we deduce that

$$v_2(x+4pb^2) + v_2(x+p^2c) = 2v_2(x),$$

which is a contradiction since the left-hand side is odd whereas the right-hand side is even. Since $v_2(x)$ is odd and $v_2(x) \geq 0$, we see that $v_2(x) \geq 1$. Since c is odd, it follows that $v_2(p^2c) = 0$. Hence $v_2(x+p^2c) = v_2(p^2c) = 0$, and thus $v_2(x+4pb^2)$ is odd. Since $p \equiv 5 \pmod 8$, the local Hilbert symbol $(p, x+4pb^2)_2$ satisfies

$$(p, x+4pb^2)_2 = \left(\frac{p}{2}\right) = -1.$$

Therefore $\mathrm{inv}_2(\mathcal{A}(P_2))$ equals $\frac{1}{2}$.

Thus, in any event, $\sum_l \mathrm{inv}_l \mathcal{A}(P_l) = \frac{1}{2}$ for any $(P_l)_l \in \mathcal{V}(\mathbb{A}_\mathbb{Q})$. Thus $\mathcal{V}(\mathbb{A}_\mathbb{Q})^{\mathrm{Br}} = \varnothing$.

Step 4. *$\mathcal{U}$ and $\mathcal{T}$ satisfy* CHP.

For any point $P_l \in \mathcal{U}(\mathbb{Q}_l)$, let $x = x(P_l)$. By the definition of $\mathcal{U}$, we see that $x+4pb^2$ is nonzero. By what we have proved in Step 3, we know that the local Hilbert symbol $(p, x+4pb^2)_l$ satisfies

$$(p, x+4pb^2)_l = \begin{cases} 1 & \text{if } l \neq 2, \\ -1 & \text{if } l = 2. \end{cases}$$

Hence it follows that $x+4pb^2$ is the norm of an element of $\mathbb{Q}_l(\sqrt{p})$ for every $l \neq 2$ including $l = \infty$, and that $x+4pb^2$ is not a local norm of any element of $\mathbb{Q}_2(\sqrt{p})$. Thus we deduce that

$$\prod_l (p, x+4pb^2)_l = -1, \tag{8}$$

where the product is taken over every prime l, including $l = \infty$. Therefore it follows from the product formula [Cohen 2007, Theorem 5.3.1] that $\mathcal{U}(\mathbb{Q})$ is empty; otherwise there exists a rational point $P \in \mathcal{U}(\mathbb{Q})$. Thus the element $x + 4pb^2$ is in $\mathbb{Q}^\times$, where $x = x(P)$. Hence, by the product formula, we see that

$$\prod_l (p, x + 4pb^2)_l = 1,$$

which is a contradiction to (8). Hence $\mathcal{U}$ satisfies CHP, and it thus follows from Step 1 that $\mathcal{T}$ satisfies CHP.

Step 5. *$\mathcal{U}$, $\mathcal{V}$ and $\mathcal{T}$ satisfy* NZC.

Note that since $\mathcal{T}(\mathbb{Q}) = \varnothing$, it follows from the Amer–Brumer theorem [Amer 1976; Brumer 1978] that $\mathcal{T}$ does not contain any zero-cycle of odd degree over $\mathbb{Q}$. Thus $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{T}$ satisfy NZC, and hence our contention follows. □

The following result plays a key role in constructing algebraic families of curves satisfying CHP and NZC.

Theorem 2.7. *Let p be a prime such that $p \equiv 5 \pmod 8$. Assume (A1), and assume further that the following is true*:

(A2) *There exists a pair (b, d) of integers such that b, d are odd, $b \not\equiv 0 \pmod 3$, $b \not\equiv 0 \pmod p$ and $q := |pd^2 - 4b^2|$ is either 1 or an odd prime.*

Let $\mathcal{Z}$ be a smooth and proper $\mathbb{Q}$-model of the smooth $\mathbb{Q}$-variety $\mathcal{X}$ in $\mathbb{A}^5_{\mathbb{Q}}$ defined by

$$\mathcal{X} : \begin{cases} 0 \neq u_1^2 - pv_1^2 = 2x, \\ 0 \neq u_2^2 - pv_2^2 = 2(x + 4pb^2)(x + p^2d^2). \end{cases} \tag{9}$$

Let $\mathcal{Y} \subset \mathbb{P}^5_{\mathbb{Q}}$ be the singular $\mathbb{Q}$-variety defined by

$$\mathcal{Y} : \begin{cases} u_1^2 - pv_1^2 = 2xy, \\ u_2^2 - pv_2^2 = 2(x + 4pb^2y)(x + p^2d^2y). \end{cases} \tag{10}$$

Then $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ satisfy CHP *and* NZC.

Remark 2.8. In Section 3, we will prove that there are infinitely many triples (p, b, d) satisfying (A1) and (A2).

Proof. Let $c = d^2$. We contend that the pair (b, c) satisfies (B) in Lemma 2.4. Indeed, we note that $\gcd(b, d) = 1$; otherwise, there exists an odd prime l such that $b = lb_1$ and $d = ld_1$ for some integers $b_1, d_1 \in \mathbb{Z}$. Hence $q = l_1^2|pd_1^2 - 4b_1^2|$, which is a contradiction to (A2). Thus $\gcd(b, d) = 1$, and it follows that $\gcd(b, c) = 1$.

We know that $q = |pc - 4b^2|$ is either 1 or an odd prime, and that $b \not\equiv 0 \pmod 3$ and $b \not\equiv 0 \pmod p$. Hence the pair (b, c) satisfies (B). Thus by Theorem 2.6, we deduce that $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ satisfy CHP and NZC. □

3. Infinitude of triples (p, b, d).

In this section, we will prove that there are infinitely many triples (p, b, d) satisfying (A1) and (A2). We begin by recalling a theorem of Iwaniec's.

Let $P(x, y)$ be a quadratic polynomial in two variables x and y. We say that P *depends essentially on two variables* if $\partial P/\partial x$ and $\partial P/\partial y$ are linearly independent as elements of the $\mathbb{Q}$-vector space $\mathbb{Q}[x, y]$.

Theorem 3.1 [Iwaniec 1974, p. 443]. *Let $P(x, y) = ax^2 + bxy + cy^2 + ex + fy + g$ be a quadratic polynomial defined over $\mathbb{Q}$, and assume that the following are true*:

(i) a, b, c, e, f, g *are in* $\mathbb{Z}$ *and* $\gcd(a, b, c, e, f, g) = 1$.

(ii) $P(x, y)$ *is irreducible in* $\mathbb{Q}[x, y]$, *represents arbitrarily large odd numbers and depends essentially on two variables.*

(iii) $D = af^2 - bef + ce^2 + (b^2 - 4ac)g = 0$ *or* $\Delta = b^2 - 4ac$ *is a perfect square.*

Then

$$\frac{N}{\log N} \ll \sum_{\substack{p \le N,\ p = P(x,y) \\ p \text{ prime}}} 1.$$

We now prove the main lemma in this section.

Lemma 3.2. *Let p be a prime such that $p \equiv 5 \pmod 8$, and assume that 3 is a quadratic nonresidue in $\mathbb{F}_p^\times$. Then there are infinitely many triples (p, b, d) satisfying* (A1) *and* (A2).

Proof. The result follows immediately by applying Theorem 3.1 to

$$P(x, y) := p(2x + 1)^2 - 4(6py + b_0)^2 \in \mathbb{Q}[x, y],$$

where b_0 is an odd integer such that $\gcd(b_0, 3p) = 1$. □

Example 3.3. Let $(p, b, d) = (5, 1, 1)$. We see that the triple (p, b, d) satisfies (A1) and (A2). Let $\mathcal{Y}_{(5,1,1)}$ be the singular $\mathbb{Q}$-threefold in $\mathbb{P}^5_{\mathbb{Q}}$ defined by

$$\mathcal{Y}_{(5,1,1)} : \begin{cases} u_1^2 - 5v_1^2 = 2xy, \\ u_2^2 - 5v_2^2 = 2(x + 20y)(x + 25y). \end{cases}$$

By Theorem 2.7, $\mathcal{Y}_{(5,1,1)}$ satisfies CHP and NZC. The threefold $\mathcal{Y}_{(5,1,1)}$ is the well-known Colliot-Thélène–Coray–Sansuc threefold [Colliot-Thélène et al. 1980, p. 186, Proposition 7.1].

Example 3.4. Let $(p, b, d) = (29, 1, 3)$. We see that

$$q = |pd^2 - 4b^2| = |29 \cdot 3^2 - 4 \cdot 1^2| = 257,$$

which is an odd prime. Hence $(29, 1, 3)$ satisfies (A1) and (A2). Let $\mathcal{Y}_{(29,1,3)}$ be the singular $\mathbb{Q}$-threefold in $\mathbb{P}^5_{\mathbb{Q}}$ defined by

$$\mathcal{Y}_{(29,1,3)} : \begin{cases} u_1^2 - 29v_1^2 = 2xy, \\ u_2^2 - 29v_2^2 = 2(x+116y)(x+7569y). \end{cases}$$

By Theorem 2.7, $\mathcal{Y}_{(29,1,3)}$ satisfies CHP and NZC.

4. Hyperelliptic curves violating the Hasse principle

In this section, we give a sufficient condition under which, for each integer $n \geq 2$ and $n \not\equiv 0 \pmod 4$, there exist hyperelliptic curves of genus n that lie on the threefolds $\mathcal{Y}$ in Theorem 2.7, and satisfy CHP and NZC. The sufficient condition is in terms of the existence of certain sextuples $(p, b, d, \alpha, \beta, \gamma)$, and obtained using the geometric construction of hyperelliptic curves due to [Coray and Manoil 1996, Proposition 4.2].

Theorem 4.1. *Let p be a prime such that $p \equiv 5 \pmod 8$, and let $(p, b, d) \in \mathbb{Z}^3$ be a triple of integers satisfying (A1) and (A2). Let n be an integer such that $n \geq 2$, and let $(\alpha, \beta, \gamma) \in \mathbb{Q}^3$ be a triple of rational numbers such that $\alpha\beta\gamma \neq 0$. Assume further that the following are true*:

(A3) *We have*

$$P := p\alpha^2 + 2\beta^2 - 2p\gamma^2 \neq 0, \tag{11}$$

$$Q := 4bdp\gamma - 4b^2\beta - d^2p\beta \neq 0, \tag{12}$$

and the conic $\mathcal{Q}_1 \subset \mathbb{P}^2_{\mathbb{Q}}$, defined by

$$\mathcal{Q}_1 : pU^2 - V^2 - (\beta P Q)T^2 = 0,$$

has a point $(u, v, t) \in \mathbb{Z}^3$ with $uvt \neq 0$ and $\gcd(u, v, t) = 1$.

(S) *The polynomial $P_{p,b,d,\alpha,\beta,\gamma}(x) \in \mathbb{Q}[x]$, defined by*

$$P_{p,b,d,\alpha,\beta,\gamma}(x) := p\alpha^2Q^2x^{2n+2} + (2b^2Px^2 + \beta Q)(d^2pPx^2 + 2\beta Q),$$

is separable; that is, $P_{p,b,d,\alpha,\beta,\gamma}(x)$ has exactly $2n+2$ distinct roots in $\mathbb{C}$.

Let $\mathcal{C}$ be the smooth projective model of the affine curve defined by

$$\mathcal{C} : z^2 = p\alpha^2Q^2x^{2n+2} + (2b^2Px^2 + \beta Q)(d^2pPx^2 + 2\beta Q). \tag{13}$$

Then $\mathcal{C}(\mathbb{Q}_l) \neq \varnothing$ for every prime $l \neq 2$, p, and $\mathcal{C}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}} = \varnothing$. Furthermore, $\mathcal{C}$ satisfies NZC.

Proof. The proof follows closely that of [Coray and Manoil 1996, Proposition 4.2]. We begin by recalling the geometric construction of hyperelliptic curves due to Coray and Manoil.

Let $\mathcal{C}_a \subset \mathbb{A}^2_K$ be the affine curve defined by $z^2 = P(x)$, where $P(x)$ is a separable polynomial of degree $2n+2$ and K is a number field. Recall from [Silverman 1986, Chapter II, Exercise 2.14] that the smooth projective model of $\mathcal{C}_a$ can be described as the closure of the image of $\mathcal{C}_a$ under the mapping

$$\begin{aligned} \mathcal{C}_a &\to \mathbb{P}^{n+2}_K, \\ (x, z) &\mapsto (1, x, \dots, x^{n+1}, z). \end{aligned}$$

Following [Coray and Manoil 1996, Proposition 4.2], we will index the coordinates of $\mathbb{P}^{n+2}_K$ in such a way that z_i corresponds to x^i for $0 \le i \le n+1$ and z_{n+2} corresponds to z.

Using the above arguments, we deduce from (13) that $\mathcal{C}$ can be smoothly embedded into the intersection of quadrics defined by

$$\begin{cases} z_{n+2}^2 = p\alpha^2 Q^2 z_{n+1}^2 + (2b^2 P z_2 + \beta Q z_0)(d^2 p P z_2 + 2\beta Q z_0), \\ z_1^2 \quad = z_2 z_0. \end{cases} \tag{14}$$

Recall that $(u, v, t) \in \mathbb{Z}^3$ is the point on the conic $\mathcal{Q}_1$ defined in (A3) that is assumed to exist. Upon letting

$$z_0 = \frac{1}{\beta Q} x, \quad z_1 = \frac{t}{u} u_1, \quad z_2 = \frac{2p}{P} y, \quad z_{n+1} = \frac{1}{\alpha Q} v_2, \quad z_{n+2} = u_2,$$

we deduce from (14) that

$$\begin{cases} \dfrac{\beta P Q t^2}{p u^2} u_1^2 = 2xy, \\ u_2^2 - p v_2^2 = 2(x + 4pb^2 y)(x + p^2 d^2 y). \end{cases} \tag{15}$$

We see that (15) defines a singular del Pezzo surface $\mathcal{D} \subseteq \mathbb{P}^4_{\mathbb{Q}}$. We contend that $\mathcal{D}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}} = \varnothing$ and $\mathcal{D}$ does not contain any zero-cycle of odd degree over $\mathbb{Q}$. Indeed, upon letting

$$v_1 = \frac{v}{pu} u_1,$$

we deduce from the first equation of (15) and (A3) that

$$u_1^2 - p v_1^2 = u_1^2 - p \frac{v^2}{p^2 u^2} u_1^2 = \frac{\beta P Q t^2}{p u^2} u_1^2 = 2xy.$$

Therefore $\mathcal{D}$ is a hyperplane section of the threefold $\mathcal{Y}$ in Theorem 2.7. Hence there exists a sequence of $\mathbb{Q}$-morphisms

$$\mathcal{C} \to \mathcal{D} \to \mathcal{Y}.$$

Hence it follows from Lemma 2.1 and Theorem 2.7 that $\mathcal{D}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}} = \varnothing$. Thus $\mathcal{C}(\mathbb{A}_{\mathbb{Q}})^{\mathrm{Br}} = \varnothing$. Furthermore, since $\mathcal{Y}$ does not contain any zero-cycle of odd degree over $\mathbb{Q}$, neither do $\mathcal{C}$ and $\mathcal{D}$.

We now prove that $\mathcal{C}$ is locally solvable at primes l with $l \neq 2$, p. We consider the following cases:

Case I. $l = \infty$ or l is an odd prime such that $l \neq p$ and $\left(\frac{p}{l}\right) = 1$.

We know that the curve $\mathcal{C}^*$, defined by

$$\mathcal{C}^* : z^2 = p\alpha^2 Q^2 x^{2n+2} + y^{2n-2}(2b^2 Px^2 + \beta Qy^2)(d^2 pPx^2 + 2\beta Qy^2),$$

is an open subscheme of $\mathcal{C}$. We see that $P_\infty = (x : y : z) = (1 : 0 : \sqrt{p}\alpha Q)$ belongs to $\mathcal{C}^*(\mathbb{Q}_l) \subset \mathcal{C}(\mathbb{Q}_l)$, and hence $\mathcal{C}$ is locally solvable at l.

Case II. l is an odd prime such that $\left(\frac{2}{l}\right) = 1$.

It follows from (13) that the point $P_1 = (x, z) = (0, \sqrt{2}\beta Q)$ belongs to $\mathcal{C}(\mathbb{Q}_l)$.

Case III. l is an odd prime such that $l \neq p$ and $\left(\frac{2p}{l}\right) = 1$.

Let $F(x, z)$ be the defining polynomial of $\mathcal{C}$, defined by

$$F(x, z) = p\alpha^2 Q^2 x^{2n+2} + (2b^2 Px^2 + \beta Q)(d^2 pPx^2 + 2\beta Q) - z^2.$$

We see that

$$\begin{aligned}
&F(1, \sqrt{2p}(\gamma Q + bdP)) \\
&\quad = (p\alpha^2 Q^2 + 2p(bdP)^2 + 4b^2\beta PQ + \beta pPQd^2 + 2\beta^2 Q^2) - 2p(\gamma Q + bdP)^2.
\end{aligned}$$

Hence, it follows from (11) and (12) that

$$p\alpha^2 Q^2 + 4b^2\beta PQ + \beta pPQd^2 + 2\beta^2 Q^2 = 2p\gamma^2 Q^2 + 4p(\gamma Q)(bdP).$$

Thus

$$p\alpha^2 Q^2 + 2p(bdP)^2 + 4b^2\beta PQ + \beta pPQd^2 + 2\beta^2 Q^2 = 2p(\gamma Q + bdP)^2.$$

Hence, we deduce that $F(1, \sqrt{2p}(\gamma Q + bdP)) = 0$, and therefore the point $P_2 = (1, \sqrt{2p}(\gamma Q + bdP))$ belongs to $\mathcal{C}(\mathbb{Q}_l)$.

Thus, in any event, $\mathcal{C}$ is locally solvable at primes l with $l \neq 2$, p, which proves our contention. $\square$

Remark 4.2. Theorem 4.1 constructs hyperelliptic curves of genus at least two such that they satisfy NZC and all conditions in CHP except the local solvability at 2 and p. The rest of this section presents certain sufficient conditions for which those hyperelliptic curves arising from Theorem 4.1 are locally solvable at 2 and p, and hence satisfy CHP and NZC.

Lemma 4.3. *Let p be a prime such that $p \equiv 5 \pmod 8$, and let $(b, d) \in \mathbb{Z}^3$ be a pair of integers satisfying (A1) and (A2). Assume that there is a triple $(\alpha, \beta, \gamma) \in \mathbb{Q}^3$ satisfying (A3) in Theorem 4.1, and assume further that $\alpha, \beta, \gamma \in \mathbb{Z}_p$. Then there is a rational number $\bar{\beta} \in \mathbb{Q}$ such that $\beta = p\bar{\beta}$ and $\bar{\beta} \in \mathbb{Z}_p$.*

Proof. Let $\mathcal{Q}_1$ be the conic defined in (A3). Assume that $(u, v, t) \in \mathbb{Z}^3$ belongs to $\mathcal{Q}_1(\mathbb{Q})$ such that $uvt \neq 0$ and $\gcd(u, v, t) = 1$. We see that

$$pu^2 - v^2 - \beta P Q t^2 = 0,$$

where P and Q are defined by (11) and (12), respectively. Taking the identity above modulo p, it follows that

$$v^2 \equiv 8b^2\beta^4t^2 \pmod p.$$

Since 2 is a quadratic nonresidue in $\mathbb{F}_p^\times$ and $b \not\equiv 0 \pmod p$, we deduce from the congruence above that

$$v \equiv \beta t \equiv 0 \pmod p.$$

Assume that $\beta \not\equiv 0 \pmod p$. Then $v \equiv t \equiv 0 \pmod p$, and hence $v = pv_1$ and $t = pt_1$ for some integers v_1, t_1. Substituting v and t into the defining equation of the conic $\mathcal{Q}_1$, we get

$$u^2 - pv_1^2 - p\beta P Q t_1^2 = 0,$$

and hence it follows that p divides u. Thus p divides $\gcd(u, v, t)$, which is a contradiction. Therefore there is a rational number $\bar{\beta} \in \mathbb{Q}$ such that $\beta = p\bar{\beta}$ and $\bar{\beta} \in \mathbb{Z}_p$. □

Remark 4.4. By Lemma 4.3, one knows that if $(\alpha, \beta, \gamma) \in \mathbb{Q}^3$ satisfies (A3) and $\alpha, \beta, \gamma \in \mathbb{Z}_p$, then there is a rational number $\bar{\beta}$ such that $\beta = p\bar{\beta}$ and $\bar{\beta} \in \mathbb{Z}_p$. Hence one sees that $P = pP_1$ and $Q = pQ_1$, where

$$P_1 := \alpha^2 + 2p\bar{\beta}^2 - 2\gamma^2,$$
$$Q_1 := 4bd\gamma - 4b^2\bar{\beta} - d^2p\bar{\beta}.$$

We also see that P_1 and Q_1 belong to $\mathbb{Z}_p$.

In the proofs of Corollaries 4.6 and 4.8 below, we will use Hensel's lemma to deduce the local solvability at primes 2 and p. For the sake of self-containedness, we recall the statement of Hensel's lemma.

Theorem 4.5 [Borevich and Shafarevich 1966, Section 5.2, Theorem 3]. *Let p be a prime. Let $F(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_p[x_1, x_2, \ldots, x_n]$ be a polynomial whose* *coefficients are p-adic integers. Let δ be a nonnegative integer. Assume that there are p-adic integers $a_1, a_2, \dots, a_n$ such that for some integer $1 \le k \le n$, we have*

$$\begin{aligned} F(a_1, a_2, \dots, a_n) &\equiv 0 \pmod{p^{2\delta+1}}, \\ \frac{\partial F}{\partial x_k}(a_1, a_2, \dots, a_n) &\equiv 0 \pmod{p^{\delta}}, \\ \frac{\partial F}{\partial x_k}(a_1, a_2, \dots, a_n) &\not\equiv 0 \pmod{p^{\delta+1}}. \end{aligned}$$

Then there exist p-adic integers $\theta_1, \theta_2, \dots, \theta_n$ such that $F(\theta_1, \theta_2, \dots, \theta_n) = 0$.

The following result provides a sufficient condition under which certain hyperelliptic curves of odd genus satisfy CHP and NZC.

Corollary 4.6. *We maintain the same notation and assumptions as in Theorem 4.1. Assume (A1)–(A3) and (S). Assume further that the following are true:*

(A4) $\alpha, \beta, \gamma \in \mathbb{Z}_2^\times$, $\alpha, \gamma, d \in \mathbb{Z}_p^\times$ *and* $\beta \in \mathbb{Z}_p$.

(A5) $\gamma Q_1 + bd P_1 \equiv 0 \pmod{p^2}$, *where $\bar{\beta}$, P_1 and Q_1 are defined as in Remark 4.4.*

(A6) $n \not\equiv -2(\gamma/\alpha)^2 \pmod{p}$, *$n \ge 3$ and n is odd.*

Let $\mathcal{C}$ be the smooth projective model of the affine curve defined by (13). Then $\mathcal{C}$ satisfies CHP *and* NZC.

Proof. By Theorem 4.1, it suffices to prove that $\mathcal{C}$ is locally solvable at 2 and p.

Step 1. *$\mathcal{C}$ is locally solvable at p.*

We will use Theorem 4.5 with the exponent $\delta = 3$ to prove the local solvability of $\mathcal{C}$ at p. We consider the system of equations

$$(16)\quad \begin{cases} F(x, z) = p\alpha^2 Q^2 x^{2n+2} + (2b^2 P x^2 + \beta Q)(d^2 p P x^2 + 2\beta Q) - z^2 \\ \qquad\qquad \equiv 0 \pmod{p^7}, \\ \dfrac{\partial F}{\partial x}(x, z) = (2n+2) p\alpha^2 Q^2 x^{2n+1} + 4b^2 P x (d^2 p P x^2 + 2\beta Q) \\ \qquad\qquad\qquad\qquad\qquad\qquad + 2d^2 p P x (2b^2 P x^2 + \beta Q) \\ \qquad\qquad \equiv 0 \pmod{p^3}, \\ \dfrac{\partial F}{\partial x}(x, z) \not\equiv 0 \pmod{p^4}. \end{cases}$$

Repeating the same arguments as in Case III of the proof of Theorem 4.1, we deduce that

$$F(1, 0) = 2p(\gamma Q + bd P)^2.$$

By Remark 4.4, one knows that $P = pP_1$ and $Q = pQ_1$. Hence

$$(17)\quad F(1, 0) = 2p^3(\gamma Q_1 + bd P_1)^2.$$

Thus it follows from (A5) and (17) that $F(1,0) \equiv 0 \pmod{p^7}$. On the other hand, we see that

$$\frac{\partial F}{\partial x}(1,0) = p^3\big((2n+2)\alpha^2 Q_1^2 + 4b^2 P_1(d^2 P_1 + 2\bar{\beta} Q_1) + 2d^2 P_1(2b^2 P_1 + p\bar{\beta} Q_1)\big). \tag{18}$$

Since α, $\bar{\beta}$, γ and P_1, Q_1 are in $\mathbb{Z}_p$, one obtains that

$$\frac{\partial F}{\partial x}(1,0) \equiv 0 \pmod{p^3}.$$

Assume that

$$\frac{1}{p^3}\Big(\frac{\partial F}{\partial x}(1,0)\Big) \equiv 0 \pmod{p}. \tag{19}$$

Since $\gamma \in \mathbb{Z}_p^\times$, it follows from (A5) that

$$Q_1 \equiv -\frac{bd}{\gamma} P_1 \pmod{p}. \tag{20}$$

Upon replacing Q_1 by $-(bd/\gamma)P_1$ in (19), we deduce that

$$\frac{2P_1^2 bd}{\gamma^2}\big((n+1)\alpha^2 bd + \gamma(4bd\gamma - 4b^2\bar{\beta} - d^2 p\bar{\beta})\big) \equiv 0 \pmod{p}.$$

Thus it follows from the definition of Q_1 in Remark 4.4 that

$$\frac{2P_1^2 bd}{\gamma^2}((n+1)\alpha^2 bd + \gamma Q_1) \equiv 0 \pmod{p}.$$

Note that $P_1 \in \mathbb{Z}_p^\times$; otherwise, we deduce from the definition of P_1 in Remark 4.4 that

$$\alpha^2 - 2\gamma^2 \equiv P_1 \equiv 0 \pmod{p}.$$

Since $\alpha, \gamma \in \mathbb{Z}_p^\times$, it follows from the congruence above that $2 \equiv (\alpha/\gamma)^2 \pmod{p}$, which is a contradiction to the fact that $p \equiv 5 \pmod{8}$. Thus $P_1 \in \mathbb{Z}_p^\times$. Since $2, b, d, \gamma$ and P_1 are in $\mathbb{Z}_p^\times$, we obtain that

$$(n+1)\alpha^2 bd + \gamma Q_1 \equiv 0 \pmod{p}.$$

Since $\gamma Q_1 \equiv -bdP_1 \pmod{p}$ and $b, d \in \mathbb{Z}_p^\times$, we deduce from this congruence that

$$(n+1)\alpha^2 \equiv P_1 \equiv (\alpha^2 - 2\gamma^2) \pmod{p}.$$

Since $\alpha, \gamma \in \mathbb{Z}_p^\times$, it follows that $n \equiv -2(\gamma/\alpha)^2 \pmod{p}$, which is a contradiction to (A6). Thus the system (16) has a solution $(x, z) = (1, 0)$. By Hensel's lemma, $\mathcal{C}$ is locally solvable at p.

Step 2. *$\mathcal{C}$ is locally solvable at* 2.

We will use Theorem 4.5 with the exponent $\delta = 1$ to prove the local solvability of $\mathcal{C}$ at 2. We consider the system of equations

$$\begin{cases} F(x,z) \equiv 0 \pmod{2^3}, \\ \dfrac{\partial F}{\partial x}(x,z) \equiv 0 \pmod{2}, \\ \dfrac{\partial F}{\partial x}(x,z) \not\equiv 0 \pmod{2^2}. \end{cases} \tag{21}$$

We see from (17) and the definitions of P_1 and Q_1 that

$$F(1,0) = 2p^3\big(\gamma(4bd\gamma - 4b^2\bar{\beta} - d^2p\bar{\beta}) + bd(\alpha^2 + 2p\bar{\beta}^2 - 2\gamma^2)\big)^2.$$

Since β is in $\mathbb{Z}_2^\times$ and $p \neq 2$, we see that $\bar{\beta}$ is also in $\mathbb{Z}_2^\times$. Since $b, d, p, \alpha, \bar{\beta}, \gamma \in \mathbb{Z}_2^\times$, we see that

$$-d^2p\bar{\beta}\gamma + bd\alpha^2 \equiv 0 \pmod{2}.$$

Let v_2 denote the 2-adic valuation. We see that

$$\begin{aligned} &v_2\big(\gamma(4bd\gamma - 4b^2\bar{\beta} - d^2p\bar{\beta}) + bd(\alpha^2 + 2p\bar{\beta}^2 - 2\gamma^2)\big) \\ &\quad = v_2\big((4\gamma(bd\gamma - b^2\bar{\beta}) + 2bd(p\bar{\beta}^2 - \gamma^2)) + (-d^2p\bar{\beta}\gamma + bd\alpha^2)\big) \\ &\quad \geq \min\big(v_2(4\gamma(bd\gamma - b^2\bar{\beta}) + 2bd(p\bar{\beta}^2 - \gamma^2)), v_2(-d^2p\bar{\beta}\gamma + bd\alpha^2)\big) \geq 1. \end{aligned}$$

Hence $F(1,0) \equiv 0 \pmod{2^3}$. On the other hand, we know from (18) that

$$\frac{\partial F}{\partial x}(1,0) \equiv 0 \pmod{2}.$$

Since n is odd, $(2n+2) \equiv 2(n+1) \equiv 0 \pmod{2^2}$. Hence, it follows from (18) that

$$\frac{\partial F}{\partial x}(1,0) \equiv 2d^2p^4\bar{\beta}P_1Q_1 \pmod{2^2}.$$

By (A4) and the definitions of P_1 and Q_1, we know that

$$d^2p^4\bar{\beta}P_1Q_1 \not\equiv 0 \pmod{2}.$$

Hence we deduce that $(\partial F/\partial x)(1,0) \not\equiv 0 \pmod{2^2}$. Thus the system (21) has a solution $(x,z) = (1,0)$. By Hensel's lemma, $\mathcal{C}$ is locally solvable at 2, and hence our contention follows. □

Remark 4.7. Assume (A1)–(A3), (A5) and (S). Following closely the proof of Corollary 4.6, we note that the following are true:

(1) If $\alpha, \beta, \gamma \in \mathbb{Z}_2^\times$ and n is odd, then $\mathcal{C}$ is locally solvable at 2.
(2) If $\alpha, \gamma, d \in \mathbb{Z}_p^\times$, $\beta \in \mathbb{Z}_p$, $n \geq 2$ and $n \not\equiv -2(\gamma/\alpha)^2 \pmod{p}$, then $\mathcal{C}$ is locally solvable at p.

We now prove a sufficient condition under which certain hyperelliptic curves of genus $n \equiv 2 \pmod{4}$ satisfy CHP and NZC.

Corollary 4.8. *We maintain the same notation as in Theorem 4.1 and Corollary 4.6. Assume (A1)–(A5) and (S). Assume further that the following are true*:

(B1) $bd - \bar{\beta}\gamma \equiv 0 \pmod 4$.

(B2) $n \not\equiv -2(\gamma/\alpha)^2 \pmod p$*,* $n \geq 2$ *and* $n \equiv 2 \pmod 4$.

Let $\mathcal{C}$ be the smooth projective model defined by (13). Then $\mathcal{C}$ satisfies CHP *and* NZC.

Proof. By Theorem 4.1 and Remark 4.7, it suffices to prove that $\mathcal{C}$ is locally solvable at 2. We will use Theorem 4.5 with the exponent $\delta = 2$ to prove the local solvability of $\mathcal{C}$ at 2. We consider the system of equations

$$\begin{cases} F(x,z) \equiv 0 \pmod{2^5}, \\ \dfrac{\partial F}{\partial x}(x,z) \equiv 0 \pmod{2^2}, \\ \dfrac{\partial F}{\partial x}(x,z) \not\equiv 0 \pmod{2^3}, \end{cases} \tag{22}$$

where $F(x,z)$ denotes the polynomial in the variables x, z defined in (16). Since $\alpha \in \mathbb{Z}_2^\times$, we know that $\alpha \equiv 1 \pmod 4$ or $\alpha \equiv 3 \pmod 4$. Hence $\alpha^2 \equiv 1 \pmod 4$. Similarly we know that $\bar{\beta}^2, \gamma^2, b^2, d^2 \equiv 1 \pmod 4$. Since $p \equiv 5 \pmod 8$, it follows that

$$\begin{aligned} P_1 &\equiv 1 \pmod 4, \\ Q_1 &\equiv -\bar{\beta} \pmod 4. \end{aligned}$$

By (B1), we know that

$$\gamma Q_1 + bdP_1 \equiv bd - \bar{\beta}\gamma \equiv 0 \pmod 4,$$

and hence we deduce from (17) that $F(1,0) \equiv 0 \pmod{2^5}$.

Since $n \equiv 2 \pmod 4$, there is a nonnegative integer l such that $n = 4l + 2$. We know that

$$\begin{aligned} 4b^2P_1(d^2P_1 + 2\bar{\beta}Q_1) &+ 2d^2P_1(2b^2P_1 + p\bar{\beta}Q_1) \\ &= 8b^2d^2P_1^2 + 8b^2\bar{\beta}P_1Q_1 + 2pd^2\bar{\beta}P_1Q_1. \end{aligned}$$

Hence, it follows from (18) that

$$\frac{\partial F}{\partial x}(1,0) \equiv 2\alpha^2Q_1^2 + 2pd^2\bar{\beta}P_1Q_1 \equiv 2 - 2\bar{\beta}^2 \equiv 0 \pmod{2^2}.$$

Similarly, one sees that

$$\frac{\partial F}{\partial x}(1,0) \equiv 5(8l+6)\alpha^2Q_1^2 + 10pd^2\bar{\beta}P_1Q_1 \pmod{2^3}.$$

Since $\alpha, \bar{\beta}, \gamma, b, d \in \mathbb{Z}_2^\times$, we deduce that $\alpha^2, \bar{\beta}^2, \gamma^2, b^2, d^2 \equiv 1 \pmod{2^3}$. Since $p \equiv 5 \pmod{2^3}$ and $bd\gamma - b^2\bar{\beta} \equiv 0 \pmod 2$, it follows from the definitions of P_1 and Q_1 that

$$\begin{aligned} P_1 &\equiv 1 \pmod{2^3}, \\ Q_1 &\equiv 4(bd\gamma - b^2\bar{\beta}) - 5\bar{\beta} \equiv -5\bar{\beta} \pmod{2^3}. \end{aligned}$$

Thus we see that

$$\frac{\partial F}{\partial x}(1,0) \equiv 30 - 250\bar{\beta}^2 \equiv 4 \not\equiv 0 \pmod{2^3}.$$

Therefore the system (22) has a solution $(x, z) = (1, 0)$. By Hensel's lemma, $\mathcal{C}$ is locally solvable at 2, which proves our contention. □

5. Infinitude of sextuples $(p, b, d, \alpha, \beta, \gamma)$

By Corollaries 4.6 and 4.8, we know that in order to construct algebraic families of hyperelliptic curves satisfying CHP and NZC, we need to find certain sextuples of rational functions in $\mathbb{Q}(T)$ that parametrize sextuples $(p, b, d, \alpha, \beta, \gamma)$ satisfying (A1)–(A5), (S) and (B1). In this section, we show how to produce infinitely many sextuples $(p, b, d, \alpha, \beta, \gamma)$ satisfying (A1)–(A5) and (B1) from the known ones.

Lemma 5.1. *Let (p, b, d) be a triple of integers satisfying (A1) and (A2). Assume that there is a triple $(\alpha_0, \beta_0, \gamma_0) \in \mathbb{Q}^3$ satisfying (A3)–(A5) and (B1). Let $(u_0, v_0, t_0) \in \mathbb{Z}^3$ be a point on the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ such that $u_0v_0t_0 \neq 0$ and* $\gcd(u_0, v_0, t_0) = 1$, *where the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ is defined by*

$$\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)} : pU^2 - V^2 - \beta_0 P_0 Q_0 T^2 = 0$$

with

$$\begin{aligned} P_0 &= p\alpha_0^2 + 2\beta_0^2 - 2p\gamma_0^2, \\ Q_0 &= 4bdp\gamma_0 - 4b^2\beta_0 - d^2 p\beta_0. \end{aligned}$$

Let $A, B \in \mathbb{Q}$ be rational numbers, and assume that the following are true:

(C1) *$A, B \in \mathbb{Z}_2$ and $B^2 - pA^2 \in \mathbb{Z}_2^\times$.*

(C2) *$A \in \mathbb{Z}_p$ and $B \in \mathbb{Z}_p^\times$.*

(C3) *$u := u_0 + AC \neq 0$ and $v := v_0 + BC \neq 0$, where*

$$C := \frac{2pu_0A - 2v_0B - 4p^3\alpha_0\beta_0 t_0^2 Q_0}{B^2 - pA^2 + 4p^5\beta_0 t_0^2 Q_0}. \tag{23}$$

Define

$$\alpha := \alpha_0 + 2p^2C, \quad \beta := \beta_0, \quad \gamma := \gamma_0.$$

Then the triple $(\alpha, \beta, \gamma) \in \mathbb{Q}^3$ satisfies (A3)–(A5) and (B1).

Remark 5.2. In order to use Theorem 4.1 to show the existence of algebraic families of hyperelliptic curves satisfying CHP and NZC, one of the crucial steps is to describe a parametrization of triples (α, β, γ) such that the conics associated to these triples in (A3) has a nontrivial rational point. Assuming the existence of one triple $(\alpha_0, \beta_0, \gamma_0)$ satisfying (A3)–(A5) and (B1), Lemma 5.1 shows how to construct families of triples (α, β, γ) satisfying the same conditions as the triple $(\alpha_0, \beta_0, \gamma_0)$.

Proof. We first prove that (α, β, γ) satisfies (A4). Since $A \in \mathbb{Z}_p$, $B \in \mathbb{Z}_p^\times$ and the triple $(\alpha_0, \beta_0, \gamma_0)$ satisfies (A4), it follows that $B^2 - pA^2 + 4p^5\beta_0 t_0^2 Q_0 \in \mathbb{Z}_p^\times$. Hence by (23) and (C2), we see that $C \in \mathbb{Z}_p$. Thus $\alpha = \alpha_0 + 2p^2 C \in \mathbb{Z}_p$. Hence it follows that

$$\alpha \equiv \alpha_0 \not\equiv 0 \pmod{p},$$

which proves that $\alpha \in \mathbb{Z}_p^\times$. By assumption, one knows that the triple $(\alpha_0, \beta_0, \gamma_0)$ satisfies (A4). Since $\beta = \beta_0$ and $\gamma = \gamma_0$, we deduce that $\beta, \gamma \in \mathbb{Z}_2^\times$, $\beta \in \mathbb{Z}_p$ and $\gamma, d \in \mathbb{Z}_p^\times$. Hence it remains to prove that $\alpha \in \mathbb{Z}_2^\times$. By assumptions and (C1), we know that $Q_0 \in \mathbb{Z}_2$ and $B^2 - pA^2 \in \mathbb{Z}_2^\times$. Hence it follows that

$$B^2 - pA^2 + 4p^5\beta_0 t_0^2 Q_0 \equiv B^2 - pA^2 \not\equiv 0 \pmod{2}.$$

Thus $B^2 - pA^2 + 4p^5\beta_0 t_0^2 Q_0 \in \mathbb{Z}_2^\times$, and hence we deduce that $C \in \mathbb{Z}_2$. Thus,

$$\alpha = \alpha_0 + 2p^2 C \equiv \alpha_0 \not\equiv 0 \pmod{2}.$$

Therefore $\alpha \in \mathbb{Z}_2^\times$, and hence (α, β, γ) satisfies (A4).

Now we prove that (α, β, γ) satisfies (A3). By what we have proved above, we know that $\alpha, \beta, \gamma \in \mathbb{Z}_2^\times$. This implies that $\alpha, \beta, \gamma \neq 0$. Let P and Q be the rational numbers defined by (11) and (12), respectively. One knows that $Q = Q_0 \neq 0$. Since $\alpha, \beta, \gamma \in \mathbb{Z}_2^\times$, it follows that $P \in \mathbb{Z}_2$. Hence we deduce that

$$P \equiv p\alpha^2 \not\equiv 0 \pmod{2},$$

which proves that $P \in \mathbb{Z}_2^\times$. Note that $P \neq 0$ since $P \in \mathbb{Z}_2^\times$.

Let $\mathcal{Q}_1 \subset \mathbb{P}^2_{\mathbb{Q}}$ be the conic defined by

$$\mathcal{Q}_1 : pU^2 - V^2 - \beta P Q T^2 = 0.$$

We prove that the point $\boldsymbol{P} := (u, v, t) \in \mathbb{Q}^3$ belongs to $\mathcal{Q}_1(\mathbb{Q})$, where u and v are defined in (C3) and $t := t_0$. Indeed, since $\beta = \beta_0$, $\gamma = \gamma_0$ and $Q = Q_0$, we deduce from (11) that

$$\begin{aligned} -\beta P Q t^2 &= -\beta_0 t_0^2 Q_0 (p(\alpha_0 + 2p^2 C)^2 + 2\beta_0^2 - 2p\gamma_0^2) \\ &= -(4p^5\beta_0 t_0^2 Q_0)C^2 - (4p^3\alpha_0\beta_0 t_0^2 Q_0)C - (\beta_0 P_0 Q_0)t_0^2. \end{aligned}$$

Hence

$$\begin{aligned}pu^2-v^2-\beta PQt^2 &= p(u_0+AC)^2-(v_0+BC)^2-(4p^5\beta_0t_0^2Q_0)C^2\\&\qquad -(4p^3\alpha_0\beta_0t_0^2Q_0)C-(\beta_0P_0Q_0)t_0^2\\&=(pA^2-B^2-4p^5\beta_0t_0^2Q_0)C^2\\&\quad +(2pu_0A-2v_0B-4p^3\alpha_0\beta_0t_0^2Q_0)C+(pu_0^2-v_0^2-\beta_0P_0Q_0t_0^2).\end{aligned}$$

Since (u_0, v_0, t_0) belongs to $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}(\mathbb{Q})$, we see that

$$pu_0^2-v_0^2-\beta_0P_0Q_0t_0^2=0.$$

Hence it follows from (23) that

$$\begin{aligned}&pu^2-v^2-\beta PQt^2\\&\qquad=(pA^2-B^2-4p^5\beta_0t_0^2Q_0)C^2+(2pu_0A-2v_0B-4p^3\alpha_0\beta_0t_0^2Q_0)C=0.\end{aligned}$$

Thus $\boldsymbol{P}\in\mathcal{Q}_1(\mathbb{Q})$. Since $\mathcal{Q}_1$ is a nonsingular conic in $\mathbb{P}^2_{\mathbb{Q}}$, $\mathcal{Q}_1(\mathbb{Q})\neq\varnothing$ and $uvt\neq 0$, it follows that (α,β,γ) satisfies (A3).

We now prove that (α,β,γ) satisfies (A5). Indeed, we have shown that (α,β,γ) satisfies (A3), (A4). This implies that $\alpha,\beta,\gamma\in\mathbb{Z}_p$. By Lemma 4.3, we know that there is a rational number $\bar\beta\in\mathbb{Q}$ such that $\beta=p\bar\beta$ and $\bar\beta\in\mathbb{Z}_p$. Similarly, since $(\alpha_0,\beta_0,\gamma_0)$ satisfies (A3) and (A4), there is a rational number $\bar\beta_0$ such that $\beta_0=p\bar\beta_0$ and $\bar\beta_0\in\mathbb{Z}_p$. Since $\beta=\beta_0$, we deduce that $\bar\beta=\bar\beta_0$.

Let P_1 and Q_1 be the rational numbers defined in Remark 4.4 and let $P_1^{(0)}$ and $Q_1^{(0)}$ be the rational numbers defined by the same equations as P_1, Q_1 with $(\alpha_0,\bar\beta_0,\gamma_0)$ in the role of $(\alpha,\bar\beta,\gamma)$. By assumption, one knows that the triple $(\alpha_0,\beta_0,\gamma_0)$ satisfies (A5), that is,

$$\gamma_0Q_1^{(0)}+bdP_1^{(0)}\equiv 0 \pmod{p^2}.$$

We will prove that

$$\gamma Q_1+bdP_1\equiv 0 \pmod{p^2}.$$

Indeed, one can check that

$$P_1=\alpha^2+2p\bar\beta^2-2\gamma^2=4p^4C^2+4p^2\alpha_0C+P_1^{(0)}$$

and $Q_1=Q_1^{(0)}$. Since $\alpha,\bar\beta,\gamma$ are in $\mathbb{Z}_p$, we deduce that $P_1\in\mathbb{Z}_p$. Recall that $C\in\mathbb{Z}_p$. Hence

$$P_1=4p^4C^2+4p^2\alpha_0C+P_1^{(0)}\equiv P_1^{(0)} \pmod{p^2},$$

and thus we deduce that

$$\gamma Q_1+bdP_1\equiv\gamma_0Q_1^{(0)}+bdP_1^{(0)}\equiv 0 \pmod{p^2}.$$

Therefore (α, β, γ) satisfies (A5).

Finally, since $(\alpha_0, \beta_0, \gamma_0)$ satisfies (B1), we see that

$$bd - \bar{\beta}\gamma = bd - \bar{\beta}_0\gamma_0 \equiv 0 \pmod 4.$$

Thus (α, β, γ) satisfies (B1), which proves our contention. □

Lemma 5.3. *Let (p, b, d) be a triple of integers satisfying (A1) and (A2). Assume that there is a triple $(\alpha_0, \beta_0, \gamma_0) \in \mathbb{Q}^3$ satisfying (A3)–(A5) and (B1). Let $(u_0, v_0, t_0) \in \mathbb{Z}^3$ be a point on the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ such that $u_0v_0t_0 \neq 0$ and $\gcd(u_0, v_0, t_0) = 1$, where P_0, Q_0 and the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ are defined as in Lemma 5.1. Let $\boldsymbol{I}$ be the set defined by*

$$\boldsymbol{I} := \{(A, B) \in \mathbb{Q}^2 : (A, B) \text{ satisfies (C1)–(C3) in Lemma 5.1}\}.$$

Then $\boldsymbol{I}$ is of infinite cardinality.

Proof. Let B_0 be an integer such that $\gcd(B_0, 2p) = 1$. For each $x \in \mathbb{Z}$, define $B = 2px + B_0$. We see that $B \in \mathbb{Z}_2^\times$ and $B \in \mathbb{Z}_p^\times$. The latter implies that $B \neq 0$. Let $A = 0$, and let C be the rational number defined by (23). Define

$$u := u_0 + AC = u_0,$$

$$v := v_0 + BC.$$

By assumption, we know that $u = u_0 \neq 0$. Assume that $v = 0$. Since $B \neq 0$, it follows from (23) and the definition of v that

$$C = -\frac{v_0}{B} = \frac{-2v_0B - 4p^3\alpha_0\beta_0t_0^2Q_0}{B^2 + 4p^5\beta_0t_0^2Q_0}.$$

Hence we deduce that B is a zero of the quadratic polynomial $\mathcal{B}(T) \in \mathbb{Q}[T]$, where $\mathcal{B}(T)$ is defined by

$$\mathcal{B}(T) := v_0T^2 + (4p^3\alpha_0\beta_0t_0^2Q_0)T - 4p^5\beta_0v_0t_0^2Q_0. \tag{24}$$

Hence, upon letting T_1 and T_2 be the zeros of $\mathcal{B}(T)$, we deduce that $(0, B)$ satisfies (C3) if and only if $B \neq T_1$ and $B \neq T_2$. The latter holds if and only if $x \neq (T_1 - B_0)/(2p)$ and $x \neq (T_2 - B_0)/(2p)$. This implies that if $T_1, T_2 \notin \mathbb{Z}$, then $(0, B)$ automatically satisfies (C3) for any integer $x \in \mathbb{Z}$. Furthermore we see that $B^2 - pA^2 = B^2 \in \mathbb{Z}_2^\times$. Hence $(0, B)$ satisfies (C1) and (C2). Thus $\boldsymbol{J}$ is a subset of $\boldsymbol{I}$, where $\boldsymbol{J}$ is defined by

$$\boldsymbol{J} := \left\{(0, B) : x \in \mathbb{Z}, x \neq \frac{T_1 - B_0}{2p} \text{ and } x \neq \frac{T_2 - B_0}{2p}\right\}.$$

Since $\boldsymbol{J}$ is of infinite cardinality, so is $\boldsymbol{I}$. Hence our contention follows. □

Using Lemmas 5.1 and 5.3, we prove the main result in this section.

Lemma 5.4. *There are infinitely many sextuples $(p, b, d, \alpha, \beta, \gamma)$ satisfying (A1)–(A5) and (B1).*

Proof. Assume that there is a sextuple $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfying (A1)–(A5) and (B1). Let $(u_0, v_0, t_0) \in \mathbb{Z}^3$ be a point on the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ such that $u_0 v_0 t_0 \neq 0$ and $\gcd(u_0, v_0, t_0) = 1$, where P_0, Q_0 and the conic $\mathcal{Q}_1^{(\alpha_0,\beta_0,\gamma_0)}$ are defined as in Lemma 5.1. Let $\boldsymbol{J}$ be the set defined in the proof of Lemma 5.3. We construct an infinite sequence $(0, B_n)_{n\in\mathbb{Z}_{\geq 0}}$ of elements of $\boldsymbol{J}$ as follows.

Let $(0, B_1)$ be an arbitrary element of $\boldsymbol{J}$, and assume that the elements $(0, B_i)$ of $\boldsymbol{J}$ with $1 \leq i \leq n$ are already constructed. Since $\boldsymbol{J}$ is infinite, we can choose an element $(0, B_{n+1})$ of $\boldsymbol{J}$ such that $B_{n+1} \neq B_i$ for $1 \leq i \leq n$ and B_{n+1} is not a zero of any of the polynomials $\boldsymbol{H}_i(T)$ for $1 \leq i \leq n$, where for each $1 \leq i \leq n$,

$$\boldsymbol{H}_i(T) = (v_0 B_i + 2p^3\alpha_0\beta_0 t_0^2 Q_0)T + 2p^3\alpha_0\beta_0 t_0^2 Q_0 B_i - 4p^5\beta_0 v_0 t_0^2 Q_0 \in \mathbb{Q}[T]. \tag{25}$$

Indeed, we see that $2p^3\alpha_0\beta_0 t_0^2 Q_0 B_i - 4p^5\beta_0 v_0 t_0^2 Q_0 \neq 0$ for every $1 \leq i \leq n$; otherwise, there is an integer $1 \leq i \leq n$ such that

$$\alpha_0 B_i = 2p^2 v_0.$$

Hence $\alpha_0 B_i \notin \mathbb{Z}_p^\times$, which is a contradiction since α_0 and B_i are in $\mathbb{Z}_p^\times$. Hence $\boldsymbol{H}_i(T)$ is nonzero and of degree at most 1 for each $1 \leq i \leq n$. Thus $\boldsymbol{H}_i(T)$ has at most one zero in $\mathbb{Z}$ for each $1 \leq i \leq n$; hence, excluding these n zeros (if existing) and the integers B_i for $1 \leq i \leq n$ out of the infinite set $\boldsymbol{J}$, one can choose an element $(0, B_{n+1})$ as desired. Therefore we have inductively constructed an infinite sequence $\{(0, B_n)\}_{n\geq 1}$ of elements of $\boldsymbol{J}$. We contend that for any two distinct members $(0, B_m)$ and $(0, B_n)$ of the sequence with $m < n$, the triples $(\alpha_m, \beta_0, \gamma_0)$ and $(\alpha_n, \beta_0, \gamma_0)$ are distinct, that is, $\alpha_m \neq \alpha_n$, where

$$\alpha_m := \alpha_0 + 2p^2 C_{(m)}, \quad \alpha_n := \alpha_0 + 2p^2 C_{(n)},$$

and $C_{(m)}$, $C_{(n)}$ are defined as in (23) with $(0, B_m)$ and $(0, B_n)$ in the role of (A, B), respectively. Assume the contrary, that is, $\alpha_m = \alpha_n$. It follows that

$$\frac{-2v_0 B_m - 4p^3\alpha_0\beta_0 t_0^2 Q_0}{B_m^2 + 4p^5\beta_0 t_0^2 Q_0} = C_{(m)} = C_{(n)} = \frac{-2v_0 B_n - 4p^3\alpha_0\beta_0 t_0^2 Q_0}{B_n^2 + 4p^5\beta_0 t_0^2 Q_0}.$$

Hence we deduce that

$$2(B_n - B_m)\big((v_0 B_m + 2p^3\alpha_0\beta_0 t_0^2 Q_0)B_n + 2p^3\alpha_0\beta_0 t_0^2 Q_0 B_m - 4p^5\beta_0 v_0 t_0^2 Q_0\big) = 0.$$

Since $B_n \neq B_m$, we deduce that B_n is a zero of $\boldsymbol{H}_m(T)$, where $\boldsymbol{H}_m(T)$ is defined by (25), which is a contradiction to the choice of B_n. Thus we have shown that there are infinitely many sextuples $(p, b, d, \alpha, \beta, \gamma)$ satisfying (A1)–(A5) and (B1) provided that there exists one sextuple $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfying (A1)–(A5) and (B1). On the other hand, in the proof of Theorem 6.8(i) below, we will show that the sextuple $(p, b, d, \alpha_0, \beta_0, \gamma_0) = (29, 1, 3, 7, 261, 15)$ satisfies (A1)–(A5) and (B1), and hence our contention follows. $\square$

6. Algebraic families of hyperelliptic curves violating the Hasse principle

Let n be an integer such that $n > 5$ and $n \not\equiv 0 \pmod 4$. In this section, using the results in the last section, we will show how to construct algebraic families of hyperelliptic curves of genus n satisfying CHP and NZC. We begin by proving:

Lemma 6.1. *Let $\boldsymbol{S}$ be a finite set of primes, and let $\boldsymbol{G}(t) \in \mathbb{Q}(t)$ be a nonzero rational function. Let $\boldsymbol{Z}$ be the finite set of rational zeros and poles of $\boldsymbol{G}(t)$, that is, $\boldsymbol{Z}$ consists of the rational numbers $z \in \mathbb{Q}$ for which $\boldsymbol{G}(z)$ is either zero or infinity. For any $z \in \boldsymbol{Z}$, let a_z, b_z be integers such that $b_z \neq 0$, $\gcd(a_z, b_z) = 1$ and $z = a_z/b_z$. Assume that the following is true*:

(D) *Let z be any element in $\boldsymbol{Z}$ such that $a_z \neq 0$. Then $a_z \not\equiv 0 \pmod l$ for each prime $l \in \boldsymbol{S}$.*

Then there is a rational function $\boldsymbol{F}(t) \in \mathbb{Q}(t)$ such that the following are true:

(1) *$\boldsymbol{F}(t_*) \in \mathbb{Z}_l^\times$ for each prime $l \in \boldsymbol{S}$ and each $t_* \in \mathbb{Q}$; and*

(2) *$\boldsymbol{G}(\boldsymbol{F}(t_*))$ is defined (that is, not infinity) and nonzero for each $t_* \in \mathbb{Q}$.*

Proof. We consider two cases:

Case 1. $\boldsymbol{Z}$ is nonempty.

By the Chinese remainder theorem, there exists an integer ϵ such that $\epsilon \equiv 2 \pmod 4$ and ϵ is a quadratic nonresidue in $\mathbb{F}_l^\times$ for each odd prime $l \in \boldsymbol{S}$ with $l \neq 2$. Let p_0 be an odd prime such that:

(i) $p_0 \notin \boldsymbol{S}$;

(ii) $b_z \not\equiv 0 \pmod{p_0}$ for every $z \in \boldsymbol{Z}$; and

(iii) for any element z in $\boldsymbol{Z}$ such that $a_z \neq 0$, we have $a_z \not\equiv 0 \pmod{p_0}$.

For each $z \in \boldsymbol{Z}$, we define

$$D_z := p_0 b_z \operatorname{sign}(a_z) \prod_{w \in \boldsymbol{Z} \setminus \{z\}} \max(1, |a_w|) \in \mathbb{Z}, \tag{26}$$

where $\operatorname{sign}(\cdot)$ denotes the usual sign function of $\mathbb{R}$, that is, $\operatorname{sign}(x) = 1$ if $x \geq 0$, and $\operatorname{sign}(x) = -1$ if $x < 0$. We see that $|D_z| \geq p_0 \geq 3$ for each $z \in \boldsymbol{Z}$. This implies that $|D_z - 1| \geq 1$ for every $z \in \boldsymbol{Z}$. We will prove that the rational function $\boldsymbol{F}(t) \in \mathbb{Q}(t)$, defined by

$$\boldsymbol{F}(t) := \left(p_0 \prod_{z \in \boldsymbol{Z}} \max(1, |a_z|)\right)\left(1 + \frac{4 \prod_{l \in \boldsymbol{S}, l \neq 2} l \prod_{z \in \boldsymbol{Z}} (D_z - 1)}{t^2 - p_0^2 \epsilon}\right), \tag{27}$$

satisfies (1) and (2) in Lemma 6.1. Indeed, take any rational number t_*, and write $t_* = t_1/t_2$, where $t_1, t_2 \in \mathbb{Z}$, $t_2 \neq 0$ and $\gcd(t_1, t_2) = 1$. For each prime l, denote by v_l the l-adic valuation of $\mathbb{Q}_l$. For each prime $l \in \boldsymbol{S}$ with $l \neq 2$, one knows that

$$v_l\left(\frac{1}{t_*^2 - p_0^2\epsilon}\right) = v_l(t_2^2) - v_l(t_1^2 - t_2^2 p_0^2 \epsilon).$$

Assume that $t_1^2 - t_2^2 p_0^2 \epsilon \equiv 0 \pmod{l}$. Since $p_0 \neq l$ and ϵ is a quadratic nonresidue in $\mathbb{F}_l^\times$, it follows that $t_1 \equiv t_2 \equiv 0 \pmod{l}$, which is a contradiction. Hence we deduce that $v_l(t_1^2 - t_2^2 p_0^2 \epsilon) = 0$. Thus we see that

$$v_l\left(\frac{1}{t_*^2 - p_0^2\epsilon}\right) = v_l(t_2^2) \geq 0.$$

Therefore $\dfrac{1}{t_*^2 - p_0^2\epsilon} \in \mathbb{Z}_l$, and hence we deduce that

$$1 + \frac{4\prod_{l \in \boldsymbol{S},\, l \neq 2} l \prod_{z \in \boldsymbol{Z}} (D_z - 1)}{t_*^2 - p_0^2 \epsilon} \in 1 + l\mathbb{Z}_l.$$

By assumption (D) and the choice of p_0, one knows that $p_0 \prod_{z \in \boldsymbol{Z}} \max(1, |a_z|) \in \mathbb{Z}_l^\times$. Hence it follows that for each prime $l \in \boldsymbol{S}$ with $l \neq 2$, $\boldsymbol{F}(t_*) \in \mathbb{Z}_l^\times$ for every $t_* \in \mathbb{Q}$. Thus we have shown that if $2 \notin \boldsymbol{S}$, then $\boldsymbol{F}(t)$ satisfies (1) in Lemma 6.1. Hence it remains to show that if $2 \in \boldsymbol{S}$, then $\boldsymbol{F}(t_*) \in \mathbb{Z}_2^\times$ for every $t_* \in \mathbb{Q}$.

Let us first assume that t_1 is even. Hence t_2 is odd, and then one sees that $t_1^2 - t_2^2 p_0^2 \epsilon \equiv 2 \pmod{4}$. Thus $v_2(t_1^2 - t_2^2 p_0^2 \epsilon) = 1$. Hence it follows that

$$v_2\left(\frac{2}{t_*^2 - p_0^2\epsilon}\right) = 1 + v_2(t_2^2) - v_2(t_1^2 - t_2^2 p_0^2 \epsilon) = 0,$$

which implies that $2/(t_*^2 - p_0^2 \epsilon) \in \mathbb{Z}_2$ for all $t_* \in \mathbb{Q}$.

Now assume that t_1 is odd. Since ϵ is even, one sees that $t_1^2 - t_2^2 p_0^2 \epsilon$ is odd. Hence it follows that

$$v_2\left(\frac{2}{t_*^2 - p_0^2\epsilon}\right) = 1 + v_2(t_2^2) - v_2(t_1^2 - t_2^2 p_0^2 \epsilon) = 1 + v_2(t_2^2) \geq 1.$$

Thus we have shown that $2/(t_*^2 - p_0^2 \epsilon) \in \mathbb{Z}_2$ for all $t_* \in \mathbb{Q}$. By the definition of $\boldsymbol{F}(t)$ and assumption (D), we deduce that $\boldsymbol{F}(t_*) \in \mathbb{Z}_2^\times$ for all $t_* \in \mathbb{Q}$. Hence the rational function $\boldsymbol{F}(t)$ satisfies Lemma 6.1(1).

Now we prove that $\boldsymbol{F}(t)$ satisfies Lemma 6.1(2). Since z is a rational zero or pole of $\boldsymbol{G}(t)$ for each $z \in \boldsymbol{Z}$, we see that if $\boldsymbol{F}(t_*) \neq z$ for every $z \in \boldsymbol{Z}$ and all $t_* \in \mathbb{Q}$, then $\boldsymbol{G}(\boldsymbol{F}(t_*))$ is defined, namely, not infinity, and nonzero for all $t_* \in \mathbb{Q}$.

Assume that there is a rational number $t_* \in \mathbb{Q}$ such that $\boldsymbol{F}(t_*) = z$ for some $z = a_z/b_z \in \boldsymbol{Z}$. We consider two subcases:

Subcase 1. $a_z \neq 0$.

We see that $\max(1, |a_z|) = |a_z|$. Hence it follows that

$$D_z\left(1 + \frac{4\prod_{l\in \boldsymbol{S}, l\neq 2} l \prod_{w\in \boldsymbol{Z}}(D_w - 1)}{t_*^2 - p_0^2\epsilon}\right) = 1.$$

Upon multiplying both sides by $t_*^2 - p_0^2\epsilon$ and simplifying, we deduce that

$$t_*^2 = p_0^2\epsilon - 4D_z \prod_{l\in \boldsymbol{S},\, l\neq 2} l \prod_{w\in \boldsymbol{Z}\setminus\{z\}} (D_w - 1).$$

Hence it follows from (26) that

$$t_*^2 = p_0\left(p_0\epsilon - 4b_z \operatorname{sign}(a_z) \prod_{w\in \boldsymbol{Z}\setminus\{z\}} \max(1, |a_w|) \prod_{l\in \boldsymbol{S},\, l\neq 2} l \prod_{w\in \boldsymbol{Z}\setminus\{z\}} (D_w - 1)\right).$$

This implies that $t_* \in \mathbb{Z}$ and $t_* \equiv 0 \pmod{p_0}$. Hence $v_{p_0}(t_*^2) = 2v_{p_0}(t_*) \geq 2$. Thus,

$$p_0\epsilon - 4b_z \operatorname{sign}(a_z) \prod_{w\in \boldsymbol{Z}\setminus\{z\}} \max(1, |a_w|) \prod_{l\in \boldsymbol{S},\, l\neq 2} l \prod_{w\in \boldsymbol{Z}\setminus\{z\}} (D_w - 1) \equiv 0 \pmod{p_0}.$$

Hence

$$4b_z \operatorname{sign}(a_z) \prod_{w\in \boldsymbol{Z}\setminus\{z\}} \max(1, |a_w|) \prod_{l\in \boldsymbol{S},\, l\neq 2} l \prod_{w\in \boldsymbol{Z}\setminus\{z\}} (D_w - 1) \equiv 0 \pmod{p_0}.$$

By (26), one knows that $D_w \equiv 0 \pmod{p_0}$ for every $w \in \boldsymbol{Z}$. Hence

$$\prod_{w\in \boldsymbol{Z}\setminus\{z\}} (D_w - 1) \equiv (-1)^{m-1} \pmod{p_0}.$$

Thus we deduce that

$$(-1)^{m-1} 4b_z \operatorname{sign}(a_z) \prod_{w\in \boldsymbol{Z}\setminus\{z\}} \max(1, |a_w|) \prod_{l\in \boldsymbol{S},\, l\neq 2} l \equiv 0 \pmod{p_0},$$

which is a contradiction to the choice of p_0. Therefore $\boldsymbol{F}(t_*) \neq z$ for all $t_* \in \mathbb{Q}$.

Subcase 2. $a_z = 0$.

We see that $\boldsymbol{F}(t_*) = a_z/b_z = 0$. Hence we deduce from the definition of $\boldsymbol{F}(t_*)$ that

$$t_*^2 = p_0^2\epsilon - 4 \prod_{l\in \boldsymbol{S},\, l\neq 2} l \prod_{w\in \boldsymbol{Z}} (D_w - 1).$$

This implies that $t_* \in \mathbb{Z}$. Hence we deduce that

$$t_*^2 = p_0^2\epsilon \pmod{l}$$

for each prime $l \in \boldsymbol{S}$ with $l \neq 2$. Since ϵ is a quadratic nonresidue in $\mathbb{F}_l^\times$, it follows that $t_* \equiv p_0 \equiv 0 \pmod{l}$, which is a contradiction to the choice of p_0. Thus, in any event, $\boldsymbol{F}(t_*) \neq z$ for all $t_* \in \mathbb{Q}$. Therefore $\boldsymbol{F}(t_*)$ satisfies Lemma 6.1(2).

Case 2. $\boldsymbol{Z} = \varnothing$.

In this case, let ϵ be the same as in Case 1, and let p_0 be an odd prime such that $p_0 \notin \boldsymbol{S}$. Let $\boldsymbol{F}(t) \in \mathbb{Q}(t)$ be the rational function defined by

$$\boldsymbol{F}(t) := 1 + \frac{4\prod_{l\in\boldsymbol{S},\, l\neq 2} l}{t^2 - p_0^2\epsilon}. \tag{28}$$

Using the same arguments as in Case 1, one can show that $\boldsymbol{F}(t)$ satisfies (1) and (2) in Lemma 6.1. □

Lemma 6.2. *Let $\boldsymbol{D}(t) \in \mathbb{Q}(t)$ be a nonzero rational function of the form*

$$\boldsymbol{D}(t) = \frac{at^4 + bt^2 + c}{dt^4 + et^2 + f},$$

where a, b, c, d, e, f are integers. Let q be an odd prime. Assume that there exists an integer t_0 such that

$$\begin{aligned} at_0^4 + bt_0^2 + c &\equiv 0 \pmod{q}, \\ at_0^4 + bt_0^2 + c &\not\equiv 0 \pmod{q^2}, \\ dt_0^4 + et_0^2 + f &\not\equiv 0 \pmod{q}. \end{aligned}$$

Then there exists a rational function $\Gamma(t) \in \mathbb{Q}(t)$ such that for all $t_* \in \mathbb{Q}$, $\boldsymbol{D}(\Gamma(t_*))$ belongs to $q\mathbb{Z}_q$, but does not belong to $q^2\mathbb{Z}_q$.

Proof. Let ϵ be an integer such that ϵ is a quadratic nonresidue in $\mathbb{F}_q^\times$. Let q_0 be an odd prime such that $q_0 \neq q$. We will show that the rational function $\Gamma(t) \in \mathbb{Q}(t)$, defined by

$$\Gamma(t) = t_0 + \frac{q^2}{t^2 - q_0^2\epsilon}, \tag{29}$$

satisfies the assertions in Lemma 6.2.

Since ϵ is not a square in $\mathbb{F}_q^\times$, it follows that $t_*^2 - q_0^2\epsilon$ is nonzero for each $t_* \in \mathbb{Q}$, and hence $\Gamma(t_*)$ is well defined, namely, not infinity for all $t_* \in \mathbb{Q}$.

We now prove that $\Gamma(t_*)$ belongs to $t_0 + q^2\mathbb{Z}_q$ for all $t_* \in \mathbb{Q}$. Indeed, take any rational number t_*, and write $t_* = t_1/t_2$, where t_1, t_2 are integers such that $t_2 \neq 0$ and $\gcd(t_1, t_2) = 1$. We see that

$$v_q\left(\frac{1}{t_*^2 - q_0^2\epsilon}\right) = v_q\left(\frac{t_2^2}{t_1^2 - q_0^2\epsilon t_2^2}\right) = v_q(t_2^2) - v_q(t_1^2 - q_0^2\epsilon t_2^2).$$

If $t_2 \equiv 0 \pmod{q}$, then it follows that $t_1 \not\equiv 0 \pmod{q}$. Hence we deduce that

$$v_q(t_1^2 - q_0^2\epsilon t_2^2) = \min(v_q(t_1^2), v_q(q_0^2\epsilon t_2^2)) = \min(0, v_q(t_2^2)) = 0,$$

and thus

$$v_q\left(\frac{1}{t_*^2 - q_0^2\epsilon}\right) = v_q(t_2^2) - v_q(t_1^2 - q_0^2\epsilon t_2^2) = 2v_q(t_2) \geq 2.$$

Therefore $1/(t_*^2 - q_0^2\epsilon)$ belongs to $\mathbb{Z}_q$, and hence it follows from (29) that $\Gamma(t_*)$ belongs to $t_0 + q^2\mathbb{Z}_q$.

If $t_2 \not\equiv 0 \pmod{q}$, then $v_q(t_2^2) = 0$. We contend that $t_1^2 - q_0^2\epsilon t_2^2 \not\equiv 0 \pmod{q}$. Assume the contrary, that is, $t_1^2 - q_0^2\epsilon t_2^2 \equiv 0 \pmod{q}$. Since $t_2 \not\equiv 0 \pmod{q}$ and $q_0 \neq q$, we deduce that

$$\epsilon \equiv \left(\frac{t_1}{q_0 t_2}\right)^2 \pmod{q},$$

which contradicts the choice of ϵ. This establishes that $t_1^2 - q_0^2\epsilon t_2^2 \not\equiv 0 \pmod{q}$, and thus

$$v_q\left(\frac{1}{t_*^2 - q_0^2\epsilon}\right) = v_q(t_2^2) - v_q(t_1^2 - q_0^2\epsilon t_2^2) = 0.$$

Therefore $1/(t_*^2 - q_0^2\epsilon)$ belongs to $\mathbb{Z}_q^\times$, and hence it follows from (29) that $\Gamma(t_*)$ belongs to $t_0 + q^2\mathbb{Z}_q$.

Since $\Gamma(t_*)$ belongs to $t_0 + q^2\mathbb{Z}_q$, we see that

$$\begin{aligned}
a(\Gamma(t_*))^4 + b(\Gamma(t_*))^2 + c &\equiv at_0^4 + bt_0^2 + c \equiv 0 \pmod{q},\\
a(\Gamma(t_*))^4 + b(\Gamma(t_*))^2 + c &\equiv at_0^4 + bt_0^2 + c \not\equiv 0 \pmod{q^2},\\
d(\Gamma(t_*))^4 + e(\Gamma(t_*))^2 + f &\equiv dt_0^4 + et_0^2 + f \not\equiv 0 \pmod{q}.
\end{aligned}$$

The last congruence shows that

$$\frac{1}{c(\Gamma(t_*))^4 + d(\Gamma(t_*))^2 + e}$$

belongs to $\mathbb{Z}_q^\times$, and hence we deduce that for every $t_* \in \mathbb{Q}$,

$$\boldsymbol{D}(\Gamma(t_*)) = \frac{a(\Gamma(t_*))^4 + b(\Gamma(t_*))^2 + c}{d(\Gamma(t_*))^4 + e(\Gamma(t_*))^2 + f}$$

belongs to $q\mathbb{Z}_q$, but does not belong to $q^2\mathbb{Z}_q$. Thus our contention follows. $\square$

The next two examples will be used in proving the main theorem in this section.

Example 6.3. Let $\boldsymbol{D}_1(T) \in \mathbb{Q}(T)$ be the rational function defined by

$$\boldsymbol{D}_1(T) := \frac{455888941732987T^4 - 1641200890885920T^2 + 14770814323798008}{-5477180725633679T^4 + 19717850612281267 6T^2 - 177460655510530271 6}, \tag{30}$$

and define

$$\begin{aligned}(31)\quad \boldsymbol{D}_1^*(T) &:= 7+1682\boldsymbol{D}_1(T)\\ &= \frac{-38340254920051483T^4+1380250355610428708T^2-12422263806891130444}{5477180725633679T^4-197178506122812676T^2+1774606555105302716}.\end{aligned}$$

Let $T_0 = 0$, and let $q = 31$. Since

$$12422263806891130444 = 2^2 \cdot 7^3 \cdot 31 \cdot 433 \cdot 3299 \cdot 10589 \cdot 19309,$$

it follows that

$$v_q(-12422263806891130444) = v_{31}(-12422263806891130444) = 1,$$

and we deduce that for $T = T_0 = 0$ the numerator of the fraction in (31) is divisible by q, but not by q^2. Since

$$1774606555105302716 = 2^2 \cdot 7^2 \cdot 47 \cdot 192640746320593,$$

we see that for $T = T_0 = 0$ the denominator in (31) is not divisible by q.

Let $\epsilon = 3$, and let $q_0 = 5$. Following the proof of Lemma 6.2, we define the rational function $\Gamma_1(T) \in \mathbb{Q}(T)$ by (29), that is,

$$\Gamma_1(T) := T_0 + \frac{q^2}{T^2 - q_0^2\epsilon} = \frac{961}{T^2 - 75}. \tag{32}$$

Applying Lemma 6.2 with $\boldsymbol{D}_1^*(T)$ in the role of $\boldsymbol{D}(t)$, we deduce that for all $T_* \in \mathbb{Q}$, $\boldsymbol{D}_1^*(\Gamma_1(T_*))$ belongs to $31\mathbb{Z}_{31}$, but does not belong to $31^2\mathbb{Z}_{31}$, where

$$\boldsymbol{D}_1^*(\Gamma_1(T)) = \frac{\Sigma_{1,1}(T)}{\Sigma_{1,2}(T)} \in \mathbb{Q}(T) \tag{33}$$

with

$$\begin{aligned}(34)\quad \Sigma_{1,1}(T) = &-12422263806891130444T^8 + 3726679142067339133200T^6\\ &+ 855438785181123078355868T^4 - 170240958125426027001880200T^2\\ &- 2592297567404672316225380003\end{aligned}$$

and

$$\begin{aligned}(35)\quad \Sigma_{1,2}(T) = &\ 1774606555105302716T^8 - 532381966531590814800T^6\\ &- 122205519918242118687196T^4 + 24320125111216714469579400T^2\\ &+ 370328399913430215308191 0439.\end{aligned}$$

Example 6.4. Let $\boldsymbol{D}_2(T) \in \mathbb{Q}(T)$ be the rational function defined by

$$\boldsymbol{D}_2(T) := \frac{-643804017087754T^4+23176936231188880T^2-208592350625 03544}{407097080892401T^4-14655494912126204T^2+131899454209147204}, \tag{36}$$

and define

$$\boldsymbol{D}_2^*(T) := 133 + 1682\boldsymbol{D}_2(T) = \frac{-54143923915434895T^4+1949179850773171028T^2-17542605965314382876}{407097080892401T^4-14655494912126204T^2+131899454209147204}. \tag{37}$$

Let $T_0 = 0$, and let $q = 11$. Since

$$17542605965314382876 = 2^2 \cdot 7 \cdot 11 \cdot 56956512874397347,$$

it follows that for $T = T_0$ the numerator in (37) is divisible by 11, but not by 11^2. Since

$$131899454209147204 \equiv 8 \not\equiv 0 \pmod{11},$$

for $T = T_0$ the denominator in (37) is not divisible by 11.

Let $\epsilon = 7$, and let $q_0 = 3$. Following the proof of Lemma 6.2, we define the rational function $\Gamma_2(T) \in \mathbb{Q}(T)$ by (29), that is,

$$\Gamma_2(T) := T_0 + \frac{q^2}{T^2 - q_0^2\epsilon} = \frac{121}{T^2 - 63}. \tag{38}$$

Applying Lemma 6.2 with $\boldsymbol{D}_2^*(T)$ in the role of $\boldsymbol{D}(t)$, we deduce that for all $T_* \in \mathbb{Q}$, $\boldsymbol{D}_2^*(\Gamma_2(T_*))$ belongs to $11\mathbb{Z}_{11}$, but does not belong to $11^2\mathbb{Z}_{11}$, where

$$\boldsymbol{D}_2^*(\Gamma_2(T)) = \frac{\Sigma_{2,1}(T)}{\Sigma_{2,2}(T)} \in \mathbb{Q}(T) \tag{39}$$

with

$$\begin{aligned}\Sigma_{2,1}(T) = &-17542605965314382876T^8 + 4420736703259224484752T^6 \\ &- 389221676262826716788116T^4 + 13950123258644442355341240T^2 \\ &- 174687125980796870729105719\end{aligned} \tag{40}$$

and

$$\begin{aligned}\Sigma_{2,2}(T) = &\ 131899454209147204T^8 - 33238662460705095408T^6 \\ &+ 2926482501528191763292T^4 - 104888292579475114826088T^2 \\ &+ 1313439132893945928914009.\end{aligned} \tag{41}$$

The next result is a mild generalization of Theorem 2.1 of [Dong Quan 2014]. The only difference between these two theorems is that in the latter, a, b, c, d, e are assumed to be integers, whereas here we only assume that a, b, c, d, e belong to $\mathbb{Z}_p$. Upon examining closely the proof of [loc. cit.], we see that it is sufficient to assume that a, b, c, d, e are in $\mathbb{Z}_p$, and hence Theorem 6.5 follows immediately from the proof of that result.

Theorem 6.5 (separability criterion [Dong Quan 2014]). *Let n, m, k be positive integers, and let a, b, c, d, e be rational numbers such that $a \neq 0$. Let p be an odd prime such that a, b, c, d, e belong to $\mathbb{Z}_p$ and $a \equiv 0 \pmod p$. Let $F(x) \in \mathbb{Q}[x]$ be the polynomial defined by*

$$F(x) := ax^{2n+2} + (bx^{2m} + c)(dx^{2k} + e). \tag{42}$$

Define

$$\begin{aligned}
n_1 &:= (m+k)(v_p(a) - v_p(bd)) + m + k - 1,\\
n_2 &:= (m+k)(v_p(a) - v_p(b)) + m - 1,\\
n_3 &:= (m+k)(v_p(a) - v_p(d)) + k - 1,\\
n_4 &:= (m+k)v_p(a) - 1,\\
n_5 &:= v_p(a) - v_p(bd) + m + k - 1.
\end{aligned}$$

Suppose that the following are true:

(S1) $n > m + k - 1$ *and* $n > \max(n_1, n_2, n_3, n_4, n_5)$.

(S2) $ce \not\equiv 0 \pmod p$, $km \not\equiv 0 \pmod p$, *and* $b^k e^m + (-1)^{m+k+1} c^k d^m \not\equiv 0 \pmod p$.

Then F is separable, that is, it has exactly $2n+2$ distinct roots in $\mathbb{C}$. □

Using Theorem 6.5, we prove the following corollaries that are crucial in constructing algebraic families of curves violating the Hasse principle.

Corollary 6.6. *We maintain the same notation as in Example 6.3. Let $\boldsymbol{D}_1(T)$, $\boldsymbol{D}_1^*(T)$, $\Gamma_1(T) \in \mathbb{Q}(T)$ be the rational functions defined by (30), (31), (32), respectively. Let n be a positive integer such that $n > 5$. For each rational number $T_* \in \mathbb{Q}$, let $\mathcal{P}_{1,T_*}(x) \in \mathbb{Q}[x]$ be the polynomial of degree $2n+2$ given by*

$$\begin{aligned}
\mathcal{P}_{1,T_*}(x) := {} & 118579927725(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 x^{2n+2}\\
& + (2(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192)x^2 - 16689645)\\
& \times (261(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192)x^2 - 33379290),
\end{aligned} \tag{43}$$

where the composition rational function $\boldsymbol{D}_1^*(\Gamma_1(T))$ of $\boldsymbol{D}_1^*(T)$ and $\Gamma_1(T)$ is given by (33). Then for all $T_* \in \mathbb{Q}$, the polynomial $\mathcal{P}_{1,T_*}(x)$ is separable, that is, it has exactly $2n+2$ distinct roots in $\mathbb{C}$.

Proof. Throughout the proof, we maintain the same notation as in Theorem 6.5. Take any rational number $T_* \in \mathbb{Q}$, and define

$$\begin{aligned}
a &:= 118579927725(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2,\\
b &:= 2(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192),\\
c &:= -16689645,\\
d &:= 261(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192),
\end{aligned}$$

$$e := -33379290.$$

Let $p = 31$, and let $m = k = 1$. Since $118579927725 \equiv 27 \not\equiv 0 \pmod{31}$, $123192 \equiv 29 \not\equiv 0 \pmod{31}$, it follows from Example 6.3 that

$$\begin{aligned}
v_p(a) &= v_{31}(118579927725(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2) = 2v_{31}(\boldsymbol{D}_1^*(\Gamma_1(T_*))) = 2,\\
v_p(b) &= v_{31}(2(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192)) = v_{31}(123192) = 0,\\
v_p(d) &= v_{31}(261(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192)) = v_{31}(123192) = 0.
\end{aligned}$$

We see that

$$\begin{aligned}
n_1 &:= (m+k)(v_p(a) - v_p(bd)) + m + k - 1 = 2v_p(a) + 1 = 5,\\
n_2 &:= (m+k)(v_p(a) - v_p(b)) + m - 1 = 2v_p(a) = 4,\\
n_3 &:= (m+k)(v_p(a) - v_p(d)) + k - 1 = 2v_p(a) = 4,\\
n_4 &:= (m+k)v_p(a) - 1 = 2v_p(a) - 1 = 3,\\
n_5 &:= v_p(a) - v_p(bd) + m + k - 1 = 2 + 1 = 3,
\end{aligned}$$

and hence

$$\max(n_1, n_2, n_3, n_4, n_5) = 5.$$

By assumption, we know that

$$n > 5 = \max(n_1, n_2, n_3, n_4, n_5),$$

and hence condition (S1) is satisfied.

It is obvious that $km = 1 \not\equiv 0 \pmod{31}$ and

$$ce = (-16689645) \cdot (-33379290) \equiv 25 \not\equiv 0 \pmod{31}.$$

Furthermore, since $\boldsymbol{D}_1^*(\Gamma_1(T_*))$ belongs to $31\mathbb{Z}_{31}$, we deduce that

$$\begin{aligned}
b^k e^m + (-1)^{m+k+1} c^k d^m &= be - cd\\
&= (2(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192))(-33379290)\\
&\qquad - (-16689645)(261(29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192))\\
&\equiv 2 \cdot 123192 \cdot (-33379290) + 16689645 \cdot 261 \cdot 123192\\
&\equiv 12 \not\equiv 0 \pmod{31}.
\end{aligned}$$

Therefore condition (S2) is satisfied, and hence the polynomial $\mathcal{P}_{1,T_*}(x)$ is separable. Since T_* is an arbitrary rational number, our contention follows. □

Corollary 6.7. *We maintain the same notation as in Example 6.4. Let $\boldsymbol{D}_2(T)$, $\boldsymbol{D}_2^*(T)$, $\Gamma_2(T) \in \mathbb{Q}(T)$ be the rational functions defined by (36), (37), (38), respectively. Let n be a positive integer such that $n > 5$. For each rational number $T_* \in \mathbb{Q}$,* *let* $\mathcal{P}_{2,T_*}(x) \in \mathbb{Q}[x]$ *be the polynomial of degree* $2n+2$ *given by*

$$\begin{aligned}(44)\qquad \mathcal{P}_{2,T_*}(x) := 84898109(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 x^{2n+2} \\ + (2(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600)x^2 + 49619) \\ \times (261(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600)x^2 + 99238),\end{aligned}$$

where the composition rational function $\boldsymbol{D}_2^*(\Gamma_2(T))$ *of* $\boldsymbol{D}_2^*(T)$ *and* $\Gamma_2(T)$ *is given by* (39)*. Then for all* $T_* \in \mathbb{Q}$, *the polynomial* $\mathcal{P}_{2,T_*}(x)$ *is separable*, *that is*, *it has exactly* $2n+2$ *distinct roots in* $\mathbb{C}$.

Proof. Throughout the proof, we maintain the same notation as in Theorem 6.5. Take any rational number $T_* \in \mathbb{Q}$, and define

$$\begin{aligned} a &:= 84898109(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2, \\ b &:= 2(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600), \\ c &:= 49619, \\ d &:= 261(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600), \\ e &:= 99238. \end{aligned}$$

Let $p = 11$, and let $m = k = 1$. Since $84898109 \equiv 10 \not\equiv 0 \pmod{11}$, $40600 \equiv 10 \not\equiv 0 \pmod{11}$, it follows from Example 6.4 that

$$\begin{aligned} v_p(a) &= v_{11}(84898109(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2) = 2v_{11}(\boldsymbol{D}_2^*(\Gamma_2(T_*))) = 2, \\ v_p(b) &= v_{11}(2(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600)) = v_{11}(40600) = 0, \\ v_p(d) &= v_{31}(261(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600)) = v_{11}(40600) = 0. \end{aligned}$$

We see that

$$\begin{aligned} n_1 &:= (m+k)(v_p(a) - v_p(bd)) + m + k - 1 = 2v_p(a) + 1 = 5, \\ n_2 &:= (m+k)(v_p(a) - v_p(b)) + m - 1 = 2v_p(a) = 4, \\ n_3 &:= (m+k)(v_p(a) - v_p(d)) + k - 1 = 2v_p(a) = 4, \\ n_4 &:= (m+k)v_p(a) - 1 = 2v_p(a) - 1 = 3, \\ n_5 &:= v_p(a) - v_p(bd) + m + k - 1 = 2 + 1 = 3, \end{aligned}$$

and hence

$$\max(n_1, n_2, n_3, n_4, n_5) = 5.$$

By assumption, we know that

$$n > 5 = \max(n_1, n_2, n_3, n_4, n_5),$$

and hence condition (S1) is satisfied.

It is obvious that $km = 1 \not\equiv 0 \pmod{11}$ and

$$ce = 49619 \cdot 99238 \equiv 8 \not\equiv 0 \pmod{11}.$$

Since $\boldsymbol{D}_2^*(\Gamma_2(T_*))$ belongs to $11\mathbb{Z}_{11}$, we deduce that

$$\begin{aligned} b^k e^m + (-1)^{m+k+1} c^k d^m &= be - cd \\ &= (2(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600))(99238) \\ &\qquad - (49619)(261(29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600)) \\ &\equiv 2 \cdot (-40600) \cdot 99238 - (49619) \cdot 261 \cdot (-40600) \\ &\equiv 8 \not\equiv 0 \pmod{11}. \end{aligned}$$

Therefore condition (S2) is satisfied, and hence the polynomial $\mathcal{P}_{2,T_*}(x)$ is separable. Since T_* is an arbitrary rational number, our contention follows. □

For the rest of this section, let

$$\begin{aligned} \boldsymbol{A}_1 &:= \{n \in \mathbb{Z} : n > 5, n \not\equiv 0 \pmod 4 \text{ and } n \not\equiv 21 \pmod{29}\}, \\ \boldsymbol{A}_2 &:= \{n \in \mathbb{Z} : n > 5, n \not\equiv 0 \pmod 4 \text{ and } n \not\equiv 8 \pmod{29}\}. \end{aligned}$$

We see that

$$\boldsymbol{A}_1 \cup \boldsymbol{A}_2 = \{n \in \mathbb{Z} : n > 5 \text{ and } n \not\equiv 0 \pmod 4\}. \tag{45}$$

We now prove the main theorem in this section.

Theorem 6.8. *For each $n \in \boldsymbol{A}_1$ and each rational number $T_* \in \mathbb{Q}$, let $\mathcal{P}_{1,T_*}(x) \in \mathbb{Q}[x]$ be the polynomial of degree $2n+2$ defined by (43). For each $n \in \boldsymbol{A}_2$ and each rational number $T_* \in \mathbb{Q}$, let $\mathcal{P}_{2,T_*}(x) \in \mathbb{Q}[x]$ be the polynomial of degree $2n+2$ defined by (44). Then*:

(i) *For each $n \in \boldsymbol{A}_1$ and each rational number $T_* \in \mathbb{Q}$, the hyperelliptic curve $\mathcal{C}^{(7,261,15)}_{n,T_*,(29,1,3)}$ of genus n satisfies* CHP *and* NZC, *where $\mathcal{C}^{(7,261,15)}_{n,T_*,(29,1,3)}$ is the smooth projective model of the affine curve defined by*

$$\mathcal{C}^{(7,261,15)}_{n,T_*,(29,1,3)} : z^2 = \mathcal{P}_{1,T_*}(x).$$

(ii) *For each $n \in \boldsymbol{A}_2$ and each rational number $T_* \in \mathbb{Q}$, the hyperelliptic curve $\mathcal{C}^{(133,29,27)}_{n,T_*,(29,1,3)}$ of genus n satisfies* CHP *and* NZC, *where $\mathcal{C}^{(133,29,27)}_{n,T_*,(29,1,3)}$ is the smooth projective model of the affine curve defined by*

$$\mathcal{C}^{(133,29,27)}_{n,T_*,(29,1,3)} : z^2 = \mathcal{P}_{2,T_*}(x).$$

Remark 6.9. By (45) and Theorem 6.8, Theorem 1.1 follows immediately.

Proof. Throughout the proof of Theorem 6.8, we will use the same notation as in Theorem 4.1 and Lemma 5.1. We first prove that Theorem 6.8(i) holds.

Let $(p, b, d, \alpha_0, \beta_0, \gamma_0) = (29, 1, 3, 7, 261, 15)$. Let n be any integer such that $n \in \boldsymbol{A}_1$. We see that $\overline{\beta_0} = 9$. One can check that the sextuple $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfies (A1)–(A5) and (B1). Indeed, (A1), (A2), (A4), (A5) and (B1) are obvious. It remains to prove that $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfies (A3). By (11) and (12), we know that

$$P_0 = 124613, \quad Q_0 = -63945.$$

The conic $\mathcal{Q}_1^{(7,261,15)}$ in (A3) of Theorem 4.1 defined by

$$\mathcal{Q}_1^{(7,261,15)} : 29U^2 - V^2 + 2079746732385T^2 = 0$$

has a point $(u_0, v_0, t_0) = (166257, 3020031, 2)$, and hence $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfies (A3).

Let $\boldsymbol{S} := \{2, 29\}$, and let $\boldsymbol{C}_1(T)$ be the rational function in $\mathbb{Q}(T)$ defined by the same equation (23) of C with $(0, T)$ in the role of (A, B), that is,

$$\boldsymbol{C}_1(T) := \frac{-2v_0T - 4p^3\alpha_0\beta_0t_0^2Q_0}{T^2 + 4p^5\beta_0t_0^2Q_0} = \frac{-6040062T + 45588900213360}{T^2 - 5477180725633680}.$$

Let $\boldsymbol{G}_1(T) \in \mathbb{Q}(T)$ be the rational function defined by

$$\begin{aligned} (46) \qquad \boldsymbol{G}_1(T) &= v_0 + T\boldsymbol{C}_1(T) \\ &= \frac{-3020031T^2 + 45588900213360T - 16541255584016208244080}{T^2 - 5477180725633680}. \end{aligned}$$

Since the numerator and denominator of $\boldsymbol{G}_1(T)$ are irreducible polynomials over $\mathbb{Q}$, the set $\boldsymbol{Z}_1$ of rational zeros and poles of $\boldsymbol{G}_1(T)$ is empty. Hence, applying Lemma 6.1 for the triple $(\boldsymbol{S}, \boldsymbol{G}_1(T), \boldsymbol{Z}_1)$, we know that $\boldsymbol{F}_1(T)$ satisfies (1) and (2) in Lemma 6.1, where $\boldsymbol{F}_1(T)$ is the rational function defined by (28) with $(p_0, \epsilon) = (3, 2)$ and $(\boldsymbol{S}, \boldsymbol{G}_1(T), \boldsymbol{Z}_1)$ in the role of $(\boldsymbol{S}, \boldsymbol{G}(T), \boldsymbol{Z})$, that is,

$$\boldsymbol{F}_1(T) := 1 + \frac{4\prod_{l \in \boldsymbol{S},\, l \neq 2} l}{T^2 - p_0^2\epsilon} = 1 + \frac{116}{T^2 - 18} = \frac{T^2 + 98}{T^2 - 18}.$$

Let $\Gamma_1(T) \in \mathbb{Q}(T)$ be the rational function defined by (32). Recall that

$$\Gamma_1(T) := \frac{961}{T^2 - 75}.$$

It is known that $\Gamma_1(T_*)$ is well-defined, namely, not infinity for all $T_* \in \mathbb{Q}$.

Take an arbitrary rational number $T_* \in \mathbb{Q}$, and let $(A, B) = (0, \boldsymbol{F}_1(\Gamma_1(T_*)))$. By Lemma 6.1, we know that $(0, \boldsymbol{F}_1(T_*))$ satisfies (C1) and (C2) in Lemma 5.1, and it thus follows that $(A, B) = (0, \boldsymbol{F}_1(\Gamma_1(T_*)))$ also satisfies (C1) and (C2).

Let $\boldsymbol{D}_1(\Gamma_1(T))$ be the rational function in $\mathbb{Q}(T)$ defined by the same equation (23) of C with $(0, \boldsymbol{F}_1(\Gamma_1(T)))$ in the role of (A, B), that is,

$$\begin{aligned}(47)\quad \boldsymbol{D}_1(\Gamma_1(T)) &:= \boldsymbol{C}_1(\boldsymbol{F}_1(\Gamma_1(T)))\\ &= \frac{45588894173298(\Gamma_1(T))^4 - 1641200890885920(\Gamma_1(T))^2 + 14770814323798008}{-5477180725633679(\Gamma_1(T))^4 + 197178506122812676(\Gamma_1(T))^2 - 1774606555105302716}.\end{aligned}$$

Note that $(A, B) = (0, \boldsymbol{F}_1(\Gamma_1(T_*)))$ satisfies (C1) and (C2). Hence, using the same arguments as in the proof of Lemma 5.1, one knows that $\boldsymbol{D}_1(\Gamma_1(T_*)) \in \mathbb{Z}_{29}$.

We see that

$$u := u_0 + A\boldsymbol{D}_1(\Gamma_1(T_*)) = u_0 = 166257 \neq 0.$$

Furthermore, it follows from Lemma 6.1(2), (46) and (47) that

$$v := v_0 + B\boldsymbol{D}_1(\Gamma_1(T_*)) = \boldsymbol{G}_1(\boldsymbol{F}_1(\Gamma_1(T_*)))$$

is well-defined, namely, not infinity and nonzero. Hence $(A, B) = (0, \boldsymbol{F}_1(\Gamma_1(T_*)))$ satisfies (C3) in Lemma 5.1.

Set

$$\begin{aligned}\alpha &:= \alpha_0 + 2p^2\boldsymbol{D}_1(\Gamma_1(T_*)) = 7 + 1682\boldsymbol{D}_1(\Gamma_1(T_*)) = \boldsymbol{D}_1^*(\Gamma_1(T_*)),\\ \beta &:= \beta_0 = 261,\\ \gamma &:= \gamma_0 = 15,\end{aligned}$$

where $\boldsymbol{D}_1^*(\Gamma_1(T)) \in \mathbb{Q}(T)$ is the rational function defined by (33). Recall from there that

$$\boldsymbol{D}_1^*(\Gamma_1(T)) = \frac{\Sigma_{1,1}(T)}{\Sigma_{1,2}(T)},$$

where $\Sigma_{1,1}(T)$, $\Sigma_{1,2}(T)$ are defined by (34), (35), respectively.

By Lemma 5.1, we know that (α, β, γ) satisfies (A1)–(A5) and (B1). By (11) and (12), we know that

$$\begin{aligned}P &= 29(7 + 1682\boldsymbol{D}_1(\Gamma_1(T_*)))^2 + 123192 = 29(\boldsymbol{D}_1^*(\Gamma_1(T_*)))^2 + 123192,\\ Q &= Q_0 = -63945.\end{aligned}$$

It is not difficult to see that the curve $\mathcal{C}^{(7,261,15)}_{n,T_*,(29,1,3)}$ defined in Theorem 6.8(i) is the smooth projective model of the affine curve defined by (13).

By Corollary 6.6, we know that $\mathcal{P}_{1,T_*}(x)$ is separable, and hence we deduce that condition (S) in Theorem 4.1 is true. Since $\boldsymbol{D}_1(\Gamma_1(T_*)) \in \mathbb{Z}_{29}$, we see that

$$-2\left(\frac{\gamma}{\alpha}\right)^2 \equiv 21 \pmod{29}.$$

Since $n \in \boldsymbol{A}_1$, we deduce that

$$n \not\equiv -2\left(\frac{\gamma}{\alpha}\right)^2 \pmod{29}.$$

Thus (A6) holds if n is odd, and (B2) holds if $n \equiv 2 \pmod 4$. By Corollaries 4.6 and 4.8, we deduce that for each $n \in \boldsymbol{A}_1$, the curve $\mathcal{C}_{n,T_*,(29,1,3)}^{(7,261,15)}$ satisfies CHP and NZC. Since T_* is an arbitrary rational number, Theorem 6.8(i) follows.

We now prove Theorem 6.8(ii). We will use the same notation as in the proof of part (i) as long as it does not cause any confusion. We will use the same arguments as in the proof of part (i) to construct an algebraic family of hyperelliptic curves of genus n satisfying CHP and NZC for each $n \in \boldsymbol{A}_2$.

Let $(p, b, d, \alpha_0, \beta_0, \gamma_0) = (29, 1, 3, 133, 29, 27)$. Let n be any integer such that $n \in \boldsymbol{A}_2$. We see that $\bar{\beta}_0 = 1$. One can check that the sextuple $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfies (A1)–(A5) and (B1). Indeed (A1), (A2), (A4), (A5) and (B1) are obvious. It remains to prove that the sextuple satisfies (A3). By (11) and (12), we know that

$$P_0 = 472381, \quad Q_0 = 1711.$$

The conic $\mathcal{Q}_1^{(133,29,27)}$ in (A3) of Theorem 4.1 defined by

$$\mathcal{Q}_1^{(133,29,27)} : 29U^2 - V^2 - 23439072839T^2 = 0$$

has a point $(u_0, v_0, t_0) = (728799, 3613777, 10)$, and thus $(p, b, d, \alpha_0, \beta_0, \gamma_0)$ satisfies (A3).

Let $\boldsymbol{S} := \{2, 29\}$, and let $\boldsymbol{C}_2(T) \in \mathbb{Q}(T)$ be the rational function defined by the same equation (23) of C with $(0, T)$ in the role of (A, B), that is,

$$\boldsymbol{C}_2(T) = \frac{-2v_0T - 4p^3\alpha_0\beta_0t_0^2Q_0}{T^2 + 4p^5\beta_0t_0^2Q_0} = \frac{-7227554T - 64380394481200}{T^2 + 407097080892400}.$$

Let $\boldsymbol{G}_2(T) \in \mathbb{Q}(T)$ be the rational function defined by

$$\begin{aligned} \boldsymbol{G}_2(T) &= v_0 + T\boldsymbol{C}_2(T) \\ &= \frac{-3613777T^2 - 64380394481200T + 1471158067669609459480}{T^2 + 407097080892400}. \end{aligned} \tag{48}$$

Since the numerator and denominator of $\boldsymbol{G}_2(T)$ are irreducible polynomials over $\mathbb{Q}$, the set $\boldsymbol{Z}_2$ of rational zeros and poles of $\boldsymbol{G}_2(T)$ is empty. Hence, applying Lemma 6.1 for the triple $(\boldsymbol{S}, \boldsymbol{G}_2(T), \boldsymbol{Z}_2)$, we know that $\boldsymbol{F}_2(T)$ satisfies (1) and (2) in Lemma 6.1, where $\boldsymbol{F}_2(T)$ is the rational function defined by (28) with $(p_0, \epsilon) = (3, 2)$ and $(\boldsymbol{S}, \boldsymbol{G}_2(T), \boldsymbol{Z}_2)$ in the role of $(\boldsymbol{S}, \boldsymbol{G}(T), \boldsymbol{Z})$, that is,

$$\boldsymbol{F}_2(T) := 1 + \frac{4\prod_{l \in \boldsymbol{S},\, l \neq 2} l}{T^2 - p_0^2\epsilon} = 1 + \frac{116}{T^2 - 18} = \frac{T^2 + 98}{T^2 - 18}.$$

Let $\Gamma_2(T) \in \mathbb{Q}(T)$ be the rational function defined by (38). Recall that

$$\Gamma_2(T) := \frac{121}{T^2 - 63}.$$

It is known that $\Gamma_2(T_*)$ is well defined, namely, not infinity for each rational number $T_* \in \mathbb{Q}$.

Now take an arbitrary rational number $T_* \in \mathbb{Q}$, and let $(A, B) = (0, \boldsymbol{F}_2(\Gamma_2(T_*)))$. By Lemma 6.1, we know that $(0, \boldsymbol{F}_2(T_*))$ satisfies (C1) and (C2) in Lemma 5.1, and it thus follows that $(A, B) = (0, \boldsymbol{F}_2(\Gamma_2(T_*)))$ also satisfies (C1) and (C2).

Let $\boldsymbol{D}_2(\Gamma_2(T))$ be the rational function in $\mathbb{Q}(T)$ defined by the same equation (23) of C with $(0, \boldsymbol{F}_2(\Gamma_2(T)))$ in the role of (A, B), that is,

$$\begin{aligned} \boldsymbol{D}_2(\Gamma_2(T)) &:= \boldsymbol{C}_2(\boldsymbol{F}_2(\Gamma_2(T))) \\ &= \frac{-64380401708754(\Gamma_2(T))^4 + 2317693623118880(\Gamma_2(T))^2 - 20859235062503544}{407097080892401(\Gamma_2(T))^4 - 14655494912126204(\Gamma_2(T))^2 + 131899454209147204}. \end{aligned} \tag{49}$$

Note that $(A, B) = (0, \boldsymbol{F}_2(\Gamma_2(T_*)))$ satisfies (C1) and (C2). Hence, using the same arguments as in the proof of Lemma 5.1, one knows that $\boldsymbol{D}_2(\Gamma_2(T_*)) \in \mathbb{Z}_{29}$.

We see that

$$u := u_0 + A\boldsymbol{D}_1(\Gamma_1(T_*)) = u_0 = 728799 \neq 0.$$

Furthermore, it follows from Lemma 6.1(2), (48) and (49) that

$$v := v_0 + B\boldsymbol{D}_2(\Gamma_2(T_*)) = \boldsymbol{G}_2(\boldsymbol{F}_2(\Gamma_2(T_*)))$$

is defined, namely, not infinity and nonzero. Hence $(A, B) = (0, \boldsymbol{F}_2(\Gamma_2(T_*)))$ satisfies Lemma 5.1(C3).

Set

$$\begin{aligned} \alpha &:= \alpha_0 + 2p^2\boldsymbol{D}_2(\Gamma_2(T_*)) = 133 + 1682\boldsymbol{D}_2(\Gamma_2(T_*)) = \boldsymbol{D}_2^*(\Gamma_2(T_*)), \\ \beta &:= \beta_0 = 29, \\ \gamma &:= \gamma_0 = 27, \end{aligned}$$

where $\boldsymbol{D}_2^*(\Gamma_2(T)) \in \mathbb{Q}(T)$ is the rational function defined by (39). Recall from there that

$$\boldsymbol{D}_2^*(\Gamma_2(T)) = \frac{\Sigma_{2,1}(T)}{\Sigma_{2,2}(T)},$$

where $\Sigma_{2,1}(T)$, $\Sigma_{2,2}(T)$ are defined by (40), (41), respectively.

By Lemma 5.1, we know that (α, β, γ) satisfies (A1)–(A5) and (B1). By (11) and (12), we know that

$$\begin{aligned} P &= 29(133 + 1682\boldsymbol{D}_2(\Gamma_2(T_*)))^2 - 40600 = 29(\boldsymbol{D}_2^*(\Gamma_2(T_*)))^2 - 40600, \\ Q &= Q_0 = 1711. \end{aligned}$$

It is not difficult to see that the curve $\mathcal{C}_{n,T_*,(29,1,3)}^{(133,29,27)}$ defined in Theorem 6.8(ii) is the smooth projective model of the affine curve defined by (13).

By Corollary 6.7, we know that $\mathcal{P}_{2,T_*}(x)$ is separable, and hence we deduce that condition (S) in Theorem 4.1 is true. Since $\boldsymbol{D}_2(\Gamma_2(T_*)) \in \mathbb{Z}_{29}$, we see that

$$-2\Big(\frac{\gamma}{\alpha}\Big)^2 \equiv 8 \pmod{29}.$$

Since $n \in \boldsymbol{A}_2$, we deduce that

$$n \not\equiv -2\Big(\frac{\gamma}{\alpha}\Big)^2 \pmod{29}.$$

Thus (A6) holds if n is odd, and (B2) holds if $n \equiv 2 \pmod 4$. By Corollaries 4.6 and 4.8, we deduce that for each $n \in \boldsymbol{A}_2$, the curve $\mathcal{C}_{n,T_*,(29,1,3)}^{(133,29,27)}$ satisfies CHP and NZC. Since T_* is an arbitrary rational number, Theorem 6.8(ii) follows. □

Acknowledgements

I am grateful to Emmanuel Kowalski for his encouragement. I thank the referee for useful comments. I was supported by a postdoctoral fellowship in the Department of Mathematics at the University of British Columbia during the writing of this paper.

References

[Amer 1976] M. Amer, *Quadratische Formen über Funktionenkörpern*, thesis, Johannes Gutenberg University, Mainz, 1976.

[Bhargava et al. 2013] M. Bhargava, B. H. Gross, and X. Wang, "Pencils of quadrics and the arithmetic of hyperelliptic curves", preprint, 2013. arXiv 1310.7692

[Borevich and Shafarevich 1966] Z. I. Borevich and I. R. Shafarevich, *Number theory*, Pure and Applied Mathematics **20**, Academic Press, New York, 1966. MR 33 #4001 Zbl 0145.04902

[Brumer 1978] A. Brumer, "Remarques sur les couples de formes quadratiques", *C. R. Acad. Sci. Paris Sér. A* **286**:16 (1978), 679–681. MR 58 #16502 Zbl 0392.10021

[Cohen 2007] H. Cohen, *Number theory, I: Tools and Diophantine equations*, Graduate Texts in Mathematics **239**, Springer, New York, 2007. MR 2008e:11001 Zbl 1119.11001

[Colliot-Thélène and Poonen 2000] J.-L. Colliot-Thélène and B. Poonen, "Algebraic families of nonzero elements of Shafarevich–Tate groups", *J. Amer. Math. Soc.* **13**:1 (2000), 83–99. MR 2000f:11067 Zbl 0951.11022

[Colliot-Thélène et al. 1980] J.-L. Colliot-Thélène, D. F. Coray, and J.-J. Sansuc, "Descente et principe de Hasse pour certaines variétés rationnelles", *J. Reine Angew. Math.* **320** (1980), 150–191. MR 82f:14020 Zbl 0434.14019

[Coray and Manoil 1996] D. F. Coray and C. Manoil, "On large Picard groups and the Hasse principle for curves and $K3$ surfaces", *Acta Arith.* **76**:2 (1996), 165–189. MR 97j:14038 Zbl 0877.14005

[Corn 2007] P. Corn, "The Brauer–Manin obstruction on del Pezzo surfaces of degree 2", *Proc. Lond. Math. Soc.* (3) **95**:3 (2007), 735–777. MR 2009a:14027 Zbl 1133.14022

[Dong Quan 2014] N. N. Dong Quan, "From separable polynomials to nonexistence of rational points on certain hyperelliptic curves", *J. Aust. Math. Soc.* **96**:3 (2014), 354–385. MR 3217721 Zbl 06324731

[Iwaniec 1974] H. Iwaniec, "Primes represented by quadratic polynomials in two variables", *Acta Arith.* **24**:5 (1974), 435–459. MR 49 #7210 Zbl 0271.10043

[Lind 1940] C.-E. Lind, *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven vom Geschlecht Eins*, thesis, University of Uppsala, 1940. MR 9,225c Zbl 0025.24802

[Poonen 2001] B. Poonen, "An explicit algebraic family of genus-one curves violating the Hasse principle", *J. Théor. Nombres Bordeaux* **13**:1 (2001), 263–274. MR 2002e:14036 Zbl 1046.11038

[Poonen 2008] B. Poonen, "Rational points on varieties", Lecture notes, MIT/UCB, 2008, http://www-math.mit.edu/~poonen/papers/Qpoints.pdf.

[Reichardt 1942] H. Reichardt, "Einige im Kleinen überall lösbare, im Grossen unlösbare Diophantische Gleichungen", *J. Reine Angew. Math.* **184** (1942), 12–18. MR 5,141c Zbl 0026.29701

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Skorobogatov 2001] A. N. Skorobogatov, *Torsors and rational points*, Cambridge Tracts in Mathematics **144**, Cambridge University Press, 2001. MR 2002d:14032 Zbl 0972.14015

[Viray 2012] B. Viray, "Failure of the Hasse principle for Châtelet surfaces in characteristic 2", *J. Théor. Nombres Bordeaux* **24**:1 (2012), 231–236. MR 2914907 Zbl 1285.11095

Received November 27, 2013.

NGUYEN NGOC DONG QUAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TX 78712
UNITED STATES

dongquan.ngoc.nguyen@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.183

F-ZIPS WITH ADDITIONAL STRUCTURE

RICHARD PINK, TORSTEN WEDHORN AND PAUL ZIEGLER

Let $\mathbb{F}_q$ be a fixed finite field of cardinality q. An F-zip over a scheme S over $\mathbb{F}_q$ is a certain object of semilinear algebra consisting of a locally free sheaf of $\mathcal{O}_S$-modules with a descending filtration and an ascending filtration and a Frob_q-twisted isomorphism between the respective graded sheaves. In this article we define and systematically investigate what might be called "F-zips with a G-structure", for an arbitrary reductive linear algebraic group G over $\mathbb{F}_q$.

These objects come in two incarnations. One incarnation is an exact $\mathbb{F}_q$-linear tensor functor from the category of finite dimensional representations of G over $\mathbb{F}_q$ to the category of F-zips over S. Locally any such functor has a type χ, which is a cocharacter of G_k for a finite extension k of $\mathbb{F}_q$ that determines the ranks of the graded pieces of the filtrations. The other incarnation is a certain G-torsor analogue of the notion of F-zips. We prove that both incarnations define stacks that are naturally equivalent to a quotient stack of the form $[E_{G,\chi}\backslash G_k]$ that was studied in our earlier paper (*Doc. Math.* 16 (2011), 253–300). By the results of this work they are therefore smooth algebraic stacks of dimension 0 over k. Using our previous work we can also classify the isomorphism classes of such objects over an algebraically closed field, describe their automorphism groups, and determine which isomorphism classes can degenerate into which others.

For classical groups we can deduce the corresponding results for twisted or untwisted symplectic, orthogonal, or unitary F-zips, a part of which has been described before by Moonen and Wedhorn (*Int. Math. Res. Not.* 2004:72, 3855–3903). The results can be applied to the algebraic de Rham cohomology of smooth projective varieties (or generalizations thereof) and to truncated Barsotti–Tate groups of level 1. In addition, we hope that our systematic group theoretical approach will help to understand the analogue of the Ekedahl–Oort stratification of the special fibers of arbitrary Shimura varieties.

MSC2010: primary 14F40, 14L30, 18D10, 20G40; secondary 14L05.
Keywords: F-zips, Ekedahl–Oort strata, De Rham cohomology.

1. Introduction

1A. ***Background.*** Let $X \to S$ be a smooth proper morphism of schemes in characteristic $p > 0$ whose Hodge spectral sequence degenerates and is compatible with base change. In [Moonen and Wedhorn 2004] it was shown that its relative De Rham cohomology $H^\bullet_{\mathrm{DR}}(X/S)$ carries the structure of a so-called *F-zip* over S, namely: It is a locally free sheaf of $\mathcal{O}_S$-modules of finite rank together with two filtrations (the "Hodge" and the "conjugate" filtration) and a Frobenius linear isomorphism between the associated graded vector spaces (the "Cartier isomorphism"). They showed that the isomorphism classes of F-zips of fixed dimension n and with a fixed type of Hodge filtration over an algebraically closed field are in natural bijection with the orbits under $\mathrm{GL}_{n,k}$ in a variant Z'_I of the varieties Z_I studied by Lusztig [2004a; 2004b]. They studied the analogous varieties Z'_I for arbitrary reductive groups G defined over a finite field and determined the G-orbits in them as analogues of the G-stable pieces in Z_I. By specializing G to classical groups they deduced from this a classification of F-zips with certain additional structure, for example, with a nondegenerate symmetric or alternating form.

In [Pink et al. 2011] the present authors showed that the quotient stack $[G\backslash Z'_I]$ is isomorphic to a quotient stack of the form $[E_\chi\backslash G]$, where E_χ is certain linear algebraic group depending on the choice of a cocharacter χ of G. We studied this quotient stack in detail, classifying the E_χ-orbits in G by a subset of the Weyl group of G and describing their closure relation using a variant of the Bruhat order.

1B. ***Main idea.*** The aim of this paper is to define and investigate what might be called "F-zips with a G-structure", for an arbitrary reductive linear algebraic group G.

As a guideline let us first review the analogous case of vector bundles. Recall that giving a vector bundle E of constant rank n on a manifold or a scheme S over a field k is equivalent to giving the associated $\mathrm{GL}_{n,k}$-torsor. For a subgroup $G \subset \mathrm{GL}_{n,k}$, the choice of a G-torsor I within this $\mathrm{GL}_{n,k}$-torsor is called a G-structure on E. The vector bundle E can be recovered as the pushout of I with the given n-dimensional representation of G, so giving a vector bundle with a G-structure is really equivalent to giving a G-torsor I.

At this point we can disregard the special role of the original representation and form the pushout of I with all finite dimensional representations of G. This yields an exact k-linear tensor functor from the Tannakian category G-$\mathtt{Rep}$ of finite dimensional representations of G over k to the category of vector bundles on S, which is known (for example by Nori [1976, Proposition 2.9]) to be again equivalent to giving I. Altogether such a functor is therefore equivalent to giving a vector bundle with a G-structure on S.

These observations suggest that objects with a G-structure in a more general exact k-linear tensor category $\mathcal{T}$ should be equivalent to, or might even be defined as, exact k-linear tensor functors $G\text{-}\mathtt{Rep} \to \mathcal{T}$, and that they should be equivalent to G-torsor analogues of the objects in $\mathcal{T}$. In the cases of graded or filtered vector bundles equivalences of this kind were in fact derived in some cases in [Saavedra Rivano 1972, IV.1–2] and in general in [Ziegler 2011]. The principle was also applied to F-isocrystals with additional structure by Rapoport and Richartz [1996].

The program for the present paper is therefore to develop this approach for the category of F-zips and to classify the ensuing objects using the results of [Pink et al. 2011].

1C. *F-zips with a G-structure.* Let $\mathbb{F}_q$ be a fixed finite field with q elements, and consider a scheme S over $\mathbb{F}_q$. Recall from [Moonen and Wedhorn 2004] that an *F-zip over S* is a tuple $\underline{\mathcal{M}} = (\mathcal{M}, C^\bullet, D_\bullet, \varphi_\bullet)$ consisting of a locally free sheaf of $\mathcal{O}_S$-modules of finite rank $\mathcal{M}$ on S, a descending filtration $C^\bullet$ and an ascending filtration $D_\bullet$ of $\mathcal{M}$, and an $\mathcal{O}_S$-linear isomorphism $\varphi_i \colon (\mathrm{gr}^i_C \mathcal{M})^{(q)} \xrightarrow{\sim} \mathrm{gr}^D_i \mathcal{M}$ for every $i \in \mathbb{Z}$, where $(\ )^{(q)}$ denotes the pullback by the Frobenius morphism $x \mapsto x^q$. In a natural way (see Section 6) the F-zips over S are the objects of an exact $\mathbb{F}_q$-linear tensor category $F\text{-}\mathtt{Zip}(S)$.

Let G be a reductive linear algebraic group over $\mathbb{F}_q$, and let k be a finite extension of $\mathbb{F}_q$. In the body of the paper we consider not necessarily connected groups, but to simplify notations in this introduction we stick to a connected group G. For simplicity we also assume that G splits over k, so that every conjugacy class of cocharacters of G over any extension field of k possesses a representative that is defined over k. Let $G\text{-}\mathtt{Rep}$ denote the $\mathbb{F}_q$-linear abelian tensor category of finite-dimensional rational representations of G over $\mathbb{F}_q$. The role of "F-zips with a G-structure" is played by the following objects:

Definition 1.1 (cf. Definition 7.1). For any scheme S over k, a *G-zip functor over S* is an exact $\mathbb{F}_q$-linear tensor functor

$$\mathfrak{z}\colon\ G\text{-}\mathtt{Rep} \to F\text{-}\mathtt{Zip}(S).$$

As S varies, these objects form a category $G\text{-}\mathtt{ZipFun}$ fibered in groupoids over the category of schemes over k. It is not hard to show that $G\text{-}\mathtt{ZipFun}$ is a stack over k (see Proposition 7.2). It possesses a natural decomposition that is indexed by conjugacy classes of cocharacters of G, defined as follows.

Let χ be a cocharacter of the group G_k obtained from G by base change. Then χ induces a grading on $V_k := V \otimes_{\mathbb{F}_q} k$ for every representation V of G and thus an $\mathbb{F}_q$-linear tensor functor γ_χ from $G\text{-}\mathtt{Rep}$ to the category of graded k-vector spaces. On the other hand, any G-zip functor $\mathfrak{z}$ over S induces an $\mathbb{F}_q$-linear tensor functor from G-Rep to the category of graded locally free sheaves of $\mathcal{O}_S$-modules on S which sends V to $\mathrm{gr}^\bullet_C \circ \mathfrak{z}(V)$.

Definition 1.2 (cf. Definitions 5.3 and 7.3). A G-zip functor $\mathfrak{z}$ over S is called *of type* χ if the graded fiber functors $\mathrm{gr}^\bullet_C \circ \mathfrak{z}$ and γ_χ are fpqc-locally isomorphic. The substack of G-ZipFun of G-zip functors of type χ is denoted $G\text{-}\mathtt{ZipFun}^\chi$.

Theorem 1.3 (cf. Corollary 7.6). *Every G-zip functor over a connected scheme has a type. Each $G\text{-}\mathtt{ZipFun}^\chi$ is an open and closed substack of G-ZipFun.*

In Section 8 we work out equivalent but simpler descriptions of G-zip functors for certain classical groups. The rough idea in all cases is that any G-zip functor is already determined up to unique isomorphism by its restriction to a certain finite subcategory of G-Rep and that, conversely, any suitable functor from this subcategory to the category of F-zips extends to a G-zip functor on all of G-Rep. For instance, giving a GL_n-zip functor is equivalent to giving an F-zip $\underline{\mathcal{M}}$ of constant rank n, and giving an SL_n-zip functor is equivalent to giving an F-zip $\underline{\mathcal{M}}$ of constant rank n together with an isomorphism between its highest exterior power $\Lambda^n\underline{\mathcal{M}}$ and the unit object $\underline{\mathbb{1}}(0)$. Similarly, giving an Sp_n-zip, resp. O_n-zip functor is equivalent to giving a symplectic, resp. orthogonal F-zip of constant rank n, by which we mean an F-zip $\underline{\mathcal{M}}$ of constant rank n together with an epimorphism of F-zips $\Lambda^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathbb{1}}(0)$, resp. $\mathrm{S}^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathbb{1}}(0)$, whose underlying pairing of locally free sheaves is nondegenerate everywhere. We also discuss the relation between U_n-zip functors and unitary F-zips, as well as twisted versions of these equivalences associated to the groups of similitudes CSp_n and CO_n and CU_n.

1D. *G-zips.* To describe the stack of G-zip functors $G\text{-}\mathtt{ZipFun}^\chi$ in detail we use the following G-torsor analogue of F-zips. Let G and χ be as above, and let $P_\pm = L \ltimes U_\pm$ be the associated pair of opposite parabolic subgroups of G_k.

Definition 1.4 (cf. Definition 3.1). A *G-zip of type* χ over a scheme S over k is a tuple $\underline{I} = (I, I_+, I_-, \iota)$ consisting of a right G_k-torsor I over S, a right P_+-torsor $I_+ \subset I$, a right $P_-^{(q)}$-torsor $I_- \subset I$, and an isomorphism of $L^{(q)}$-torsors $\iota\colon I_+^{(q)}/U_+^{(q)} \xrightarrow{\sim} I_-/U_-^{(q)}$.

As S varies, these objects form a category $G\text{-}\mathtt{Zip}^\chi$ fibered in groupoids over the category of schemes over k. It is not hard to show that $G\text{-}\mathtt{Zip}^\chi$ is a stack over k (see Proposition 3.2).

To G and χ we can also associate a natural *algebraic zip datum* in the sense of [Pink et al. 2011] (see Definition 3.6). The associated *zip group* is the linear algebraic group

$$E_{G,\chi} := \{(\ell u_+, \ell^{(q)}u_-) \mid \ell \in L, u_+ \in U_+, u_- \in U_-^{(q)}\} \subset P_+ \times_k P_-^{(q)}$$

which acts from the left hand side on G_k by $(p_+, p_-)\cdot g := p_+ g p_-^{-1}$. We can thus form the algebraic quotient stack $[E_{G,\chi}\backslash G_k]$.

Theorem 1.5 (cf. Proposition 3.11, Theorem 7.13, and Corollary 3.12). *The stacks $G\text{-}\mathtt{ZipFun}^\chi$ and $G\text{-}\mathtt{Zip}^\chi$ and $[E_{G,\chi}\backslash G_k]$ are naturally equivalent. They are smooth algebraic stacks of dimension* 0 *over k.*

In particular, it is equivalent to give a G-zip functor of type χ over S, or a G-zip of type χ over S, or a morphism $S \to [E_{G,\chi}\backslash G_k]$ over k. The equivalences are in fact obtained by explicit constructions (see Sections 3D and 7B).

1E. *Classification.* Using the results of [Pink et al. 2011] we can now describe the stack of G-zips of type χ in detail. By Theorem 1.5, all the following statements hold equivalently for G-zip functors of type χ. By the results of Section 8 they also hold for symplectic, orthogonal, resp. unitary F-zips, and so on.

Let W be the Weyl group of G, let $I \subset W$ be the subset of simple reflections corresponding to P_+, and let W_I be the subgroup of W generated by I. Let ${}^I W$ be the set of elements $w \in W$ that are of minimal length in their right coset $W_I w$. We endow ${}^I W$ with a certain partial order $\preceq$ which is somewhat complicated to describe (and in general strictly finer than the Bruhat order; see (3.16) and Example 3.23). This turns ${}^I W$ into a finite topological space (see Proposition 2.1), which we can compare with the topological space underlying the algebraic stack $G\text{-}\mathtt{Zip}^\chi$ (see Section 2B).

Theorem 1.6 (cf. Theorem 3.20). *The topological space underlying $G\text{-}\mathtt{Zip}^\chi$ is naturally homeomorphic to ${}^I W$. In particular, there is a natural bijection between the set of isomorphism classes of G-zips of type χ over an algebraically closed field K containing k and the set ${}^I W$.*

For any G-zip $\underline{I}$ of type χ over a scheme S over k we thus obtain a finite stratification of S by the isomorphism type of $\underline{I}$. This generalizes the F-zip stratification defined in [Moonen and Wedhorn 2004] in the case $G = \mathrm{GL}_{n,\mathbb{F}_q}$, as well as the Ekedahl–Oort stratification of the moduli space of g-dimensional principally polarized abelian varieties in the case $G = \mathrm{CSp}_{2g,\mathbb{F}_q}$. The partial order $\preceq$ yields information on the closure relations between these strata (see (3.29) and Proposition 3.30). Using a result from [Wedhorn and Yatsyshyn 2014] we can also deduce a purity result (see Proposition 3.33). Furthermore, the description of point stabilizers in $E_{G,\chi}$ from [Pink et al. 2011, Theorem 8.1] yields information on automorphism groups of G-zips; in particular:

Theorem 1.7 (cf. Proposition 3.34). *The automorphism group scheme of the G-zip of type χ over an algebraically closed field K containing k corresponding to $w \in {}^I W$ is an extension of a finite group* (*see Proposition 3.34 for its precise description*) *by* *a connected unipotent group of dimension* $\dim(G/P_+) - \ell(w)$, *where* $\ell(\,)$ *denotes the length function on the Coxeter group* W.

1F. *Applications.* In Section 9 we study the de Rham cohomology of a smooth proper Deligne–Mumford stack $\mathscr{X} \to S$ whose Hodge spectral sequence degenerates and is compatible with arbitrary base change. For all $d \geq 0$ we obtain an F-zip $\underline{H}^d_{\mathrm{DR}}(\mathscr{X}/S)$. If $\mathscr{X}$ is a scheme, the cup product induces for all $d, e \geq 0$ a morphism of F-zips

$$\cup\colon \underline{H}^d_{\mathrm{DR}}(\mathscr{X}/S) \otimes \underline{H}^e_{\mathrm{DR}}(\mathscr{X}/S) \longrightarrow \underline{H}^{d+e}_{\mathrm{DR}}(\mathscr{X}/S).$$

If $\mathscr{X} \to S$ is a smooth proper morphism of schemes with geometrically connected fibers of dimension n, the cup product turns $\underline{H}^n_{\mathrm{DR}}(\mathscr{X}/S)$ into a twisted symplectic or orthogonal F-zip, depending on the parity of n.

In Section 9C we attach an F-zip to any truncated Barsotti–Tate group of level 1 over a scheme S of characteristic p. This construction improves the one given in [Moonen and Wedhorn 2004] where S was assumed to be perfect.

In addition, we hope that our results can be applied to the special fibers of arbitrary Shimura varieties, where G is the reduction modulo p of the connected reductive linear algebraic group over $\mathbb{Q}$ that gives rise to the Shimura variety. In that case our systematic group theoretical approach should prove especially valuable. For good reductions of Shimura varieties of Hodge type some progress has already been made. C. Zhang [2013] has defined a smooth morphism from the special fiber of Kisin's integral model to the stack of G-zips of a certain type χ yielding a description of Ekedahl–Oort strata for the special fiber. This has been used by D. Wortmann [2013] to prove that the conjectured candidate for the generic Newton stratum is indeed open and dense in the special fiber.

1G. *Contents of the paper.* As a preparation we begin by recalling some properties of quotient stacks in Section 2.

In Section 3 we first introduce some general notation used throughout the rest of the article. We mostly work with a not necessarily connected linear algebraic group $\hat{G}$ over $\mathbb{F}_q$ whose identity component G is reductive. Besides a cocharacter χ of G_k, the basic data also requires the choice of a subgroup Θ of the group of connected components of the stabilizer of χ. We then define the general notion of $\hat{G}$-zips of type (χ, Θ) and prove that they form a smooth algebraic stack of dimension 0 over k that is naturally isomorphic to a quotient stack of the form $[E_{\hat{G},\chi,\Theta}\backslash \hat{G}_k]$ that was studied in [Pink et al. 2011]. The remainder of Section 3 contains an assortment of results on the topological space underlying this stack, on the associated stratification, and on automorphisms.

We use Section 4 to collect some generalities concerning locally free sheaves, gradings, filtrations, and alternating and symmetric powers. In Section 5 we recall some results of the third author on filtered and graded fiber functors on the Tannakian category $\hat{G}$-Rep.

In Section 6 we endow the category of F-zips over a scheme S with the structure of an exact rigid tensor category. Section 7 then contains the definition of $\hat{G}$-zip functors. Here we use the results recalled in Section 5 to prove that every $\hat{G}$-zip functor over a connected scheme S has a type χ. With the unique maximal possible choice of Θ we then establish a natural isomorphism between the stack of $\hat{G}$-zip functors of type χ and the stack of $\hat{G}$-zips of type (χ, Θ). Consequently the stack of $\hat{G}$-zip functors of type χ is also a smooth algebraic stack of dimension 0 that is naturally isomorphic to $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$.

In Section 8 we go through a list of eight classical groups, in each case describing an equivalence between $\hat{G}$-zip functors and F-zips of a given rank with a certain embellishment such as an alternating or symmetric or hermitian form within the category of F-zips.

Finally, in Section 9 we discuss applications to the algebraic de Rham cohomology of certain Deligne–Mumford stacks and to truncated Barsotti–Tate groups of level 1.

2. General properties of quotient stacks

As a preparation we discuss some general properties of algebraic group actions and quotient stacks.

2A. ***Closure relation for an algebraic group action.*** First recall that a topological space Z is called T_0 (or Kolmogorov) if for any two distinct points, at least one of them possesses a neighborhood that does not contain the other. Abbreviating the closure of a subset by $\overline{(\ )}$, as usual, this is equivalent to saying that for any $z', z \in Z$ we have $z' = z$ if and only if both $z' \in \overline{\{z\}}$ and $z \in \overline{\{z'\}}$. On the other hand, recall that a partial order $\preceq$ on a set Z is a transitive binary relation which is antisymmetric in the sense that $z' = z$ if and only if both $z' \preceq z$ and $z \preceq z'$. With these observations the following well-known fact is easy to prove:

Proposition 2.1. *For any finite T_0 topological space Z the relation $z' \preceq z :\Leftrightarrow z' \in \overline{\{z\}}$ is a partial order on Z. Conversely, any partial order on a finite set Z arises in this way from a unique T_0 topology on Z. Moreover, a map between finite T_0 spaces is continuous if and only if it preserves the associated partial orders.*

Next consider a field k with an algebraic closure $\bar{k}$ and the associated absolute Galois group $\Gamma := \mathrm{Aut}(\bar{k}/k)$. Let X be a scheme of finite type over k, and let H be a linear algebraic group over k which acts on X from the left by a morphism $H \times_k X \to X$. Then every $H(\bar{k})$-orbit $\mathbb{O} \subset X(\bar{k})$ is locally closed for the Zariski topology on $X(\bar{k})$. Moreover, its closure $\overline{\mathbb{O}}$ is again $H(\bar{k})$-invariant and therefore a union of orbits, and we have $\dim(\overline{\mathbb{O}} \setminus \mathbb{O}) < \dim(\mathbb{O})$. From this it follows that the set of orbits

$$\Xi := H(\bar{k})\backslash X(\bar{k})$$

with the induced quotient topology is a T_0 space.

Assume now that Ξ is finite. Then by Proposition 2.1 the topology on Ξ corresponds to the partial order $\preceq$ on Ξ defined by $\mathcal{O}' \preceq \mathcal{O} :\Leftrightarrow \mathcal{O}' \subset \bar{\mathcal{O}}$. Also, the Galois group Γ acts on Ξ preserving the topology and the partial order. Thus the quotient set $\Gamma\backslash\Xi$ again inherits a quotient topology which is T_0 and corresponds to a partial order described in the same fashion. The set $\Gamma\backslash\Xi$ is in natural bijection with the set of algebraic H-orbits in X, that is, with the set of nonempty H-invariant locally closed reduced subschemes that do not possess nonempty H-invariant locally closed proper subschemes.

2B. ***Quotient stacks.*** This article will require a certain familiarity with the notion of stacks. Recall that a stack over a scheme S is a category fibered in groupoids over the category $((\boldsymbol{Sch}/S))$ of schemes over S which satisfies effective descent with respect to any fpqc morphism. The morphisms of stacks are functors, and so instead of equality of morphisms one often only has isomorphisms of functors. For the technical definition of an algebraic stack see [Laumon and Moret-Bailly 2000, Definition 4.1]. Let us recall only that every scheme can be considered as an algebraic stack, and for every algebraic stack $\mathcal{X}$ there exists a smooth surjective morphism from a scheme $X \to \mathcal{X}$. Many concepts and properties of schemes and morphisms of schemes have analogues for algebraic stacks. For example, there exist natural fiber products and pullbacks of algebraic stacks, and many properties of algebraic stacks and of morphisms of algebraic stacks are tested using pullbacks to schemes.

Every algebraic stack $\mathcal{X}$ possesses an underlying topological space $|\mathcal{X}|$, defined in [Laumon and Moret-Bailly 2000, Section 5]. An element of $|\mathcal{X}|$ is an equivalence class of morphisms $\operatorname{Spec} K \to \mathcal{X}$ for fields K, where two morphisms $\operatorname{Spec} K_1 \to \mathcal{X}$ and $\operatorname{Spec} K_2 \to \mathcal{X}$ are equivalent if and only if there exists a common field extension K such that the composite morphisms $\operatorname{Spec} K \to \operatorname{Spec} K_i \to \mathcal{X}$ are isomorphic. The open subsets of $|\mathcal{X}|$ are the subsets $|\mathcal{U}|$ for all open substacks $\mathcal{U} \subset \mathcal{X}$. If $\mathcal{X}$ is represented by a scheme X, then $|\mathcal{X}|$ is homeomorphic to the topological space underlying X.

Consider now the situation of Section 2A, where H acts from the left hand side on a scheme X over k. The quotient stack $[H\backslash X]$ is then defined as follows. For any scheme S over k the category $[H\backslash X](S)$ has as objects the pairs consisting of a left H-torsor $T \to S$ and an H-equivariant morphism $f : T \to X$ over k. A morphism $(T, f) \to (T', f')$ in $[H\backslash X](S)$ is a morphism $g : T \to T'$ of H-torsors over S such that $f' \circ g = f$, and composition is defined in the obvious way. With the evident notion of pullback under morphisms $S' \to S$ this turns the whole collection of categories $[H\backslash X](S)$ into a stack over k that is denoted $[H\backslash X]$. This is an algebraic stack by [Laumon and Moret-Bailly 2000, Proposition 10.13.1], and it possesses a natural surjective morphism $X \to [H\backslash X]$.

Assume that the set Ξ of orbits over $\bar{k}$ is finite, so that it and its quotient $\Gamma\backslash\Xi$ carry the natural topologies described in Section 2A. As a reformulation of [Wedhorn 2001, Section 4.4] we then have:

Proposition 2.2. *There is a natural homeomorphism* $\Gamma\backslash\Xi \cong \bigl|[H\backslash X]\bigr|$.

Now consider any algebraic H-orbit $Y \subset X$. Then Y is a locally closed reduced subscheme of X, and so $[H\backslash Y]$ is a locally closed reduced substack of $[H\backslash X]$. Varying Y we thus obtain a stratification

$$[H\backslash X] \text{“="} \bigsqcup_{\Gamma\backslash\Xi} [H\backslash Y] \tag{2.3}$$

in the sense that for any scheme S and any morphism $S \to [H\backslash X]$ we obtain a disjoint decomposition of S into locally closed subschemes $S \times_{[H\backslash X]} [H\backslash Y]$.

Moreover, assume that k is perfect, and consider any point $y \in Y(\bar{k})$. Then $Y_{\bar{k}}$ is again reduced and hence the disjoint union of the reduced $H_{\bar{k}}$-orbit $\mathbb{O}(y)$ of y and a (possibly empty) finite collection of Γ-conjugates thereof. Being a reduced orbit $\mathbb{O}(y)$ is smooth over $\bar{k}$; hence Y is smooth over k, and so $[H\backslash Y]$ is a smooth algebraic stack over k by definition (see [Laumon and Moret-Bailly 2000, Définition 4.14]). Furthermore, as the smooth morphism $X \to [H\backslash X]$ preserves codimension, we have

$$\operatorname{codim}([H\backslash Y], [H\backslash X]) = \operatorname{codim}(Y, X) = \operatorname{codim}(\mathbb{O}(y), X_{\bar{k}}). \tag{2.4}$$

Also, the automorphisms of an object of a quotient stack can be described as follows. Consider a scheme S over k, a point $x \in X(S)$, and let $\bar{x} \in [H\backslash X](S)$ denote the image of x under the canonical morphism $X \to [H\backslash X]$. Denote by $\underline{\mathrm{Aut}}(\bar{x})$ the sheaf of groups on the category on schemes over S that attaches to $S' \to S$ the group of automorphisms of the base change $\bar{x}_{S'}$ in the category $[H\backslash X](S')$. On the other hand let $\mathrm{Stab}_{H_S}(x)$ denote the closed subgroup scheme of $H_S := H \times_k S$ whose S'-valued points consist of those $h \in H(S')$ which satisfy $h \cdot x_{S'} = x_{S'}$.

Proposition 2.5. *There is a natural isomorphism* $\underline{\mathrm{Aut}}(\bar{x}) \cong \mathrm{Stab}_{H_S}(x)$.

Proof. In the construction of $[H\backslash X]$, the point $\bar{x} \in [H\backslash X](S)$ is represented by the trivial H-torsor $H_S \to S$ together with the morphism $H_S \to X$, $h \mapsto hx$. The automorphisms of the trivial left H-torsor $H_{S'} \to S'$ are precisely the morphisms $h \mapsto hg$ for all sections $g \in H(S')$. Thus $\underline{\mathrm{Aut}}(\bar{x})(S') = \mathrm{Aut}_{[H\backslash X](S')}(\bar{x}_{S'})$ is the group of automorphisms $h \mapsto hg$ of $H_{S'} \to S'$ with $g \in H(S')$, such that the two morphisms $H_{S'} \to X$ given by $h \mapsto hx$ and $h \mapsto hg \mapsto hgx$ coincide. But these conditions are equivalent to $g \in \mathrm{Stab}_{H_S}(x)(S')$. $\square$

3. $\hat{G}$-zips

3A. ***General notation.*** Let $\mathbb{F}_q$ be a finite field of order q, and let k be a finite overfield of $\mathbb{F}_q$. By a linear algebraic group over k we mean a reduced affine group scheme of finite type over k. We do not generally assume it to be connected. Throughout we denote a linear algebraic group over k by $\hat{H}$, its identity component by H, and the finite étale group scheme of connected components by $\pi_0(\hat{H}) := \hat{H}/H$; and similarly for other letters of the alphabet. Note that the unipotent radical R_uH of H is a normal subgroup of $\hat{H}$. Any homomorphism of algebraic groups $\hat{\varphi}\colon \hat{G} \to \hat{H}$ restricts to a homomorphism $\varphi\colon G \to H$.

Let S be a scheme over k. By an $\hat{H}$-torsor I over S we will mean a right $\hat{H}$-torsor over S for the fpqc-topology, unless mentioned otherwise. In other words I is a scheme over S together with a right action $I \times_k \hat{H} \to I$ written $(i, h) \mapsto ih$, such that the morphism $I \times_k \hat{H} \to I \times_S I$, $(i, h) \mapsto (i, ih)$ is an isomorphism and there exists an fpqc-covering $S' \to S$ such that $I(S') \neq \varnothing$. Any section in $I(S')$ then induces an isomorphism $\hat{H} \times_k S' \xrightarrow{\sim} I \times_S S'$ over S'. By faithfully flat descent for $S' \to S$ we can therefore deduce that every $\hat{H}$-torsor over S is affine and faithfully flat over S. Moreover, since k is perfect, the reduced group scheme $\hat{H}$ is automatically smooth, and hence I is smooth over S. Thus by [Grothendieck 1967, Corollaire (17.16.3)] there already exists a surjective étale morphism $S' \to S$ such that $I(S') \neq \varnothing$.

Any scheme S over k possesses a natural q-th power Frobenius morphism $S \to S$, which is the identity on the underlying topological space and the map $x \mapsto x^q$ on the structure sheaf. The pullback of a scheme or a sheaf or a morphism over S under this Frobenius morphism is denoted by $(\ )^{(q)}$. For example, the pullback of a linear algebraic group $\hat{H}$ over k is a linear algebraic group $\hat{H}^{(q)}$ over k, and the pullback of an $\hat{H}$-torsor I over S is an $\hat{H}^{(q)}$-torsor $I^{(q)}$ over S.

3B. ***The basic data.*** Let $\hat{G}$ be a linear algebraic group over $\mathbb{F}_q$ such that G is reductive, and let $\hat{G}_k$ denote its base extension to k. Let $\chi : \mathbb{G}_{m,k} \to G_k$ be a cocharacter over k, and let L denote its centralizer in G_k. There exist unique opposite parabolic subgroups $P_\pm = L \ltimes U_\pm \subset G_k$ with common Levi component L and unipotent radicals $U_\pm$, such that $\operatorname{Lie} U_+$ is the sum of the weight spaces of weights > 0, and $\operatorname{Lie} U_-$ is the sum of the weight spaces of weights < 0 in $\operatorname{Lie} G_k$ under $\operatorname{Ad} \circ \chi$. Note that the groups L and $P_\pm$ and $U_\pm$ are all connected.

By definition we have $L = \operatorname{Cent}_{\hat{G}_k}(\chi) \cap G_k$, and since L is connected, we have a canonical inclusion $\pi_0(\operatorname{Cent}_{\hat{G}_k}(\chi)) = \operatorname{Cent}_{\hat{G}_k}(\chi)/L \hookrightarrow \pi_0(\hat{G}_k)$. Let Θ be a subgroup scheme of $\pi_0(\operatorname{Cent}_{\hat{G}_k}(\chi))$, and $\hat{L}$ denote its inverse image in $\operatorname{Cent}_{\hat{G}_k}(\chi)$. Then L is the identity component of $\hat{L}$, and $\pi_0(\hat{L}) = \Theta \subset \pi_0(\hat{G}_k)$. Also, since χ is centralized by $\hat{L}$, and the subgroups $U_\pm$ depend only on χ, these subgroups are normalized by $\hat{L}$. Thus $\hat{P}_\pm := \hat{L} \ltimes U_\pm$ are algebraic subgroups of $\hat{G}_k$ with identity components $P_\pm$ and $\pi_0(\hat{P}_\pm) \cong \pi_0(\hat{L}) = \Theta$.

This data will remain fixed throughout the article.

For schemes S over k we are interested in (right) torsors over S with respect to the above algebraic groups. Consider any $\hat{G}_k$-torsor I over S. By a $\hat{P}_\pm$-torsor $I_\pm \subset I$ we mean a subscheme which is a $\hat{P}_\pm$-torsor with respect to the induced action of $\hat{P}_\pm$. For any $\hat{P}_\pm$-torsor $I_\pm$ over S, the quotient $I_\pm/U_\pm$ is a $\hat{P}_\pm/U_\pm$-torsor over S, which we can view as an $\hat{L}$-torsor under the canonical isomorphism $\hat{L} \xrightarrow{\sim} \hat{P}_\pm/U_\pm$.

On the other hand, the definition of $\hat{G}_k$ as a base extension from $\mathbb{F}_q$ induces a natural isomorphism $\hat{G}_k^{(q)} \cong \hat{G}_k$. Via this isomorphism we can consider $\chi^{(q)}$ again as a cocharacter of G_k, with associated subgroups $\hat{P}_\pm^{(q)} = \hat{L}^{(q)} \ltimes U_\pm^{(q)}$. Likewise, the $\hat{G}_k^{(q)}$-torsor $I^{(q)}$ becomes a $\hat{G}_k$-torsor in a natural way. Moreover, the pullback of a $\hat{P}_\pm$-torsor $I_\pm \subset I$ is a $\hat{P}_\pm^{(q)}$-torsor $I_\pm^{(q)} \subset I^{(q)}$.

3C. *The stack of $\hat{G}$-zips.*

Definition 3.1. Let S be a scheme over k.

(a) A *$\hat{G}$-zip of type* (χ, Θ) *over* S is a tuple $\underline{I} = (I, I_+, I_-, \iota)$ consisting of a (right) $\hat{G}_k$-torsor I over S, a $\hat{P}_+$-torsor $I_+ \subset I$, a $\hat{P}_-^{(q)}$-torsor $I_- \subset I$, and an isomorphism of $\hat{L}^{(q)}$-torsors $\iota\colon I_+^{(q)}/U_+^{(q)} \xrightarrow{\sim} I_-/U_-^{(q)}$.

(b) A *morphism* $(I, I_+, I_-, \iota) \to (I', I'_+, I'_-, \iota')$ of $\hat{G}$-zips of type (χ, Θ) over S consists of equivariant morphisms $I \to I'$ and $I_\pm \to I'_\pm$ that are compatible with the inclusions and the isomorphisms ι and ι'.

(c) The resulting *category of $\hat{G}$-zips of type* (χ, Θ) *over* S is denoted $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}(S)$.

If $\hat{G}$ is connected, we necessarily have $\Theta = 1$ and drop it from the notation, speaking simply of $\hat{G}$-zips of type χ over S and denoting their category by $\hat{G}\text{-}\mathtt{Zip}_k^{\chi}(S)$.

With the evident notion of pullback the $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}(S)$ form a fibered category over the category $((Sch/k))$ of schemes over k, which we denote $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$.

Proposition 3.2. *$\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ is a stack.*

Proof. Any morphism of $\hat{G}$-zips is an isomorphism; hence $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ is a category fibered in groupoids. As $\hat{G}_k$ and $\hat{P}_\pm$ are affine over k, the torsors I and $I_\pm$ are affine over S, and so the data in a $\hat{G}$-zip satisfy effective descent with respect to any fpqc morphism $S' \to S$. □

Remark 3.3. For any finite field extension k'/k the given data $\chi, \Theta, L, P_\pm, \ldots$ induces corresponding data $\chi_{k'}, \Theta_{k'}, \ldots$ over k' by base change. The definition of $\hat{G}$-zips then immediately implies that $\hat{G}\text{-}\mathtt{Zip}_{k'}^{\chi_{k'},\Theta_{k'}}$ is just the pullback of $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ under $\operatorname{Spec} k' \to \operatorname{Spec} k$. One can use this to deduce properties of $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ from the corresponding properties of $\hat{G}\text{-}\mathtt{Zip}_{k'}^{\chi_{k'},\Theta_{k'}}$. In particular, one can apply this to a finite extension k'/k for which $G_{k'}$ splits and $\pi_0(\hat{G}_{k'})$ is a constant group scheme. Thus over k' the added complexity induced by the Galois action in Sections 3E and 3F disappears.

3D. ***Realization as a quotient stack.***

Construction 3.4. Let S be a scheme over k. To any section $g \in \hat{G}(S)$ we associate a $\hat{G}$-zip of type (χ, Θ) over S, as follows. Let $I_g := S \times_k \hat{G}_k$ and $I_{g,+} := S \times_k \hat{P}_+ \subset I_g$ be the trivial torsors. Then $I_g^{(q)} \cong S \times_k \hat{G}_k = I_g$ canonically, and we define $I_{g,-} \subset I_g$ as the image of $S \times_k \hat{P}_-^{(q)} \subset S \times_k \hat{G}_-$ under left multiplication by g. Then left multiplication by g induces an isomorphism of $\hat{L}^{(q)}$-torsors

$$\begin{aligned}\iota_g \colon I_{g,+}^{(q)}/U_+^{(q)} = S \times_k \hat{P}_+^{(q)}/U_+^{(q)} \cong S \times_k \hat{P}_-^{(q)}/U_-^{(q)} \\ \xrightarrow{\sim} g(S \times_k \hat{P}_-^{(q)})/U_-^{(q)} = I_{g,-}/U_-^{(q)}.\end{aligned}$$

We thus obtain a $\hat{G}$-zip of type (χ, Θ) over S, which we denote by

$$\underline{I}_g := (I_g, I_{g,+}, I_{g,-}, \iota_g).$$

Lemma 3.5. *Every $\hat{G}$-zip of type (χ, Θ) is étale locally isomorphic to one of the form $\underline{I}_g$.*

Proof. Let $\underline{I} = (I, I_+, I_-, \iota)$ be a $\hat{G}$-zip of type (χ, Θ) over S. By Section 3A, after replacing S by an étale covering there exist sections $i_\pm \in I_\pm(S)$. These sections induce two sections $i_- U_-^{(q)}$ and $\iota(i_+^{(q)} U_+^{(q)})$ in $(I_-/U_-^{(q)})(S)$; hence there exists a unique section $\ell \in \hat{L}^{(q)}(S)$ such that $i_- U_-^{(q)} \cdot \ell = \iota(i_+^{(q)} U_+^{(q)})$. After replacing i_- by $i_- \ell$ we may therefore assume that the induced sections of $I_-/U_-^{(q)}$ coincide. Then i_- and i_+ induce two sections of I; hence there exists a unique $g \in \hat{G}(S)$ such that $i_- = i_+ g$.

We claim that $\underline{I} \cong \underline{I}_g$. Indeed, using i_+ to trivialize I_+ and I, we may without loss of generality assume that $I_+ = I_{g,+} \subset I = I_g$ and that i_+ is the identity section. Then $i_- = i_+ g$ corresponds to the section g of I_g. This implies that $I_- = i_- \hat{P}_-^{(q)} = I_{g,-}$. Furthermore, since the $\hat{L}^{(q)}$-equivariant isomorphism $\iota \colon I_+^{(q)}/U_+^{(q)} \xrightarrow{\sim} I_-/U_-^{(q)}$ sends the section $i_+^{(q)} U_+^{(q)} = U_+^{(q)}$ to the section $i_- U_-^{(q)} = g(S \times_k U_-^{(q)})$, it must coincide with ι_g. Thus we find that $\underline{I} = \underline{I}_g$ and are done. $\square$

Definition 3.6. The *algebraic zip datum* associated to $\hat{G}$ and (χ, Θ) is the tuple $\mathfrak{Z}_{\hat{G},\chi,\Theta} := (\hat{G}_k, \hat{P}_+, \hat{P}_-^{(q)}, \hat{\varphi})$ where $\hat{\varphi}$ is the composite isogeny

$$\hat{P}_+/U_+ \cong \hat{L} \xrightarrow{\mathrm{Frob}_q} \hat{L}^{(q)} \cong \hat{P}_-^{(q)}/U_-^{(q)}.$$

The associated *zip group* is the linear algebraic group over k

$$E_{\hat{G},\chi,\Theta} := \{(\ell u_+, \ell^{(q)} u_-) : \ell \in \hat{L}, u_+ \in U_+, u_- \in U_-^{(q)}\} \subset \hat{P}_+ \times_k \hat{P}_-^{(q)}. \tag{3.7}$$

It acts from the left hand side on $\hat{G}_k$ by the formula

$$(p_+, p_-) \cdot g := p_+ g p_-^{-1}. \tag{3.8}$$

If $\hat{G}$ is connected and thus $\Theta = 1$, we abbreviate $\mathscr{Z}_{\hat{G},\chi} := \mathscr{Z}_{\hat{G},\chi,\Theta}$ and $E_{\hat{G},\chi} := E_{\hat{G},\chi,\Theta}$.

Remark 3.9. In [Pink et al. 2011, Definition 10.1], we defined algebraic zip data over algebraically closed fields, whereas here k is finite. But the natural base extension of the above tuple $\mathscr{Z}_{\hat{G},\chi,\Theta}$ to an algebraic closure $\bar{k}$ of k is an algebraic zip datum in the sense of loc. cit., and the base extension of the above zip group $E_{\hat{G},\chi,\Theta}$ and its action on $\hat{G}$ are those of [loc. cit.], so all the results there have direct consequences here. For example, by [Pink et al. 2011, Proposition 7.3], the zip datum over $\bar{k}$ is orbitally finite, and so the group $E_{\hat{G},\chi,\Theta}$ acts with only finitely many orbits on $\hat{G}_{\bar{k}}$.

Lemma 3.10. *For any two sections $g, g' \in \hat{G}(S)$ there is a natural bijection between the transporter*

$$\mathrm{Transp}_{E_{\hat{G},\chi,\Theta}(S)}(g, g') := \{(p_+, p_-) \in E_{\hat{G},\chi,\Theta}(S) \mid p_+ g p_-^{-1} = g'\}$$

and the set of morphisms of $\hat{G}$-zips $\underline{I}_g \to \underline{I}_{g'}$ over S, under which (p_+, p_-) corresponds to the morphisms $I_g \to I_{g'}$ and $I_{g,+} \to I_{g',+}$ given by left multiplication with p_+ and the morphism $I_{g,-} \to I_{g',-}$ given by left multiplication with $g' p_- g^{-1}$.

Proof. By definition a morphism $\underline{I}_g \to \underline{I}_{g'}$ consists of equivariant isomorphisms $f\colon I_g \to I_{g'}$ and $f_\pm\colon I_{g,\pm} \to I_{g',\pm}$ satisfying certain compatibilities, which we analyze in turn. First, since $I_{g,+} = S \times_k \hat{P}_+ = I_{g',+}$, the isomorphism f_+ must be left multiplication by a unique section $p_+ \in \hat{P}_+(S)$. Next, since $I_g = S \times_k \hat{G} = I_{g'}$, the compatibility with f_+ implies that f, too, is left multiplication by p_+.

On the other hand, since $I_{g,-} = g(S \times_k \hat{P}_-^{(q)})$ and $I_{g',-} = g'(S \times_k \hat{P}_-^{(q)})$ within $S \times_k \hat{G}_-$, the isomorphism f_- must be left multiplication by $g' p_- g^{-1}$ for a unique section $p_- \in \hat{P}_-^{(q)}(S)$. This isomorphism must be compatible with the isomorphism $f\colon I_g \xrightarrow{\sim} I_{g'}$, which is left multiplication by p_+. The compatibility thus amounts to the equation $g' p_- g^{-1} = p_+$.

The last compatibility is the commutativity of the diagram of isomorphisms

$$\begin{tikzcd}
I_{g,+}^{(q)}/U_+^{(q)} \arrow[r, "p_+^{(q)}"] \arrow[d, "g"'] & I_{g',+}^{(q)}/U_+^{(q)} \arrow[d, "g'"'] \\
I_{g,-}/U_-^{(q)} \arrow[r, "g' p_- g^{-1}"] & I_{g',-}/U_-^{(q)},
\end{tikzcd}$$

where each arrow is defined as left multiplication by the indicated element. This amounts to the equation $p_+^{(q)} U_+^{(q)} = p_- U_-^{(q)}$ in $\hat{P}_+^{(q)}/U_+^{(q)} \cong \hat{L}^{(q)} \cong \hat{P}_-^{(q)}/U_-^{(q)}$. That in turn is equivalent to $p_+ = \ell u_+$ and $p_- = \ell^{(q)} u_-$ with $\ell \in \hat{L}(S)$, $u_+ \in U_+(S)$, and $u_- \in U_-^{(q)}(S)$, or in other words to $(p_+, p_-) \in E_{\hat{G},\chi,\Theta}(S)$.

Combined with the earlier relation $g' = p_+ g p_-^{-1}$ this means that (p_+, p_-) lies in the transporter $\mathrm{Transp}_{E_{\hat{G},\chi,\Theta}(S)}(g, g')$. Thus the map in the lemma defines a bijection between this transporter and the set of morphisms $\underline{I}_g \to \underline{I}_{g'}$, as desired. □

Proposition 3.11. *The stack $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ of $\hat{G}$-zips of type (χ, Θ) is isomorphic to the algebraic quotient stack $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$. In particular, the isomorphism classes of $\hat{G}$-zips of type (χ, Θ) over any algebraically closed field K containing k are in bijection with the $E_{\hat{G},\chi,\Theta}(K)$-orbits on $\hat{G}(K)$.*

Proof. Consider the category $\mathscr{X}$ fibered in groupoids over $((\boldsymbol{Sch}/k))$ defined as follows: For any scheme S over k the class of objects of $\mathscr{X}(S)$ is the set $\hat{G}(S)$, and for any elements $g, g' \in \hat{G}(S)$ the set of morphisms from g to g' is the transporter $\mathrm{Transp}_{E_{\hat{G},\chi,\Theta}}(g, g')$, with composition given by the multiplication in $E_{\hat{G},\chi,\Theta}$. For any morphism $S' \to S$ of schemes over k, the pullback of objects and morphisms is given by the canonical maps $\hat{G}(S) \to \hat{G}(S')$ and $E_{\hat{G},\chi,\Theta}(S) \to E_{\hat{G},\chi,\Theta}(S')$. Since $E_{\hat{G},\chi,\Theta}$ is a scheme, this is a prestack, that is, it satisfies effective descent for morphisms. By [Laumon and Moret-Bailly 2000, 3.4.3], the stackification (for this notion see [Laumon and Moret-Bailly 2000, 3.2]) of this prestack is the quotient stack $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$.

As can be verified directly from its description, the bijection in Lemma 3.10 is compatible with pullback and composition and sends $1 \in \mathrm{Transp}_{E_{\hat{G},\chi,\Theta}}(g, g)$ to the identity morphism $\mathrm{id}\colon \underline{I}_g \to \underline{I}_g$ for all $g \in \hat{G}(S)$. Thus there is a fully faithful morphism $\mathscr{X} \to \hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ which sends $g \in \mathscr{X}(S) = \hat{G}(S)$ to $\underline{I}_g$ and which acts on morphisms by the bijection of Lemma 3.10. Lemma 3.5 is then equivalent to saying that this morphism induces an isomorphism from the stackification of $\mathscr{X}$ to $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$. Since the former is $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$, the proposition follows. □

Corollary 3.12. *$\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ is a smooth algebraic stack of dimension 0 over k.*

Proof. The quotient stack $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$ it is algebraic by [Laumon and Moret-Bailly 2000, Proposition 10.13.1], and the canonical morphism $\hat{G}_k \to [E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$ is a torsor over the group scheme $E_{\hat{G},\chi,\Theta}$. As $\hat{G}_k$ and $E_{\hat{G},\chi,\Theta}$ are smooth of the same dimension, this quotient stack is smooth of dimension 0 over k. The corollary thus follows from Proposition 3.11. □

3E. ***The topological space underlying $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$.*** We recall some notation and facts from [Pink et al. 2011], especially from Sections 2.2, 6, and 10.

Choose an algebraic closure $\bar{k}$ of k and let $\Gamma := \mathrm{Gal}(\bar{k}/k)$ be the corresponding Galois group of k. Let $T \subset B \subset G_{\bar{k}}$ be a maximal torus, respectively a Borel subgroup of $G_{\bar{k}}$. Consider the finite groups

$$W := \mathrm{Norm}_{G(\bar{k})}(T(\bar{k}))/T(\bar{k}), \qquad \hat{W} := \mathrm{Norm}_{\hat{G}(\bar{k})}(T(\bar{k}))/T(\bar{k}),$$
$$\Omega := \big(\mathrm{Norm}_{\hat{G}(\bar{k})}(T(\bar{k})) \cap \mathrm{Norm}_{\hat{G}(\bar{k})}(B(\bar{k}))\big)/T(\bar{k}).$$

The fact that W acts simply transitively on the set of Borel subgroups containing $T_{\bar{k}}$ implies that $\hat{W} = W \rtimes \Omega$, and the fact that $G(\bar{k})$ acts transitively on the set of all maximal tori of $G_{\bar{k}}$ implies that $\Omega \cong \hat{W}/W \cong \pi_0(\hat{G})(\bar{k})$. Also, let $S \subset W$ be the set of simple reflections associated to the pair (T, B). As this pair is unique up to conjugation by $G(\bar{k})$, and $\mathrm{Norm}_{G(\bar{k})}(T(\bar{k})) \cap \mathrm{Norm}_{G(\bar{k})}(B(\bar{k})) = T(\bar{k})$, the Coxeter system (W, S) and the groups $\hat{W}$ and Ω are, up to unique isomorphism, independent of the choice of T and B. The inner automorphism of $\hat{W}$ induced by an element $x \in \hat{W}$ will be denoted by $\mathrm{int}(x)\colon \hat{w} \mapsto {}^x\hat{w} := x\hat{w}x^{-1}$.

Recall that the length of an element $w \in W$ is the smallest number $\ell(w)$ such that w can be written as a product of $\ell(w)$ simple reflections. For any subsets $K, K' \subseteq S$, we denote by W_K the subgroup of W generated by K and by ${}^KW$ (resp. $W^{K'}$, resp. ${}^KW^{K'}$) the set of $w \in W$ that are of minimal length in the left coset W_Kw (resp. in the right coset $wW_{K'}$, resp. in the double coset $W_KwW_{K'}$). We let $w_0 \in W$ denote the unique element of maximal length in W, and $w_{0,K}$ the unique element of maximal length in W_K.

The Frobenius isogeny $\hat{\varphi}\colon \hat{G} \to \hat{G}$ relative to $\mathbb{F}_q$ induces an automorphism $\bar{\varphi}$ of $\hat{W}$ which preserves W and Ω. Its restriction to W is an automorphism of Coxeter systems $(W, S) \xrightarrow{\sim} (W, S)$. Therefore $\bar{\varphi}$ preserves the length of elements in W and in particular satisfies $\bar{\varphi}(w_0) = w_0$.

Let $I \subseteq S$ be the type of the parabolic subgroup P_+, and $J \subseteq S$ the type of $P_-^{(q)}$. The fact that P_- is opposite to P_+ implies that $J = \bar{\varphi}({}^{w_0}I) = {}^{w_0}\bar{\varphi}(I)$. We write these equations in the form $J = \bar{\varphi}({}^yI) = {}^x\bar{\varphi}(I)$, where $x \in {}^JW^{\bar{\varphi}(I)}$ is the unique element of minimal length in $W_Jw_0W_{\bar{\varphi}(I)}$ and $y := \bar{\varphi}^{-1}(x)$. Then $\hat{\psi} := \mathrm{int}(x)\circ\bar{\varphi} = \bar{\varphi}\circ\mathrm{int}(y)$ is an automorphism of $\hat{W}$ which induces an isomorphism of Coxeter systems

$$(W_I, I) \xrightarrow{\sim} (W_J, J).$$

From [Pink et al. 2011, Proposition 2.7] we can deduce that

$$y = w_0w_{0,I} = w_{0,\bar{\varphi}^{-1}(J)}w_0. \tag{3.13}$$

Via the isomorphism $\pi_0(\hat{G})(\bar{k}) \cong \Omega$ we can view $\Theta(\bar{k})$ (resp. $\Theta^{(q)}(\bar{k})$) as a subgroup of Ω, which by abuse of notation we will again denote by Θ (resp. $\Theta^{(q)}$). Note that, since $\Theta \subseteq \mathrm{Norm}_{\hat{G}}(P_+)/P_+$, conjugation by elements of Θ preserves the type I of P_+ and thus the subgroup W_I. Therefore $W_I\Theta = W_I \rtimes \Theta$ is a subgroup of $\hat{W}$. Since $\Theta^{(q)} \subseteq \mathrm{Norm}_{\hat{G}}(P_-^{(q)})/P_-^{(q)}$, the same observation holds for $W_J\Theta^{(q)} = W_J \rtimes \Theta^{(q)}$, and the automorphism $\hat{\psi}$ sends the subgroup $\Theta \subseteq \Omega$ to $\Theta^{(q)} \subseteq \Omega$. By [Pink et al. 2011, Lemma 10.4], the map

$$(\theta, \hat{w}) \mapsto \theta\hat{w}\hat{\psi}(\theta)^{-1} \tag{3.14}$$

defines a left action of Θ on the subset ${}^IW\Omega \subseteq \hat{W}$.

Recall that the Bruhat order $\leq$ on W is defined by $w' \leq w$ if for some (and equivalently for any) expression of w as a product of $\ell(w)$ simple reflections, by leaving out certain factors one can obtain an expression of w' as a product of $\ell(w')$ simple reflections. We extend the Bruhat order to $\hat{W}$ by setting

$$w'\omega' \leq w\omega \text{ if and only if } w' \leq w \text{ and } \omega' = \omega \tag{3.15}$$

for any $w, w' \in W$ and $\omega, \omega' \in \Omega$. Also, for any $\hat{w}, \hat{w}' \in {}^I W\Omega$ we write

$$\hat{w}' \preceq \hat{w} \text{ if and only if there exists } \hat{v} \in W_I\Theta \text{ with } \hat{v}\hat{w}'\hat{\psi}(\hat{v})^{-1} \leq \hat{w}. \tag{3.16}$$

By [Pink et al. 2011, Theorem 10.9], see also [He 2007], this defines a partial order on ${}^I W\Omega$. We also extend the length function from W to $\hat{W}$ by setting

$$\ell(w\omega) := \ell(w) \tag{3.17}$$

for any $w \in W$ and $\omega \in \Omega$.

Lemma 3.18. *The action (3.14) preserves the extended Bruhat order $\leq$ on $\hat{W}$, the partial order $\preceq$ on ${}^I W\Omega$, and the extended length function ℓ on $\hat{W}$.*

Proof. Consider any elements $\theta \in \Theta$ and $w', w \in W$ and $\omega', \omega \in \Omega$. First assume that $w'\omega' \leq w\omega$, in other words, that $w' \leq w$ and $\omega' = \omega$. Then $\theta w'\theta^{-1} \leq \theta w\theta^{-1}$ and $\theta\omega'\hat{\psi}(\theta)^{-1} = \theta\omega\hat{\psi}(\theta)^{-1}$, and the latter is again an element of Ω, because $\Theta^{(q)} = \hat{\psi}(\Theta) \subset \Omega$. By (3.15) we therefore find that

$$\theta w'\omega'\hat{\psi}(\theta)^{-1} = \theta w'\theta^{-1} \cdot \theta\omega'\hat{\psi}(\theta)^{-1} \leq \theta w\theta^{-1} \cdot \theta\omega\hat{\psi}(\theta)^{-1} = \theta w\omega\hat{\psi}(\theta)^{-1}.$$

Thus the action (3.14) preserves the extended Bruhat order $\leq$ on $\hat{W}$.

Next, the last equality above and the fact that $\theta\omega\hat{\psi}(\theta)^{-1} \in \Omega$ also imply that the length of $\theta w\omega\hat{\psi}(\theta)^{-1}$ is equal to that of $\theta w\theta^{-1}$. Since $\theta \in \Omega$, that length is equal to the length of w and hence of $w\omega$, proving that the action (3.14) preserves the extended length function on $\hat{W}$.

Now assume that $w, w' \in {}^I W$ and $w'\omega' \preceq w\omega$, which means that $\hat{v}w'\omega'\hat{\psi}(\hat{v})^{-1} \leq w\omega$ for some $\hat{v} \in W_I\Theta$. Then we have just shown that

$$\theta\hat{v}\theta^{-1} \cdot \theta w'\omega'\hat{\psi}(\theta)^{-1} \cdot \hat{\psi}(\theta\hat{v}\theta^{-1})^{-1} = \theta\hat{v}w'\omega'\hat{\psi}(\hat{v})^{-1}\hat{\psi}(\theta)^{-1} \leq \theta w\omega\hat{\psi}(\theta)^{-1}.$$

Since θ normalizes W_I, it follows that $\hat{u} := \theta\hat{v}\theta^{-1}$ is an element of $W_I\Theta$ which satisfies $\hat{u} \cdot \theta w'\omega'\hat{\psi}(\theta)^{-1} \cdot \hat{\psi}(\hat{u})^{-1} \leq \theta w\omega\hat{\psi}(\theta)^{-1}$ and thus by (3.16) shows that $\theta w'\omega'\hat{\psi}(\theta)^{-1} \preceq \theta w\omega\hat{\psi}(\theta)^{-1}$. Therefore the action (3.14) preserves the partial order $\preceq$, and we are done. □

As a consequence of Lemma 3.18, the partial order $\preceq$ from (3.16) induces a partial order on the set of Θ-orbits

$$\Xi^{\chi,\Theta} := \Theta\backslash{}^I W\Omega. \tag{3.19}$$

By Proposition 2.1 this in turn defines a T_0 topology on the finite set $\Xi^{\chi,\Theta}$.

Now observe that since the subgroups $P_\pm \subset G_k$ and $\hat{P}_\pm \subset \hat{G}_k$ are defined over k, there is a natural continuous action of $\Gamma := \mathrm{Gal}(\bar{k}/k)$ on everything discussed above. In particular this action preserves the decomposition $\hat{W} = W \rtimes \Omega$, the subsets $S, I, J, {}^I W, \ldots$, the partial orders $\leq$ and $\preceq$, the length function ℓ, the subgroup Θ and its action on ${}^I W\Omega$, and so it induces an action on the topological space $\Xi^{\chi,\Theta}$.

Theorem 3.20. *The topological space underlying $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ is naturally homeomorphic to the quotient space $\Gamma\backslash\Xi^{\chi,\Theta}$.*

Proof. By Proposition 3.11 the stack $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ is isomorphic to $[E_{\hat{G},\chi,\Theta}\backslash\hat{G}_k]$. By [Pink et al. 2011, Proposition 7.3], (see also Remark 3.9) the zip datum $\mathscr{Z}_{\hat{G},\chi,\Theta,\bar{k}}$ is orbitally finite, that is, the number of $E_{\hat{G},\chi,\Theta}(\bar{k})$-orbits in $\hat{G}(\bar{k})$ is finite. We can therefore apply Proposition 2.2. The description of the topological space now follows from the description of $E_{\hat{G},\chi,\Theta,\bar{k}}$-orbits in $\hat{G}_{\bar{k}}$ and their closures from [Pink et al. 2011, Theorems 10.9 and 10.10]. □

Remark 3.21. If we replace k by a suitable finite extension k' within $\bar{k}$, the Galois group Γ is replaced by a subgroup which acts trivially on $\hat{W}$ and everything else above. Then Theorem 3.20 asserts that the topological space underlying $\hat{G}\text{-}\mathtt{Zip}_{k'}^{\chi,\Theta}$ is naturally homeomorphic to $\Xi^{\chi,\Theta}$. In particular, for any algebraically closed extension field K of $\bar{k}$ we obtain a natural bijection

$$\Xi^{\chi,\Theta} \xrightarrow{\sim} \left\{\begin{matrix}\text{isomorphism classes of } \hat{G}\text{-zips} \\ \text{of type } (\chi,\Theta) \text{ over } K\end{matrix}\right\}. \tag{3.22}$$

By [Pink et al. 2011] this can be made more explicit, as follows. As the choice of (T, B) was arbitary, we may without loss of generality assume that $T \subset L_K$ and $B \subset P_{-,K}$. Then we may identify $W = \mathrm{Norm}_G(T)(K)/T(K)$ and $\hat{W} = \mathrm{Norm}_{\hat{G}}(T)(K)/T(K)$. Choose a representative $g \in \mathrm{Norm}_G(T)(K)$ of the element $y = \bar{\varphi}^{-1}(x) \in W$. Then by [Pink et al. 2011, Lemma 12.11] the triple (B, T, g) is a frame of the connected zip datum $(G_K, P_{+,K}, P_{-,K}^{(q)}, \varphi\colon L_K \to L_K^{(q)})$ in the sense of [Pink et al. 2011, Definition 3.6]. Also, for every element $\hat{w} \in {}^I W\Omega$ choose a representative $\dot{\hat{w}} \in \mathrm{Norm}_{\hat{G}}(T)(K)$, and let $\underline{I}_{g\dot{\hat{w}}}$ denote the $\hat{G}$-zip of type (χ,Θ) over K attached to $g\dot{\hat{w}} \in \hat{G}(K)$ by Construction 3.4. Combining the isomorphism in Proposition 3.11 with [Pink et al. 2011, Theorem 10.10] then shows that the bijection (3.22) sends the orbit of $\hat{w}$ in $\Xi^{\chi,\Theta} = \Theta\backslash{}^I W\Omega$ to the isomorphism class of $\underline{I}_{g\dot{\hat{w}}}$.

Example 3.23. Assume that $\hat{G} = G$ is a connected split reductive group over $\mathbb{F}_q$. Then $\Theta = \Omega = 1$, and $\Xi^{\chi,\Theta} = {}^I W$ with the trivial action of $\mathrm{Gal}(\bar{k}/\mathbb{F}_q)$. All the formulas then simplify accordingly. In particular, by Theorem 3.20 the topological space underlying $G\text{-}\mathtt{Zip}_k^\chi$ is naturally homeomorphic to ${}^IW$ for every finite extension k of $\mathbb{F}_q$.

Moreover, the automorphism $\bar\varphi$ induced by Frobenius is the identity on W. Thus $\hat\psi = \mathrm{int}(x)$, where $x \in {}^JW^I$ is the unique element of minimal length in $W_J w_0 W_I$. The partial order $\preceq$ on ${}^IW$ is therefore given by

$$w' \preceq w \text{ if and only if there exists } v \in W_I \text{ with } vw'xv^{-1}x^{-1} \leq w. \tag{3.24}$$

If in addition the Dynkin diagram of G has no component of type A_n with $n \geq 2$, of type D_n with $n \geq 5$ odd, or of type E_6, then w_0 is central in W, and so $I = J$. Also, by (3.13) we then have $x = w_0 w_{0,I}$, and since w_0 is central and $w_{0,I}^{-1} = w_{0,I}$, the partial order can then be written equivalently in the form

$$w' \preceq w \text{ if and only if there exists } v \in W_I \text{ with } vw'w_{0,I}v^{-1}w_{0,I} \leq w. \tag{3.25}$$

3F. *The stratification of* $\hat G$-$\mathtt{Zip}_k^{\chi,\Theta}$. For any orbit in $\Gamma\backslash\Xi^{\chi,\Theta} = \Gamma\backslash(\Theta\backslash{}^IW\Omega)$ represented by an element $\hat w \in {}^IW\Omega$, let $[\hat w]$ denote the corresponding point in the topological space underlying $\hat G\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ via the homeomorphism in Theorem 3.20.

Theorem 3.26. *The point* $[\hat w]$ *underlies a smooth locally closed substack of the category* $\hat G\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ *of pure codimension* $\dim(G/P_+) - \ell(\hat w)$, *where* $\ell(\,)$ *denotes the extended length function from* (3.17).

Proof. As in the proof of Theorem 3.20 this translates into an assertion for the quotient stack $[E_{\hat G,\chi,\Theta}\backslash\hat G_k]$. Let $\hat G_{\bar k}^{\hat w}$ denote the $E_{\hat G,\chi,\Theta,\bar k}$-orbit in $\hat G_{\bar k}$ corresponding to $\hat w$ by [Pink et al. 2011, Theorem 10.10]. Since k is perfect, by the remarks following Proposition 2.2 this determines a smooth locally closed substack of $[E_{\hat G,\chi,\Theta}\backslash\hat G_k]$ with underlying point $[\hat w]$. By (2.4) the codimension of this substack is equal to the codimension of $\hat G_{\bar k}^{\hat w}$ in $\hat G_{\bar k}$, which by [Pink et al. 2011, Theorem 5.11 and Lemma 10.3] is given by the desired formula. $\square$

Let S be a scheme over k, and let $\underline{I}$ be a $\hat G$-zip of type (χ,Θ) over S. Then $\underline{I}$ defines a classifying morphism

$$\zeta : S \longrightarrow \hat G\text{-}\mathtt{Zip}_k^{\chi,\Theta}. \tag{3.27}$$

Let $S_{\underline{I}}^{\hat w}$ denote the pullback under ζ of the substack corresponding to $[\hat w]$. This is a locally closed subscheme of S. As $[\hat w]$ varies, these subschemes form a finite stratification of S, in other words S is the set-theoretic disjoint union

$$S = \bigsqcup_{[\hat w]\in\Gamma\backslash\Xi^{\chi,\Theta}} S_{\underline{I}}^{[\hat w]}. \tag{3.28}$$

The description of the topology in Theorem 3.20 implies that for any $\hat{w}$ we have

$$\overline{S_{\underline{I}}^{[\hat{w}]}} \subset \bigsqcup_{\substack{[\hat{w}'] \in \Gamma \backslash \Xi^{\chi,\Theta} \\ \hat{w}' \preceq \hat{w}}} S_{\underline{I}}^{[\hat{w}]} \tag{3.29}$$

For the next result recall that any open or flat morphism of schemes is generizing.

Proposition 3.30. *If the morphism ζ in (3.27) is generizing, the inclusion (3.29) is an equality. If in addition S is locally noetherian, then $S_{\underline{I}}^{[\hat{w}]}$ is of pure of codimension* $\dim(G/P_+) - \ell(\hat{w})$. *If ζ is smooth, then $S_{\underline{I}}^{[\hat{w}]}$ is smooth as a scheme over k.*

Proof. If ζ is generizing, then $\zeta^{-1}(\overline{\Upsilon}) = \overline{\zeta^{-1}(\Upsilon)}$ for any locally closed substack Υ of $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$, so the first assertion follows from Theorem 3.20. If in addition S is locally noetherian, the codimension is well defined and preserved by ζ; so the second assertion follows from Theorem 3.26. If ζ is smooth, then $S_{\underline{I}}^{[\hat{w}]}$, being the pullback of a smooth stack under a smooth morphism, is smooth as a scheme over k, proving the third assertion. $\square$

Instead of the above construction, the subscheme $S_{\underline{I}}^{[\hat{w}]}$ can also be characterized by a construction directly involving $\underline{I}$. For simplicity we discuss this only in a special case (but compare Remark 3.3):

Proposition 3.31. *Assume that G splits over k and that $\pi_0(\hat{G}_k)$ is a constant group scheme. Then a morphism of schemes $f : S' \to S$ factors through $S_{\underline{I}}^{[\hat{w}]}$ if and only if $f^*\underline{I}$ is locally for the fppf-topology on S' isomorphic to the constant G-zip $\underline{I}_{g\dot{\hat{w}}} \times_k S$ with $\underline{I}_{g\dot{\hat{w}}}$ as in Remark 3.21.*

Proof. The assumptions imply that the $E_{\hat{G},\chi,\Theta}$-orbit used to prove Theorem 3.26 is really defined over k; let us denote it by $\hat{G}_k^{\hat{w}}$. Define S'' and g' by the cartesian diagram

$$\begin{tikzcd} S'' \arrow[rr, "g'"] \arrow[d] & & \hat{G}_k \arrow[d] \\ S' \arrow[r, "f"] & S \arrow[r, "\zeta"] & {[E_{\hat{G},\chi,\Theta} \backslash \hat{G}_k] \cong \hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}} \end{tikzcd}$$

where the vertical morphisms are fppf. Then by the definition of the quotient stack and the construction of $S_{\underline{I}}^{[\hat{w}]}$, the morphism f factors through $S_{\underline{I}}^{[\hat{w}]}$ if and only if g' factors through $\hat{G}_k^{\hat{w}}$. As the orbit $\hat{G}_k^{\hat{w}}$ is smooth, the morphism $E_{\hat{G},\chi,\Theta} \to \hat{G}_k^{\hat{w}}$, $e \mapsto e \cdot g\dot{\hat{w}}$ is fppf. Thus g' factors through $\hat{G}_k^{\hat{w}}$ if and only if there exists an fppf-covering $S''' \to S''$ and an $e : S''' \to E_{\hat{G},\chi,\Theta}$ such that $g' = e \cdot g\dot{\hat{w}}$. By Lemma 3.10 the latter condition is equivalent to saying that the $\hat{G}$-zip $\underline{I}_{g'}$ is fppf-locally isomorphic to $\underline{I}_{g\dot{\hat{w}}}$, or again that $f^*\underline{I}$ is fppf-locally isomorphic to $\underline{I}_{g\dot{\hat{w}}}$, as desired. $\square$

Remark 3.32. It is shown in [Wedhorn and Yatsyshyn 2014] that all $E_{\hat{G},\chi,\Theta}$-orbits in $\hat{G}_k$ are affine. This implies that the inclusion into $\hat{G}\text{-}\mathtt{Zip}_k^{\chi,\Theta}$ of the substack associated to $[\hat{w}]$ is an affine morphism, and so the inclusion $S_{\underline{I}}^{[\hat{w}]} \hookrightarrow S$ is an affine morphism. In particular this implies the following purity result:

Proposition 3.33. *Let S be a locally noetherian scheme over k, and let Z be a closed subscheme of codimension ≥ 2. Assume that Z contains no embedded component of S (which is automatic if S is reduced). Let $\underline{I}$ be a $\hat{G}$-zip over S whose restriction to $S \smallsetminus Z$ is fppf-locally constant. Then $\underline{I}$ is fppf-locally constant.*

Proof. By Proposition 3.31 there exists $[\hat{w}]$ such that the open immersion $S \smallsetminus Z \hookrightarrow S$ factors through the subscheme $S_{\underline{I}}^{[\hat{w}]}$. By assumption $S \smallsetminus Z$ and hence $S_{\underline{I}}^{[\hat{w}]}$ is schematically dense in S; being locally closed $S_{\underline{I}}^{[\hat{w}]}$ is therefore an open subscheme of S. On the other hand its complement Z' is of codimension ≥ 2. Since the inclusion $S_{\underline{I}}^{[\hat{w}]} \hookrightarrow S$ is affine, this implies that $Z' = \varnothing$. $\square$

3G. *Automorphisms of G-zips.* Let K be an algebraically closed extension field of $\bar{k}$, and let $\underline{I}$ be a $\hat{G}$-zip of type (χ, Θ) over K. Let T, B, g, $\dot{\hat{w}}$ be as in Remark 3.21. Then $\underline{I}$ is isomorphic to $\underline{I}_{g\dot{\hat{w}}}$ for some $\hat{w} \in {}^IW\Omega$. Its automorphism group scheme is therefore $\underline{\mathrm{Aut}}(\underline{I}) \cong \underline{\mathrm{Aut}}(\underline{I}_{g\dot{\hat{w}}})$. By Proposition 2.5 the latter is isomorphic to the stabilizer $\mathrm{Stab}_{E_{\hat{G},\chi,\Theta,K}}(g\dot{\hat{w}})$.

Since the results on stabilizers in [Pink et al. 2011] were formulated only for connected zip data, we now assume that $\hat{G} = G$ is connected. Then $\Theta = \Omega = 1$, and we can write $\hat{w} = w \in {}^IW$ and $\dot{\hat{w}} = \dot{w} \in \mathrm{Norm}_G(T)(K)$. As in [Pink et al. 2011, Section 5.1], let H_w be the Levi subgroup of G_K containing T whose set of simple reflections is the unique largest subset K_w of $J \cap {}^{w^{-1}}I$ such that $(\mathrm{int}(x) \circ \bar{\varphi} \circ \mathrm{int}(w))(K_w) = K_w$.

Proposition 3.34. (a) *The identity component of $\underline{\mathrm{Aut}}(\underline{I}_{g\dot{w}})$ is a unipotent group scheme of dimension* $\dim(G/P_+) - \ell(w)$.

(b) *Let v be the unique element of minimal length in the double coset $W_I w W_J$. Then the Lie algebra of $\underline{\mathrm{Aut}}(\underline{I}_{g\dot{w}})$ has dimension* $\dim(G/P_+) - \ell(v)$.

(c) *The group of connected components of $\underline{\mathrm{Aut}}(\underline{I}_{g\dot{w}})$ is isomorphic to the constant group scheme over K associated to the finite group*

$$\Pi := \{h \in H_w(\bar{k}) : h = \varphi(g\dot{w}h(g\dot{w})^{-1})\}.$$

Proof. The group $A := \underline{\mathrm{Aut}}(\underline{I}_{g\dot{w}}) \cong \mathrm{Stab}_{E_{G,\chi,\Theta,K}}(g\dot{w})$ is isomorphic to a semidirect product of the group Π in (c) with a connected unipotent group scheme U [Pink et al. 2011, Theorem 8.1]. As the zip datum is orbitally finite, the group Π is finite by [Pink et al. 2011, Proposition 7.1]. This shows (c) and that the identity component of A is unipotent. Moreover, the orbit $o(g\dot{w}) \subset G_K$ of $g\dot{w}$ has dimension

$\dim P_+ + \ell(w)$ by [Pink et al. 2011, Theorem 7.5]. As the definition of $E_{G,\chi,\Theta}$ implies that $\dim E_{G,\chi,\Theta} = \dim G$, it follows that

$$\dim A = \dim E_{G,\chi,\Theta,K} - \dim o(g\dot{w}) = \dim G - \dim(P_+) - \ell(w),$$

proving (a). Assertion (b) follows from [Pink et al. 2011, Theorem 8.5]. □

Remark 3.35. Since a group scheme is smooth if and only if its dimension is equal to the dimension of its Lie algebra, Proposition 3.34 (a) and (b) imply that $\underline{\mathrm{Aut}}(\underline{I}_{g\dot{w}})$ is smooth if and only if w is of minimal length in its double coset $W_I w W_J$. This condition will often not be satisfied.

4. Generalities on filtrations

In this section we briefly review some standard definitions and notations for gradings and filtrations of locally free sheaves of finite rank.

4A. ***Locally free sheaves of finite rank.*** Let S be a scheme over a ring k. The category of locally free sheaves of $\mathcal{O}_S$-modules of finite rank on S with all $\mathcal{O}_S$-linear homomorphisms between them is denoted $\mathtt{LF}(S)$. It is a k-linear additive category, but in general not abelian. A homomorphism in $\mathtt{LF}(S)$ is called *admissible* if its image in the category of sheaves of finite rank is a locally direct summand. This notion turns $\mathtt{LF}(S)$ into an exact category in the sense of Quillen. It is also idempotent complete, that is, any endomorphism $f : \mathcal{M} \to \mathcal{M}$ in $\mathtt{LF}(S)$ satisfying $f^2 = f$ is admissible and corresponds to a direct sum decomposition $\mathcal{M} = \ker(f) \oplus \mathrm{im}(f)$ within $\mathtt{LF}(S)$.

For a useful overview of exact categories see [Bühler 2010]. Every admissible homomorphism in an exact category has a kernel and a cokernel, and they satisfy a number of axioms: see [Bühler 2010]. An additive functor between exact categories is exact if it sends admissible homomorphisms to admissible homomorphisms and commutes with their kernels and cokernels.

Endowed with the usual tensor product of sheaves of finite rank $\mathcal{M} \otimes \mathcal{N}$, the usual dual $\mathcal{M}^\vee := \mathcal{H}om(\mathcal{M}, \mathcal{O}_S)$, and the usual associativity and commutativity constraints $\mathtt{LF}(S)$ is a rigid tensor category in the sense of [Saavedra Rivano 1972, I.5.1]. Moreover, the tensor product and the dual define exact functors in the indicated sense.

4B. ***Gradings.*** By a *graded locally free sheaf of finite rank on S* we mean a locally free sheaf of finite rank $\mathcal{M}$ together with a decomposition $\mathcal{M} = \bigoplus_{i \in \mathbb{Z}} \mathcal{M}^i$, whose *graded pieces* $\mathcal{M}^i$ vanish for almost all i. A *homomorphism* of graded locally free sheaves of finite rank $f : \mathcal{M} \to \mathcal{N}$ is a homomorphism of the underlying sheaves that satisfies $f(\mathcal{M}^i) \subset \mathcal{N}^i$ for all $i \in \mathbb{Z}$. The category of graded locally free sheaves of finite rank on S is denoted $\mathtt{GrLF}(S)$. It is a k-linear additive category, but in general not abelian.

A homomorphism in $\mathtt{GrLF}(S)$ is *admissible* if it is admissible in each degree. This turns $\mathtt{GrLF}(S)$ into an exact category that is idempotent complete.

The *tensor product* of graded locally free sheaves of finite rank is the usual tensor product of sheaves with the grading $(\mathcal{M}\otimes\mathcal{N})^i := \bigoplus_{j\in\mathbb{Z}} \mathcal{M}^j\otimes\mathcal{N}^{i-j}$. The *dual* of a graded locally free sheaf of finite rank $\mathcal{M}$ is the usual dual sheaf with the grading $(\mathcal{M}^\vee)^i := (\mathcal{M}^{-i})^\vee$. These notions turn $\mathtt{GrLF}(S)$ into a rigid tensor category together with a (forgetful) exact tensor functor $\mathrm{forg}\colon \mathtt{GrLF}(S)\to\mathtt{LF}(S)$ sending graded locally free sheaves to their underlying locally free sheaves.

4C. ***Descending filtrations.*** By a *descending filtration* $C^\bullet$ of a locally free sheaf of finite rank $\mathcal{M}$ on S we mean a family of quasicoherent subsheaves $C^i\mathcal{M}$ for $i\in\mathbb{Z}$, which are locally direct summands and satisfy $C^{i+1}\mathcal{M}\subset C^i\mathcal{M}$ for all i and $C^i\mathcal{M}=0$ for all $i\gg 0$ and $C^i\mathcal{M}=\mathcal{M}$ for all $i\ll 0$. A homomorphism of sheaves of finite rank $f\colon\mathcal{M}\to\mathcal{N}$ endowed with a descending filtration is *compatible with the filtrations* if it satisfies $f(C^i\mathcal{M})\subset C^i\mathcal{N}$ for all $i\in\mathbb{Z}$. This defines a category of locally free sheaves of finite rank on S endowed with a descending filtration, which we denote $\mathtt{FilLF}^\bullet(S)$. It is a k-linear additive category, but in general not abelian. It possesses an evident forgetful functor $\mathrm{forg}\colon \mathtt{FilLF}^\bullet(S)\to\mathtt{LF}(S)$.

The assumptions imply that the subquotients $\mathrm{gr}^i_C\mathcal{M} := C^i\mathcal{M}/C^{i+1}\mathcal{M}$ are again locally free sheaves of finite rank on S which vanish for almost all i. Also, any homomorphism $f\colon\mathcal{M}\to\mathcal{N}$ in $\mathtt{FilLF}^\bullet(S)$ induces natural homomorphisms $\mathrm{gr}^i_C f\colon \mathrm{gr}^i_C\mathcal{M}\to\mathrm{gr}^i_C\mathcal{N}$. Together this defines a natural functor $\mathrm{gr}^\bullet_C\colon \mathtt{FilLF}^\bullet(S)\to \mathtt{GrLF}(S)$.

Reciprocally, any graded locally free sheaf of finite rank $\mathcal{M}$ carries a natural descending filtration $C^i\mathcal{M} := \bigoplus_{j\geqslant i}\mathcal{M}^j$, which defines a natural functor $\mathrm{fil}^\bullet\colon \mathtt{GrLF}(S)\to\mathtt{FilLF}^\bullet(S)$.

A homomorphism $f\colon\mathcal{M}\to\mathcal{N}$ in $\mathtt{FilLF}^\bullet(S)$ is called *admissible* if for all i the sheaf $f(C^i\mathcal{M})$ is equal to $f(\mathcal{M})\cap C^i\mathcal{N}$ and a locally direct summand of $\mathcal{N}$. This is equivalent to saying that locally on S, the morphism possesses a factorization of the form $\mathcal{M}\cong\mathcal{M}'\oplus\mathcal{L}\twoheadrightarrow\mathcal{L}\hookrightarrow\mathcal{L}\oplus\mathcal{N}'\cong\mathcal{N}$ in the category of filtered locally free sheaves of finite rank. With this notion $\mathtt{FilLF}^\bullet(S)$ is an exact category that is idempotent complete.

Descending filtrations of $\mathcal{M}$ and $\mathcal{N}$ induce a natural descending filtration of $\mathcal{M}\otimes\mathcal{N}$ by the formula $C^i(\mathcal{M}\otimes\mathcal{N}) := \sum_{j\in\mathbb{Z}} C^j\mathcal{M}\otimes C^{i-j}\mathcal{N}$. The graded subquotients inherit natural isomorphisms $\mathrm{gr}^i_C(\mathcal{M}\otimes\mathcal{N})\cong\bigoplus_{j\in\mathbb{Z}}\mathrm{gr}^j_C\mathcal{M}\otimes\mathrm{gr}^{i-j}_C\mathcal{N}$. Moreover, a descending filtration of $\mathcal{M}$ induces a descending filtration of $\mathcal{M}^\vee$ by the formula $C^i(\mathcal{M}^\vee) := (\mathcal{M}/C^{1-i}\mathcal{M})^\vee$, and the graded subquotients possess natural isomorphisms $\mathrm{gr}^i_C(\mathcal{M}^\vee)\cong(\mathrm{gr}^{-i}_C\mathcal{M})^\vee$. These notions turn $\mathtt{FilLF}^\bullet(S)$ into a rigid tensor category, such that all three functors above are tensor functors. These functors, as well as tensor product and dual, are also exact.

4D. ***Ascending filtrations.*** An *ascending filtration* $D_\bullet$ of $\mathcal{M}$ is a family of subsheaves $D_i\mathcal{M}$ such that the $D_{-i}\mathcal{M}$ form a descending filtration of $\mathcal{M}$. Thus everything in Section 4C has a direct analogue for ascending filtrations. Descending filtrations are generally indexed by upper indices, ascending filtrations by lower indices, while gradings can be indexed in both fashions. In particular the graded subquotients of an ascending filtration are denoted $\mathrm{gr}_i^D\mathcal{M} := D_i\mathcal{M}/D_{i-1}\mathcal{M}$. The category of locally free sheaves of $\mathcal{O}_S$-modules with an ascending filtration is denoted $\mathtt{FilLF}_\bullet(S)$. There are natural exact tensor functors $\mathrm{gr}_\bullet^D\colon \mathtt{FilLF}_\bullet(S) \to \mathtt{GrLF}(S)$ and $\mathrm{fil}_\bullet\colon \mathtt{GrLF}(S) \to \mathtt{FilLF}_\bullet(S)$ and $\mathrm{forg}\colon \mathtt{FilLF}_\bullet(S) \to \mathtt{LF}(S)$. The functors introduced so far are summarized in the following diagram.

$$
\begin{tikzcd}
 & \mathtt{FilLF}^\bullet(S) \arrow[dl, shift right, "\mathrm{gr}_C^\bullet"'] \arrow[dr, "\mathrm{forg}"] & \\
\mathtt{GrLF}(S) \arrow[ur, shift right, "\mathrm{fil}^\bullet"'] \arrow[rr, "\mathrm{forg}"] \arrow[dr, shift left, "\mathrm{fil}_\bullet"] & & \mathtt{LF}(S) \\
 & \mathtt{FilLF}_\bullet(S) \arrow[ul, shift left, "\mathrm{gr}_\bullet^D"] \arrow[ur, "\mathrm{forg}"'] &
\end{tikzcd}
\tag{4.1}
$$

4E. ***Types.*** Let $\underline{n} = (n_i)_{i\in\mathbb{Z}}$ be a family of nonnegative integers which vanish for almost all i. We say that a graded locally free sheaf of finite rank $\mathcal{M}$ is *of type* $\underline{n}$ if each $\mathcal{M}^i$ is locally free of constant rank n_i. We call a locally free sheaf of finite rank endowed with a descending or ascending filtration of type $\underline{n}$ if its associated graded sheaf is of type $\underline{n}$. In all these cases, the sheaf itself is then locally free of constant rank $\sum_i n_i$. If S is connected (and hence nonempty!), every graded or filtered locally free sheaf of finite rank on S possesses a unique type.

4F. ***Pullback.*** All the above notions possess evident pullbacks under a morphism $S' \to S$, which are compatible with all the given constructions. We generally denote the pullback of $f:\mathcal{M}\to\mathcal{N}$ by $f_{S'}:\mathcal{M}_{S'}\to\mathcal{N}_{S'}$. This defines an exact tensor functor $\mathtt{FilLF}^\bullet(S) \to \mathtt{FilLF}^\bullet(S')$ and similar functors on the other categories.

Many properties and invariants such as the rank of a locally free sheaf of finite rank are local for the fpqc topology. In particular:

Lemma 4.2. *For a homomorphism of graded, filtered, or naked locally free sheaves of finite rank, the property of being admissible is local for the fpqc topology.*

Proof. The subsheaf $f(\mathcal{M}) \subset \mathcal{N}$ is a locally direct summand if and only if the quotient $\mathcal{N}/f(\mathcal{M})$ is locally free. Since the latter property is local for the fpqc topology, so is the former, and the lemma follows for naked and graded locally free sheaves of finite rank. For filtered ones observe that the formation of $f(C^i\mathcal{M})$ and $f(\mathcal{M})\cap C^i\mathcal{N}$ commutes with flat pullback and their equality is local for the fpqc topology. By the same argument as before their being a locally direct summand is local for the fpqc topology, too, and so the lemma follows in the filtered case. ☐

4G. ***Alternating and symmetric powers.*** Consider an object X of an exact additive tensor category $\mathscr{C}$. For any integer $m \geqslant 0$ let $X^{\otimes m}$ denote the tensor product of m copies of X with itself, which carries a natural action of the symmetric group S_m. Thus there is a homomorphism

$$A_m(X)\colon X^{\otimes m} \longrightarrow X^{\otimes m}, \quad x \mapsto \sum_{\sigma\in S_m} \operatorname{sgn}(\sigma)\cdot\sigma(x). \tag{4.3}$$

If this homomorphism is admissible, its image $\Lambda^m X := \operatorname{im} A_m(X)$ is called the *m-th alternating, or exterior, power of X*. Likewise, there is a homomorphism

$$B_m(X)\colon \bigoplus_{\sigma\in S_m} X^{\otimes m} \longrightarrow X^{\otimes m}, \quad (x_\sigma)_\sigma \mapsto \sum_{\sigma\in S_m} (\sigma-1)(x_\sigma). \tag{4.4}$$

If this homomorphism is admissible, its cokernel $\mathrm{S}^m X := \operatorname{coker} B_m(X)$ is called the *m-th symmetric power of X*.

For any morphism $f\colon X \to Y$ in $\mathscr{C}$ the above constructions are compatible with the induced morphism $f^{\otimes m}\colon X^{\otimes m} \to Y^{\otimes m}$; hence they induce morphisms $\Lambda^m f\colon \Lambda^m X \to \Lambda^m Y$ and $\mathrm{S}^m f\colon \mathrm{S}^m X \to \mathrm{S}^m Y$ whenever the respective powers exist. Evidently this sends the identity on X to the identity on $\Lambda^m X$ and $\mathrm{S}^m X$ and commutes with composition, that is, it is functorial in X.

As a direct consequence of this construction, alternating and symmetric powers commute with any exact tensor functor between exact additive tensor categories $F\colon \mathscr{C} \to \mathscr{D}$. More precisely, if $\Lambda^m X$ exists, then $\Lambda^m F(X)$ exists and is canonically isomorphic to $F(\Lambda^m X)$, and similarly for S^m.

In each of the categories $\texttt{LF}(S)$, $\texttt{GrLF}(S)$, $\texttt{FilLF}^\bullet(S)$, and $\texttt{FilLF}_\bullet(S)$ above, all alternating and symmetric powers exist and have the usual local descriptions, essentially because every object is Zariski locally on S a direct sum of objects of rank 1. Also, the m-th alternating power of an object of constant rank n has constant rank $\binom{n}{m}$, and the m-th symmetric power of an object of constant rank n has constant rank $\binom{n+m-1}{m}$. In particular, the n-th exterior power of an object X of constant rank n is an object of constant rank 1, also called the *highest exterior power of X*.

5. Filtered fiber functors and cocharacters

Let $\hat{G}$ be a (not necessarily connected) linear algebraic group over an arbitrary field k_0, and let $\hat{G}\text{-}\texttt{Rep}$ denote the tensor category of finite dimensional representations of $\hat{G}$ over k_0. As usual, by a *fiber functor* over a scheme S over k_0 we mean an exact k_0-linear tensor functor $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{LF}(S)$. Similarly, by a *graded fiber functor* we mean an exact k_0-linear tensor functor $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$, and by a *filtered fiber functor* an exact k_0-linear tensor functor $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$. In this section we collect some results from [Saavedra Rivano 1972] and [Ziegler 2011] on graded and filtered fiber functors. We consider only descending filtrations; the corresponding results for ascending filtrations follow directly by renumbering.

For any graded fiber functor $\gamma\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ and any morphism $S' \to S$ we let $\gamma_{S'}\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S')$ denote the graded fiber functor obtained by pullback. We call two graded fiber functors $\gamma_1, \gamma_2\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ fpqc-locally isomorphic if their pullbacks under some fpqc morphism $S' \to S$ are isomorphic. In general we let $\underline{\mathrm{Isom}}^{\otimes}(\gamma_1, \gamma_2)$ denote the fpqc-sheaf on $((\mathit{Sch}/S))$ sending $S' \to S$ to the set of isomorphisms $\gamma_{1,S'} \xrightarrow{\sim} \gamma_{2,S'}$. By composition of isomorphisms it carries a natural right action of the sheaf of groups $\underline{\mathrm{Aut}}^{\otimes}(\gamma_1) := \underline{\mathrm{Isom}}^{\otimes}(\gamma_1, \gamma_1)$. The same notation will be used for filtered fiber functors $\psi, \psi_1, \psi_2\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$.

We first consider the special case that $S = \operatorname{Spec} k$ for an overfield k of k_0. Since a locally free sheaf of finite rank on $\operatorname{Spec} k$ is just a finite dimensional k-vector space, we abbreviate

$$\mathtt{Vec}(k) := \mathtt{LF}(\operatorname{Spec} k),$$
$$\mathtt{GrVec}(k) := \mathtt{GrLF}(\operatorname{Spec} k),$$
$$\mathtt{FilVec}^\bullet(k) := \mathtt{FilLF}^\bullet(\operatorname{Spec} k).$$

Let $\omega_{0,k}\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{Vec}(k)$ denote the tautological fiber functor that sends each representation V to the vector space $V_k := V \otimes_{k_0} k$.

Consider a cocharacter $\chi : \mathbb{G}_{m,k} \to \hat{G}_k$. Let $\hat{L}$ denote its centralizer in $\hat{G}_k$, let U denote the unique connected smooth unipotent subgroup of $\hat{G}_k$ that is normalized by $\hat{L}$ and whose Lie algebra is the sum of the weight spaces of weights > 0 under $\mathrm{Ad} \circ \chi$, and set $\hat{P} := \hat{L} \ltimes U$. (If the identity component of $\hat{G}$ is reductive, the identity component of $\hat{P}$ is a parabolic subgroup P and the identity component of $\hat{L}$ is a Levi subgroup of P.)

For any representation $V \in \hat{G}\text{-}\mathtt{Rep}$, the cocharacter χ determines a grading $V_k = \bigoplus_{i\in\mathbb{Z}} V_k^i$. This grading is k_0-linearly functorial in V, exact in short exact sequences, and compatible with tensor product, and the same holds for the associated descending filtration. Thus χ induces a graded fiber functor γ_χ and a filtered fiber functor $\mathrm{fil}^\bullet \circ \gamma_\chi$ such that the composite

$$\hat{G}\text{-}\mathtt{Rep} \xrightarrow{\gamma_\chi} \mathtt{GrVec}(k) \xrightarrow{\mathrm{fil}^\bullet} \mathtt{FilVec}^\bullet(k) \xrightarrow{\mathrm{forg}} \mathtt{Vec}(k)$$

is equal to $\omega_{0,k}$.

Proposition 5.1. (a) *The action of* $\hat{L}$ *on* γ_χ *induces a natural isomorphism*

$$\hat{L} \xrightarrow{\sim} \underline{\mathrm{Aut}}^{\otimes}(\gamma_\chi).$$

(b) *The action of* $\hat{P}$ *on* $\mathrm{fil}^\bullet \circ \gamma_\chi$ *induces a natural isomorphism*

$$\hat{P} \xrightarrow{\sim} \underline{\mathrm{Aut}}^{\otimes}(\mathrm{fil}^\bullet \circ \gamma_\chi).$$

Proof. Part (a) is [Ziegler 2011, Corollary 3.7]. Part (b) is a consequence of [Saavedra Rivano 1972, 2.1.4 and 2.1.5]. □

Now let $\bar{k}_0$ be an algebraic closure of k_0, and let k_0^s denote the separable closure of k_0 in $\bar{k}_0$. Let $\mathscr{C}_{\hat{G}}$ denote the set of $\hat{G}(\bar{k}_0)$-conjugacy classes of cocharacters $\mathbb{G}_{m,\bar{k}_0} \to \hat{G}_{\bar{k}_0}$. The Galois group $\mathrm{Gal}(k_0^s/k_0) \cong \mathrm{Aut}(\bar{k}_0/k_0)$ acts naturally on $\mathscr{C}_{\hat{G}}$. For any $c \in \mathscr{C}_{\hat{G}}$ we let $k_c \subset k_0^s$ denote the fixed field of the stabilizer of c in $\mathrm{Gal}(k_0^s/k_0)$. The fact that c contains a cocharacter which is defined over a finite separable extension of k_0 implies that k_c is finite separable over k_0.

Definition 5.2. We call k_c the *field of definition* of the conjugacy class c.

Next observe that conjugate cocharacters χ, χ' give rise to isomorphic functors γ_χ, γ'_χ, so the following definition depends only on the conjugacy class of χ.

Definition 5.3. Let $c \in \mathscr{C}_{\hat{G}}$ and let S be a scheme over k_c. A graded fiber functor $\gamma\colon\ \hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ is called *of type* c, or *of type* χ for any $\chi \in c$, if the pullbacks of the functors γ and γ_χ to $S \times_{k_c} \bar{k}_0$ are fpqc-locally isomorphic. A filtered fiber functor $\psi\colon\ \hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$ is called *of type* c if the associated graded fiber functor $\mathrm{gr}^\bullet_C \circ \psi\colon\ G\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ is of type c.

Theorem 5.4 [Ziegler 2011, Theorem 3.25]. *Let S be a connected scheme over k_0, and let γ be a graded fiber functor $\hat{G}$-*$\mathtt{Rep} \to \mathtt{GrLF}(S)$. *Then there exist a unique $c \in \mathscr{C}_{\hat{G}}$ and a unique morphism $S \to$* $\mathrm{Spec}\, k_c$ *over k_0, such that γ is of type c. The same assertion holds for any filtered fiber functor* $\psi\colon\ \hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$.

In general, a conjugacy class $c \in \mathscr{C}_{\hat{G}}$ does not have a representative which is defined over k_c. For the following results, we therefore fix a field extension $k_c \subset k \subset \bar{k}_0$ and a representative $\chi \in c$ that is defined over k. Let $\hat{L}$, U, and $\hat{P} = \hat{L} \ltimes U$ be the associated subgroups of $\hat{G}_k$. Let S be a scheme over k.

Theorem 5.5 [Ziegler 2011, Theorem 3.27]. *There is a natural equivalence of categories from the category of graded fiber functors* $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ *of type c to the category of right $\hat{L}$-torsors over S, given by*

$$\gamma \mapsto \underline{\mathrm{Isom}}^{\otimes}(\gamma_{\chi,S}, \gamma).$$

Theorem 5.6 [Ziegler 2011, Theorem 4.43]. *There is a natural equivalence of categories from the category of filtered fiber functors $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$ of type c to the category of right $\hat{P}$-torsors over S, given by*

$$\psi \mapsto \underline{\mathrm{Isom}}^{\otimes}(\mathrm{fil}^\bullet \circ \gamma_{\chi,S}, \psi).$$

Theorem 5.7 [Ziegler 2011, Theorem 4.39]. *For any filtered fiber functor ψ : $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}^\bullet(S)$ of type c, the functor $\mathrm{gr}^\bullet_C$ induces a natural isomorphism of right $\hat{L}$-torsors*

$$\underline{\mathrm{Isom}}^{\otimes}(\mathrm{fil}^\bullet \circ \gamma_{\chi,S}, \psi)/U \cong \underline{\mathrm{Isom}}^{\otimes}(\gamma_{\chi,S}, \mathrm{gr}^\bullet_C \circ \psi).$$

6. F-zips

Definition 6.1. (a) An *F-zip over S* is a tuple $\underline{\mathcal{M}} = (\mathcal{M}, C^\bullet, D_\bullet, \varphi_\bullet)$ consisting of a locally free sheaf of $\mathcal{O}_S$-modules of finite rank $\mathcal{M}$ on S, a descending filtration $C^\bullet$ and an ascending filtration $D_\bullet$ of $\mathcal{M}$, and an $\mathcal{O}_S$-linear isomorphism $\varphi_i : (\mathrm{gr}^i_C \mathcal{M})^{(q)} \xrightarrow{\sim} \mathrm{gr}^D_i \mathcal{M}$ for every $i \in \mathbb{Z}$.

(b) A *homomorphism* $f : \underline{\mathcal{M}} \to \underline{\mathcal{N}}$ of F-zips over S is a homomorphism of the underlying sheaves of $\mathcal{O}_S$-modules $\mathcal{M} \to \mathcal{N}$ which for all $i \in \mathbb{Z}$ satisfies $f(C^i\mathcal{M}) \subset C^i\mathcal{N}$ and $f(D_i\mathcal{M}) \subset D_i\mathcal{N}$ and makes the following diagram commute:

$$\begin{tikzcd}
(\mathrm{gr}^i_C \mathcal{M})^{(q)} \arrow[r, "\varphi_i", "\sim"'] \arrow[d, "(\mathrm{gr}^i_C f)^{(q)}"'] & \mathrm{gr}^D_i \mathcal{M} \arrow[d, "\mathrm{gr}^D_i f"] \\
(\mathrm{gr}^i_C \mathcal{N})^{(q)} \arrow[r, "\varphi_i", "\sim"'] & \mathrm{gr}^D_i \mathcal{N}.
\end{tikzcd}$$

(c) The resulting *category of F-zips over S* is denoted $F\text{-}\mathtt{Zip}(S)$.

The category $F\text{-}\mathtt{Zip}(S)$ is additive, but due to the presence of the Frobenius pullback it is only $\mathbb{F}_q$-linear in general, not $\mathcal{O}_S$-linear. Easy examples show that it is not abelian if $S \neq \varnothing$.

Definition 6.2. A homomorphism of F-zips $\underline{\mathcal{M}} \to \underline{\mathcal{N}}$ is *admissible* if the underlying morphisms of filtered locally free sheaves $\mathcal{M} \to \mathcal{N}$ for both filtrations $C^\bullet$ and $D_\bullet$ is admissible.

This notion turns $F\text{-}\mathtt{Zip}(S)$ into an exact category that is idempotent complete.

Definition 6.3. An F-zip is called *of rank n*, or *of height n*, if its underlying sheaf of $\mathcal{O}_S$-modules is of constant rank n. Let $\underline{n} = (n_i)_{i\in\mathbb{Z}}$ be a family of nonnegative integers which vanish for almost all i. An F-zip $\underline{\mathcal{M}}$ is *of type $\underline{n}$* if $\mathrm{gr}^i_C\mathcal{M}$, or equivalently $\mathrm{gr}^D_i\mathcal{M}$, is locally free of constant rank n_i for all i.

Any F-zip of type $\underline{n}$ is of rank $\sum_i n_i$. If its rank is 1 there exists an integer d such that $n_d = 1$ and $n_i = 0$ for $i \neq d$. In this case we say briefly that the F-zip is of type d. If S is connected, every F-zip over S possesses a unique type.

Definition 6.4. The *tensor product* of F-zips $\underline{\mathcal{M}}$ and $\underline{\mathcal{N}}$ over S is the F-zip $\underline{\mathcal{M}} \otimes \underline{\mathcal{N}}$ consisting of the tensor product $\mathcal{M} \otimes \mathcal{N}$ with the induced descending filtration $C^\bullet$ and the induced ascending filtration $D_\bullet$ of $\mathcal{M} \otimes \mathcal{N}$ and the induced isomorphisms

$$\begin{tikzcd}
(\mathrm{gr}^i_C(\mathcal{M} \otimes \mathcal{N}))^{(q)} \arrow[r, equal, "\sim"] \arrow[d, "\cong"'] & \bigoplus_{j\in\mathbb{Z}} (\mathrm{gr}^j_C \mathcal{M})^{(q)} \otimes (\mathrm{gr}^{i-j}_C \mathcal{N})^{(q)} \arrow[d, "\cong"', "\bigoplus_j (\varphi_j \otimes \varphi_{i-j})"] \\
\mathrm{gr}^D_i(\mathcal{M} \otimes \mathcal{N}) \arrow[r, equal, "\sim"] & \bigoplus_{j\in\mathbb{Z}} \mathrm{gr}^D_j \mathcal{M} \otimes \mathrm{gr}^D_{i-j} \mathcal{N}.
\end{tikzcd}$$

There is also a straightforward definition of tensor product of morphisms of F-zips, we leave it to the reader to verify that this is a homomorphism of F-zips. The tensor product thus defines a functor $F\text{-}\mathtt{Zip}(S) \times F\text{-}\mathtt{Zip}(S) \to F\text{-}\mathtt{Zip}(S)$, which is $\mathbb{F}_q$-bilinear and exact.

Comparing Definition 6.2 with the construction in Section 4G we find that all symmetric powers $\mathrm{S}^m \underline{\mathcal{M}}$ and all alternating powers $\Lambda^m \underline{\mathcal{M}}$ of F-zips exist. They have evident descriptions in terms of the symmetric and alternating powers of the underlying filtered and graded locally free sheaves since symmetric and alternating powers of filtered and graded locally free sheaves are compatible with pullbacks under Frobenius and the functor gr.

Definition 6.5. The *dual* of an F-zip $\underline{\mathcal{M}}$ over S is the F-zip $\underline{\mathcal{M}}^\vee$ consisting of the dual sheaf of $\mathcal{O}_S$-modules $\mathcal{M}^\vee$ equipped with the duals of the filtrations $C^\bullet(\mathcal{M})$ and $D_\bullet(\mathcal{M})$ whose terms are given by $\mathrm{gr}^i_C(\mathcal{M}^\vee) = (\mathrm{gr}^{-i}_C \mathcal{M})^\vee$ and $\mathrm{gr}^D_i(\mathcal{M}^\vee) = (\mathrm{gr}^D_{-i} \mathcal{M})^\vee$, and the induced isomorphisms

$$\begin{tikzcd}
(\mathrm{gr}^i_C(\mathcal{M}^\vee))^{(q)} \arrow[r, equal] \arrow[d, "\cong"'] & ((\mathrm{gr}^{-i}_C \mathcal{M})^\vee)^{(q)} \arrow[d, "\cong"', "(\varphi_{-i}^{-1})^\vee"] \\
\mathrm{gr}^D_i(\mathcal{M}^\vee) \arrow[r, equal, "\sim"] & (\mathrm{gr}^D_{-i} \mathcal{M})^\vee.
\end{tikzcd}$$

There is an evident notion of the dual of a homomorphism of F-zips, so that we obtain a functor $F\text{-}\mathtt{Zip}(S)^{\mathrm{op}} \to F\text{-}\mathtt{Zip}(S)$, which is $\mathbb{F}_q$-linear and exact.

As usual the tensor product and the dual yields the notion of an internal Hom of two F-zips $\underline{\mathcal{M}}$ and $\underline{\mathcal{N}}$ over S by setting

$$\underline{\mathrm{Hom}}(\underline{\mathcal{M}}, \underline{\mathcal{N}}) := \underline{\mathcal{M}}^\vee \otimes \underline{\mathcal{N}}.$$

Example 6.6. The *Tate F-zip of weight $d \in \mathbb{Z}$* is $\underline{\mathbb{1}}(d) := (\mathcal{O}_S, C^\bullet, D_\bullet, \varphi_\bullet)$, where

$$C^i = \begin{cases} \mathcal{O}_S & \text{for } i \leq d; \\ 0 & \text{for } i > d; \end{cases} \qquad D_i = \begin{cases} 0 & \text{for } i < d; \\ \mathcal{O}_S & \text{for } i \geq d; \end{cases}$$

and φ_d is the identity on $\mathcal{O}_S^{(q)} = \mathcal{O}_S$. Thus $\underline{\mathbb{1}}(d)$ is an F-zip of rank 1 and type d. There are natural isomorphisms $\underline{\mathbb{1}}(d) \otimes \underline{\mathbb{1}}(d') \cong \underline{\mathbb{1}}(d+d')$ and $\underline{\mathbb{1}}(d)^\vee \cong \underline{\mathbb{1}}(-d)$. The *$d$-th Tate twist* of an F-zip $\underline{\mathcal{M}}$ is defined as $\underline{\mathcal{M}}(d) := \underline{\mathcal{M}} \otimes \underline{\mathbb{1}}(d)$, and there is a natural isomorphism $\underline{\mathcal{M}}(0) \cong \underline{\mathcal{M}}$.

With the above tensor product and dual and the unit object $\underline{\mathbb{1}}(0)$ the category $F\text{-}\mathtt{Zip}(S)$ is a rigid tensor category. It is endowed with the following natural exact $\mathbb{F}_q$-linear tensor functors:

$$\begin{aligned} \mathrm{forg}\colon \; & F\text{-}\mathtt{Zip}(S) \to \mathtt{LF}(S), & \underline{\mathcal{M}} &\mapsto \mathcal{M}, \\ \mathrm{fil}^\bullet\colon \; & F\text{-}\mathtt{Zip}(S) \to \mathtt{FilLF}^\bullet(S), & \underline{\mathcal{M}} &\mapsto (\mathcal{M}, C^\bullet), \\ \mathrm{fil}_\bullet\colon \; & F\text{-}\mathtt{Zip}(S) \to \mathtt{FilLF}_\bullet(S), & \underline{\mathcal{M}} &\mapsto (\mathcal{M}, D_\bullet). \end{aligned}$$

The isomorphism $\varphi_\bullet\colon (\mathrm{gr}^\bullet_C \mathcal{M})^{(q)} \xrightarrow{\sim} \mathrm{gr}^D_\bullet \mathcal{M}$ that is part of an F-zip induces an isomorphism of tensor functors $\varphi \circ (\;)^{(q)}\colon \mathrm{gr}^\bullet_C \circ \mathrm{fil}^\bullet \to \mathrm{gr}^D_\bullet \circ \mathrm{fil}_\bullet$. Combined with some of the functors from (4.1) we obtain the following diagram

(6.7)
$$\begin{tikzcd}
 & \mathtt{FilLF}^\bullet(S) \arrow[dl, "\mathrm{gr}^\bullet_C"'] \arrow[drr, bend left, "\mathrm{forg}"] & & \\
\mathtt{GrLF}(S) & \varphi \circ (\;)^{(q)} \circlearrowright & F\text{-}\mathtt{Zip}(S) \arrow[ul, "\mathrm{fil}^\bullet"'] \arrow[r, "\mathrm{forg}"] \arrow[dl, "\mathrm{fil}_\bullet"] & \mathtt{LF}(S) \\
 & \mathtt{FilLF}_\bullet(S), \arrow[ul, "\mathrm{gr}^D_\bullet"] \arrow[urr, bend right, "\mathrm{forg}"'] & &
\end{tikzcd}$$

which is commutative except that the left hand side commutes only up to $\varphi \circ (\;)^{(q)}$.

Also, there is an evident notion of pullback of F-zips under morphisms $S' \to S$, compatible with everything discussed above. Lemma 4.2 directly implies:

Lemma 6.8. *For a homomorphism of F-zips, the property of being admissible is local for the fpqc topology.*

7. $\hat{G}$-zip functors

Throughout this section we fix a (not necessarily connected) linear algebraic group $\hat{G}$ over $\mathbb{F}_q$. Let S be a scheme over $\mathbb{F}_q$. It is known (for example by [Nori 1976, Proposition 2.9]) that giving an exact $\mathbb{F}_q$-linear tensor functor $\hat{G}\text{-}\mathtt{Rep} \to \mathtt{LF}(S)$ is equivalent to giving a $\hat{G}$-torsor over S. This suggests the idea that an exact

$\mathbb{F}_q$-linear tensor functor from $\hat{G}$-$\mathtt{Rep}$ to an arbitrary exact $\mathbb{F}_q$-linear tensor category $\mathscr{C}$ may be viewed as a "$\hat{G}$-torsor in $\mathscr{C}$", which underlies Section 8 of Deligne's article [1990]. In the present section we apply this point of view to the category of F-zips and describe an equivalence between exact $\mathbb{F}_q$-linear tensor functors $\hat{G}$-$\mathtt{Rep} \to F$-$\mathtt{Zip}(S)$ and $\hat{G}$-zips.

7A. *The stack of $\hat{G}$-zip functors.*

Definition 7.1. (a) A *$\hat{G}$-zip functor over S* is an exact $\mathbb{F}_q$-linear tensor functor

$$\mathfrak{z}\colon\ \hat{G}\text{-}\mathtt{Rep} \to F\text{-}\mathtt{Zip}(S).$$

(b) A *morphism* of $\hat{G}$-zip functors over S is a natural transformation that is compatible with the tensor product.

(c) The resulting category of $\hat{G}$-zip functors over S is denoted $\hat{G}$-$\mathtt{ZipFun}(S)$.

With the evident notion of pullback the $\hat{G}$-$\mathtt{ZipFun}(S)$ form a fibered category over the category $((Sch/\mathbb{F}_q))$ of schemes over $\mathbb{F}_q$, which we denote $\hat{G}$-$\mathtt{ZipFun}$.

Proposition 7.2. *$\hat{G}$-$\mathtt{ZipFun}$ is a stack.*

Proof. Since $\hat{G}$-$\mathtt{Rep}$ and F-$\mathtt{Zip}(S)$ are rigid tensor categories, any morphism of $\hat{G}$-zip functors is an isomorphism (see [Saavedra Rivano 1972, I.5.2.3]); hence $\hat{G}$-$\mathtt{ZipFun}$ is fibered in groupoids. It remains to prove that $\hat{G}$-$\mathtt{ZipFun}$ satisfies effective descent for morphisms and objects. For this let $S' \to S$ be an fpqc covering and set $S'' := S' \times_S S'$.

First consider objects $\mathfrak{z}_1, \mathfrak{z}_2 \in \hat{G}$-$\mathtt{ZipFun}(S)$ and a morphism $\lambda'\colon \mathfrak{z}_{1,S'} \to \mathfrak{z}_{2,S'}$ whose two pullbacks to S'' coincide. Since morphisms of F-zips satisfy effective descent with respect to the fpqc topology, for any $V \in \hat{G}$-$\mathtt{Rep}$ the homomorphism $\lambda'(V)\colon \mathfrak{z}_1(V)_{S'} \to \mathfrak{z}_2(V)_{S'}$ comes from a unique homomorphism $\lambda(V)\colon \mathfrak{z}_1(V) \to \mathfrak{z}_2(V)$. In order for λ to be a tensor morphism, certain diagrams in F-$\mathtt{Zip}(S)$ need to commute. But as λ' is a tensor morphism, these diagrams commute after pullback to S'; hence by descent they commute over S and thus λ is a tensor morphism. Therefore $\hat{G}$-$\mathtt{ZipFun}$ satisfies effective descent for morphisms.

Now consider an object $\mathfrak{z}'$ of $\hat{G}$-$\mathtt{ZipFun}(S')$ equipped with a descent datum. For each $V \in \hat{G}$-$\mathtt{Rep}$ this descent datum induces a descent datum on $\mathfrak{z}'(V)$; hence it yields an object $\mathfrak{z}(V)$ of F-$\mathtt{Zip}(S)$ with $\mathfrak{z}'(V) = \mathfrak{z}(V)_{S'}$. Next, the descent datum on $\mathfrak{z}'(V)$ depends functorially on V. Thus for each morphism $f\colon V \to V'$ in $\hat{G}$-$\mathtt{Rep}$ the two pullbacks of $\mathfrak{z}'(f)\colon \mathfrak{z}(V)_{S'} \to \mathfrak{z}(V')_{S'}$ coincide and therefore come from a unique morphism $\mathfrak{z}(f)\colon \mathfrak{z}(V) \to \mathfrak{z}(V')$. The uniquess of $\mathfrak{z}(f)$ implies that $\mathfrak{z}\colon\ \hat{G}$-$\mathtt{Rep} \to F$-$\mathtt{Zip}(S)$ is a functor. Making $\mathfrak{z}$ into a tensor functor requires functorial isomorphisms $\mathfrak{z}(\dot{-}) \cong \dot{-}$ and $\mathfrak{z}(V) \otimes \mathfrak{z}(V') \cong \mathfrak{z}(V \otimes V')$ for all V, $V' \in \hat{G}$-$\mathtt{Rep}$ which are compatible with the associativity, commutativity and unit constraints of the tensor category F-$\mathtt{Zip}(S)$. These are again obtained by descent from the corresponding isomorphisms for $\mathfrak{z}'$, and the compatibility with the constraints holds because it holds after pullback to S'. Finally, by Lemma 6.8 the exactness of $\mathfrak{z}$ follows from the exactness of $\mathfrak{z}'$. Altogether $\mathfrak{z}$ is an element of $\hat{G}\text{-}\mathtt{ZipFun}(S)$ which gives rise to $\mathfrak{z}'$ with its descent datum. Thus $\hat{G}\text{-}\mathtt{ZipFun}$ satisfies effective descent for objects and we are done. □

To analyze a zip functor we will compose it with the functors forg, $\mathrm{fil}^\bullet$, and $\mathrm{fil}_\bullet$ from (6.7). First we look at the numerical invariants obtained from the filtrations. Let $\bar{\mathbb{F}}_q$ be an algebraic closure of $\mathbb{F}_q$. As in Section 5 let $\mathscr{C}_{\hat{G}}$ denote the set of $\hat{G}(\bar{\mathbb{F}}_q)$-conjugacy classes of cocharacters $\mathbb{G}_{m,\bar{\mathbb{F}}_q} \to \hat{G}_{\bar{\mathbb{F}}_q}$, and let $k_c \subset \bar{\mathbb{F}}_q$ denote the field of definition of an element $c \in \mathscr{C}_{\hat{G}}$, which is a finite extension of $\mathbb{F}_q$.

Definition 7.3. Let $c \in \mathscr{C}_{\hat{G}}$ and let S be a scheme over k_c.

(a) A $\hat{G}$-zip functor $\mathfrak{z}$ over S is called *of type c*, or *of type $\chi \in c$*, if the associated functor $\mathrm{gr}_C^\bullet \circ \mathrm{fil}^\bullet \circ \mathfrak{z}\colon \hat{G}\text{-}\mathtt{Rep} \to \mathtt{GrLF}(S)$ is of type c in the sense of Definition 5.3.

(b) The full subcategory of $\hat{G}\text{-}\mathtt{ZipFun}(S)$ whose objects are the G-zip functors of type c is denoted $\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}(S)$.

With the evident notion of pullback the categories $\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}(S)$ form a fibered category over the category $((Sch/k_c))$, which we denote $\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}$. Since Definition 5.3 is local for the fpqc topology, Proposition 7.2 and Definition 7.3 directly imply:

Proposition 7.4. *$\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}$ is a substack of $\hat{G}\text{-}\mathtt{ZipFun}_{k_c}$.*

The next result says that every zip functor over a connected scheme has a type.

Proposition 7.5. *Let S be a connected scheme over $\mathbb{F}_q$ and $\mathfrak{z}$ a $\hat{G}$-zip functor over S. Then there exist a unique $c \in \mathscr{C}_{\hat{G}}$ and a unique morphism $S \to \operatorname{Spec} k_c$ over $\mathbb{F}_q$ such that $\mathfrak{z}$ is of type c.*

Proof. Direct consequence of Definition 7.3 and Theorem 5.4. □

Corollary 7.6. (a) *Each $\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}$, viewed as a stack over $\mathbb{F}_q$ by Grothendieck restriction, is an open and closed substack of $\hat{G}\text{-}\mathtt{ZipFun}$.*

(b) *$\hat{G}\text{-}\mathtt{ZipFun}$ is the disjoint union of the $\hat{G}\text{-}\mathtt{ZipFun}^c_{k_c}$ taken over all $c \in \mathscr{C}_{\hat{G}}$.*

Theorem 7.7 [Ziegler 2011, Theorem 3.32]. *For any $c \in \mathscr{C}_{\hat{G}}$ there exists an inner form*

$$(\hat{G}', \tau\colon \hat{G}'_{\bar{k}} \xrightarrow{\sim} \hat{G}_{\bar{k}})$$

of $\hat{G}$ defined over k_c and a cocharacter $\chi\colon \mathbb{G}_{m,k_c} \to \hat{G}'$ such that $\tau \circ \chi_{\bar{k}}$ lies in c.

Remark 7.8. From the Tannakian viewpoint, replacing $\hat{G}$ by an inner form does not change the category $\hat{G}$-$\mathtt{Rep}$; it merely endows it with a different fiber functor. In particular it does not change the stack of $\hat{G}$-zip functors. Thus Theorem 7.7 implies that to study zip functors of a given type c, we may without loss of generality assume that c has a representative χ which is defined over k_c.

7B. *Equivalence with $\hat{G}$-zips.* We now assume that the identity component G of $\hat{G}$ is reductive. We fix a finite field extension k of $\mathbb{F}_q$ and a cocharacter $\chi : \mathbb{G}_{m,k} \to \hat{G}_k$. We let $\hat{L} \subset \hat{G}_k$ denote the centralizer of χ and set $\Theta := \pi_0(\hat{L}) \subset \pi_0(\hat{G}_k)$. Then we are in the situation of Section 3B with the maximal possible choice of Θ. We will use all the pertaining notation from Section 3. Let c, $c^{(q)} \in \mathscr{C}_{\hat{G}}$ denote the conjugacy classes of χ, $\chi^{(q)}$.

Construction 7.9. For any finite dimensional representation V of $\hat{G}$ over $\mathbb{F}_q$, the cocharacter χ determines a grading $V_k = \bigoplus_{i\in\mathbb{Z}} V_k^i$. This grading induces a descending filtration $C^\bullet(V_k)$. Also, the definition of V_k by base extension induces a natural isomorphism $V_k^{(q)} \cong V_k$. Thus we may consider the decomposition $\bigoplus_{i\in\mathbb{Z}} (V_k^i)^{(q)}$ as another grading of V_k, namely that induced by the cocharacter $\chi^{(q)}$. This grading induces an ascending filtration $D_\bullet(V_k)$. Then for all $i \in \mathbb{Z}$ we obtain natural isomorphisms $\varphi_i(V_k)\colon (\mathrm{gr}^i_C(V_k))^{(q)} \xrightarrow{\sim} (V_k^i)^{(q)} \xrightarrow{\sim} \mathrm{gr}^D_i(V_k)$. Altogether this data defines an F-zip over k, denoted

$$\mathfrak{z}_1(V) := \big(V_k, C^\bullet(V_k), D_\bullet(V_k), \varphi_\bullet(V_k)\big).$$

Clearly this construction is $\mathbb{F}_q$-linearly functorial in V and compatible with tensor product. It therefore defines a $\hat{G}$-zip functor over k

$$\mathfrak{z}_1\colon\ \hat{G}\text{-}\mathtt{Rep} \to F\text{-}\mathtt{Zip}(\operatorname{Spec} k).$$

By pullback we obtain a zip functor $\mathfrak{z}_{1,S}$ over any scheme S over k. We will measure an arbitrary zip functor over S by how it differs from this basic zip functor $\mathfrak{z}_{1,S}$.

Lemma 7.10. *There are natural isomorphisms*

(a) $\underline{\mathrm{Aut}}^\otimes(\mathrm{forg} \circ \mathfrak{z}_1) \cong \hat{G}_k$,

(b) $\underline{\mathrm{Aut}}^\otimes(\mathrm{fil}^\bullet \circ \mathfrak{z}_1) \cong \hat{P}_+$,

(c) $\underline{\mathrm{Aut}}^\otimes(\mathrm{fil}_\bullet \circ \mathfrak{z}_1) \cong \hat{P}_-^{(q)}$.

Proof. Assertion (a) is an instance of the main theorem of Tannaka duality [Deligne 1990, Theorem 1.12], and (b) and (c) are instances of Proposition 5.1. □

Construction 7.11. Let S be a scheme over k, and $\mathfrak{z}$ a $\hat{G}$-zip functor of type c over S. Then

(a) $I_\mathfrak{z} := \underline{\mathrm{Isom}}^\otimes(\mathrm{forg} \circ \mathfrak{z}_{1,S}, \mathrm{forg} \circ \mathfrak{z})$ is a right $\hat{G}_k$-torsor over S,

(b) $I_{\mathfrak{z},+} := \underline{\mathrm{Isom}}^\otimes(\mathrm{fil}^\bullet \circ \mathfrak{z}_{1,S}, \mathrm{fil}^\bullet \circ \mathfrak{z})$ is a right $\hat{P}_+$-torsor over S,

(c) $I_{\mathfrak{z},-} := \underline{\mathrm{Isom}}^{\otimes}(\mathrm{fil}_\bullet \circ \mathfrak{z}_{1,S}, \mathrm{fil}_\bullet \circ \mathfrak{z})$ is a right $\hat{P}_-^{(q)}$-torsor over S.

Indeed, (a) follows from Lemma 7.10 (a) and [Deligne 1990, Theorem 1.12], and (b) results from combining Lemma 7.10 (b) with Theorem 5.6 above. Also, the commutativity of (6.7) shows that $(\ )^q \circ \mathrm{gr}_C^\bullet \circ \mathrm{fil}^\bullet \circ \mathfrak{z} \cong \mathrm{gr}_\bullet^D \circ \mathrm{fil}_\bullet \circ \mathfrak{z}$ is of type $c^{(q)}$, and so (c) follows from Lemma 7.10 (c) and Theorem 5.6.

Moreover, composition with the functors forgetting the filtration induces a natural $\hat{P}_+$-equivariant embedding

$$\begin{tikzcd} I_{\mathfrak{z},+} \arrow[r, hook] & \underline{\mathrm{Isom}}^{\otimes}(\mathrm{forg} \circ \mathrm{fil}^\bullet \circ \mathfrak{z}_{1,S}, \mathrm{forg} \circ \mathrm{fil}^\bullet \circ \mathfrak{z}) \arrow[d, equal, "\wr"] \\ I_{\mathfrak{z}} \arrow[r, equal, "\sim"] & \underline{\mathrm{Isom}}^{\otimes}(\mathrm{forg} \circ \mathfrak{z}_{1,S}, \mathrm{forg} \circ \mathfrak{z}) \end{tikzcd}$$

and likewise a natural $\hat{P}_-^{(q)}$-equivariant embedding $I_{\mathfrak{z},-} \hookrightarrow I_{\mathfrak{z}}$. Furthermore, by Theorem 5.7 we have natural isomorphisms of $\hat{L}^{(q)}$-torsors in the rows of the following diagram, where the vertical isomorphism is induced by the isomorphism of tensor functors $\varphi\colon (\ )^{(q)} \circ \mathrm{gr}_C^\bullet \circ \mathrm{fil}^\bullet \xrightarrow{\sim} \mathrm{gr}_\bullet^D \circ \mathrm{fil}_\bullet$ from (6.7):

$$\begin{tikzcd} I_{\mathfrak{z},+}^{(q)}/U_+^{(q)} \arrow[r, equal, "\sim"] & \underline{\mathrm{Isom}}^{\otimes}((\ )^{(q)} \circ \mathrm{gr}_C^\bullet \circ \mathrm{fil}^\bullet \circ \mathfrak{z}_{1,S}, (\ )^{(q)} \circ \mathrm{gr}_C^\bullet \circ \mathrm{fil}^\bullet \circ \mathfrak{z}) \arrow[d, "\wr"] \\ I_{\mathfrak{z},-}/U_-^{(q)} \arrow[r, equal, "\sim"] & \underline{\mathrm{Isom}}^{\otimes}(\mathrm{gr}_\bullet^D \circ \mathrm{fil}_\bullet \circ \mathfrak{z}_{1,S}, \mathrm{gr}_\bullet^D \circ \mathrm{fil}_\bullet \circ \mathfrak{z}) \end{tikzcd}$$

The composite is therefore an isomorphism of $\hat{L}^{(q)}$-torsors $\iota_{\mathfrak{z}}\colon I_{\mathfrak{z},+}^{(q)}/U_+^{(q)} \xrightarrow{\sim} I_{\mathfrak{z},-}/U_-^{(q)}$. Together this data defines a $\hat{G}$-zip of type (χ, Θ) over S

$$\underline{I}_{\mathfrak{z}} := (I_{\mathfrak{z}}, I_{\mathfrak{z},+}, I_{\mathfrak{z},-}, \iota_{\mathfrak{z}}).$$

Clearly this construction is $\mathbb{F}_q$-linearly functorial in $\mathfrak{z}$ and compatible with pullback. Thus it defines a morphism of stacks

$$\hat{G}\text{-}\mathtt{ZipFun}_k^c \to \hat{G}\text{-}\mathtt{Zip}_k^{(\chi,\Theta)}, \mathfrak{z} \mapsto \underline{I}_{\mathfrak{z}}. \tag{7.12}$$

Theorem 7.13. *The morphism (7.12) is an isomorphism.*

Proof. We construct a morphism in the other direction, as follows. Consider a $\hat{G}$-zip $\underline{I} = (I, I_+, I_-, \iota)$ of type (χ, Θ) over S. The essential surjectivity in Theorem 5.6 shows that $I_- \cong \underline{\mathrm{Isom}}^{\otimes}(\mathrm{fil}_\bullet \circ \mathfrak{z}_{1,S}, \psi_-)$ for some exact $\mathbb{F}_q$-linear tensor functor

$$\psi_-\colon\ \hat{G}\text{-}\mathtt{Rep} \to \mathtt{FilLF}_\bullet(S), \qquad V \mapsto (\mathcal{M}(V), D_\bullet).$$

The embedding $I_- \hookrightarrow I$ and the fullness in Theorem 5.6 then yield an isomorphism $I \cong \underline{\mathrm{Isom}}^{\otimes}(\mathrm{forg} \circ \mathfrak{z}_{1,S}, \omega)$ with $\omega := \mathrm{forg} \circ \psi_-\colon V \mapsto \mathcal{M}(V)$. The essential surjectivity in Theorem 5.6 also shows that $I_+ \cong \underline{\mathrm{Isom}}^{\otimes}(\mathrm{fil}^\bullet \circ \mathfrak{z}_{1,S}, \psi_+)$ for some exact $\mathbb{F}_q$-linear tensor functor $\psi_+\colon\ \hat{G}\text{-}\mathtt{Rep}\to\mathtt{FilLF}^\bullet(S)$. The embedding $I_+\hookrightarrow I$ and the fullness in Theorem 5.6 then yield an isomorphism $\mathrm{forg}\circ\psi_+\cong\omega$. After replacing ψ_+ by an isomorphic functor we may therefore assume that ψ_+ has the form $V\mapsto(\mathcal{M}(V),C^\bullet)$. Moreover, Theorem 5.7 and ι yield isomorphisms

$$\begin{array}{ccc} I_+^{(q)}/U_+^{(q)} & \overset{\sim}{=} & \underline{\mathrm{Isom}}^{\otimes}((\ )^{(q)}\circ\mathrm{gr}_C^\bullet\circ\mathrm{fil}^\bullet\circ\mathfrak{z}_{1,S},\ (\ )^{(q)}\circ\mathrm{gr}_C^\bullet\circ\psi_+) \\ {\scriptstyle\iota}\big\downarrow{\scriptstyle\wr} & & \\ I_-/U_-^{(q)} & \overset{\sim}{=\!=\!=\!=} & \underline{\mathrm{Isom}}^{\otimes}(\mathrm{gr}_\bullet^D\circ\mathrm{fil}_\bullet\circ\mathfrak{z}_{1,S},\ \mathrm{gr}_\bullet^D\circ\psi_-). \end{array}$$

Thus the isomorphism $\varphi_\bullet\colon\ (\ )^{(q)}\circ\mathrm{gr}\circ\mathrm{fil}^\bullet\circ\mathfrak{z}_{1,S}\xrightarrow{\sim}\mathrm{gr}\circ\mathrm{fil}_\bullet\circ\mathfrak{z}_{1,S}$ from Construction 7.9 and the fullness in Theorem 5.5 yield an isomorphism $(\ )^{(q)}\circ\mathrm{gr}_C^\bullet\circ\psi_+\xrightarrow{\sim}\mathrm{gr}_\bullet^D\circ\psi_-$. This amounts to graded isomorphisms $\varphi_\bullet\colon\ (\mathrm{gr}_C^\bullet\mathcal{M}(V))^{(q)}\xrightarrow{\sim}\mathrm{gr}_\bullet^D\mathcal{M}(V)$ for all $V\in\hat{G}\text{-}\mathtt{Rep}$ that are functorial and compatible with tensor product. The assembled data thus determines a $\hat{G}$-zip functor

$$\mathfrak{z}_{\underline{I}}\colon\ V\mapsto(\mathcal{M}(V),C^\bullet,D_\bullet,\varphi_\bullet)$$

of type c over S. By definition it satisfies $\mathrm{fil}^\bullet\circ\mathfrak{z}=\psi_+$ and $\mathrm{fil}_\bullet\circ\mathfrak{z}=\psi_-$; comparing this construction with Construction 7.11 therefore yields an isomorphism $\underline{I}_{\mathfrak{z}_{\underline{I}}}\cong\underline{I}$.

The faithfulness in Theorems 5.6 and 5.5 implies that $\mathfrak{z}$ is unique up to unique isomorphism. It is therefore functorial in $\underline{I}$. As the construction is clearly compatible with pullback, it thus defines a morphism of stacks $\hat{G}\text{-}\mathtt{Zip}_k^{(\chi,\Theta)}\to\hat{G}\text{-}\mathtt{ZipFun}_k^c$. Again by faithfulness the isomorphism $\underline{I}_{\mathfrak{z}_{\underline{I}}}\cong\underline{I}$ is functorial in $\underline{I}$ and compatible with pullback; hence $\underline{I}\mapsto\mathfrak{z}_{\underline{I}}$ is a right inverse of $\mathfrak{z}\mapsto\underline{I}_{\mathfrak{z}}$. Moreover, applying the above construction to $\underline{I}_{\mathfrak{z}}$ for a $\hat{G}$-zip functor $\mathfrak{z}$ one easily shows that $\mathfrak{z}_{\underline{I}_{\mathfrak{z}}}\cong\mathfrak{z}$, and so the morphism is also a left inverse. Thus the morphism (7.12) has a two-sided inverse and is therefore an isomorphism, as desired. □

8. *F*-zips with additional structure

An important tool in the study of vector bundles is the equivalence between vector bundles of constant rank n on a scheme S and the associated GL_n-torsors. One also uses the equivalence between vector bundles with a nondegenerate symmetric, alternating, resp. hermitian pairing and the associated torsors with respect to the orthogonal, symplectic, resp. unitary group. In this section we describe similar equivalences between F-zips of constant rank n and GL_n-zips, and between F-zips with additional structure such as a pairing and G-zips for certain associated linear algebraic groups G.

Let $\underline{n}=(n_i)_{i\in\mathbb{Z}}$ be a family of nonnegative integers which vanish for almost all i, such that $n:=\sum_i n_i\geqslant 1$.

8A. ***F*-zips versus GL_n-zips.** For any scheme S over $k := \mathbb{F}_q$ we let $F\text{-}\mathtt{Zip}_k^{\underline{n}}(S)$ denote the category whose objects are all F-zips of type $\underline{n}$ over S according to Definition 6.3 and whose morphisms are all isomorphisms. For varying S this defines a category $F\text{-}\mathtt{Zip}_k^{\underline{n}}$ fibered in groupoids over $((\mathit{Sch}/k))$. Since F-zips consist of quasicoherent sheaves and homomorphisms thereof, they satisfy effective descent with respect to any fpqc morphism $S' \to S$. Therefore $F\text{-}\mathtt{Zip}_k^{\underline{n}}$ is a stack.

Choose a cocharacter $\chi : \mathbb{G}_{m,k} \to \mathrm{GL}_{n,k}$ whose weights on the standard representation k^n of $\mathrm{GL}_{n,k}$ are i with multiplicity n_i for all i. This determines a grading of k^n, whose associated descending and ascending filtrations we denote by $C^\bullet$ and $D_\bullet$. Since $k = \mathbb{F}_q$, there is a natural isomorphism $\varphi_{1\bullet} \colon (\mathrm{gr}_C^\bullet k^n)^{(q)} = \mathrm{gr}_C^\bullet k^n \xrightarrow{\sim} \mathrm{gr}_\bullet^D k^n$ turning

$$\underline{\mathcal{M}}_1 := (k^n, C^\bullet, D_\bullet, \varphi_{1\bullet}) \tag{8.1}$$

into an F-zip of type $\underline{n}$ over k. As in Section 7B, we compare arbitrary F-zips of type $\underline{n}$ with this basic one.

Set $G := \mathrm{GL}_{n,k}$, and let $P_\pm = L \ltimes U_\pm$ be the parabolics of G associated to χ, as in Section 3B. Thus P_+ is the stabilizer of the filtration $C^\bullet$, and P_- is the stabilizer of $D_\bullet$. Since χ is defined over $\mathbb{F}_q$, we have $\chi^{(q)} = \chi$ and $P_\pm^{(q)} = L^{(q)} \ltimes U_\pm^{(q)} = P_\pm = L \ltimes U_\pm$. Also, since G is connected, we have $\Theta = 1$ in this case.

In the following, for any graded, filtered, or naked sheaves of $\mathcal{O}_S$-modules $\mathcal{M}_1$ and $\mathcal{M}_2$ we let $\underline{\mathrm{Isom}}(\mathcal{M}_1, \mathcal{M}_2)$ denote the fpqc-sheaf on $((\mathit{Sch}/S))$ sending $S' \to S$ to the set of (graded, filtered, resp. neither) isomorphisms $\mathcal{M}_{1,S'} \xrightarrow{\sim} \mathcal{M}_{2,S'}$. By composition of isomorphisms it carries a natural right action of the sheaf of groups $\underline{\mathrm{Isom}}(\mathcal{M}_1, \mathcal{M}_1) = \underline{\mathrm{Aut}}(\mathcal{M}_1)$. This sheaf is representable by a smooth affine group scheme over S if $\mathcal{M}_1$ is locally free of finite rank.

Construction 8.2. Let $\underline{\mathcal{M}} = (\mathcal{M}, C^\bullet, D_\bullet, \varphi_\bullet)$ be an F-zip of type $\underline{n}$ over S. Then $\mathcal{M}$ is a locally free sheaf of rank n, and the filtered sheaves $(\mathcal{M}, C^\bullet)$ and $(\mathcal{M}, D_\bullet)$ are Zariski locally isomorphic to $(k^n, C^\bullet)_S$ and $(k^n, D_\bullet)_S$, respectively. Thus

(a) $I := \underline{\mathrm{Isom}}((k^n)_S, \mathcal{M})$ is a right GL_n-torsor over S,

(b) $I_+ := \underline{\mathrm{Isom}}\big((k^n, C^\bullet)_S, (\mathcal{M}, C^\bullet)\big)$ is a right P_+-torsor over S,

(c) $I_- := \underline{\mathrm{Isom}}\big((k^n, D_\bullet)_S, (\mathcal{M}, D_\bullet)\big)$ is a right $P_-^{(q)}$-torsor over S.

Forgetting the filtration induces natural equivariant embeddings $I_\pm \hookrightarrow I$. Also, the functors $\mathrm{gr}_C^\bullet$ and $\mathrm{gr}_\bullet^D$ induce natural isomorphisms $I_+/U_+ \cong \underline{\mathrm{Isom}}((\mathrm{gr}_C^\bullet k^n)_S, \mathrm{gr}_C^\bullet \mathcal{M})$ and $I_-/U_-^{(q)} \cong \underline{\mathrm{Isom}}((\mathrm{gr}_\bullet^D k^n)_S, \mathrm{gr}_\bullet^D \mathcal{M})$. Moreover, the isomorphisms $\varphi_{1\bullet} \colon (\mathrm{gr}_C^\bullet k^n)^{(q)} \xrightarrow{\sim} \mathrm{gr}_\bullet^D k^n$ and $\varphi_\bullet \colon (\mathrm{gr}_C^\bullet \mathcal{M})^{(q)} \xrightarrow{\sim} \mathrm{gr}_\bullet^D \mathcal{M}$ induce an isomorphism ι of L-torsors making the following diagram commute:

$$\begin{array}{ccc}
I_+^{(q)}/U_+^{(q)} & \overset{\sim}{=} & \underline{\mathrm{Isom}}\big((\mathrm{gr}^\bullet_C k^n)^{(q)}_S, (\mathrm{gr}^\bullet_C \mathcal{M})^{(q)}\big) \\
{\scriptstyle\iota}\big\downarrow & & \big\downarrow{\scriptstyle\wr} \\
I_-/U_-^{(q)} & \overset{\sim}{=\!=} & \underline{\mathrm{Isom}}\big((\mathrm{gr}^D_\bullet k^n)_S, \mathrm{gr}^D_\bullet \mathcal{M}\big).
\end{array}$$

Together this data defines a GL_n-zip $\underline{I} := (I, I_+, I_-, \iota)$ of type χ over S. Clearly this construction is k-linearly functorial in $\underline{\mathcal{M}}$ and compatible with pullback. Thus it defines a morphism of stacks

$$F\text{-}\mathtt{Zip}_k^{\underline{n}} \longrightarrow \mathrm{GL}_n\text{-}\mathtt{Zip}_k^{\chi}\ . \tag{8.3}$$

Proposition 8.4. *This morphism is an isomorphism.*

Proof. A morphism in the other direction is obtained by evaluating the inverse of the morphism (7.12) at the standard representation $V = k^n$ of $\mathrm{GL}_{n,k}$. It is straightforward to check that these two morphisms are mutually inverse. □

Combined with Theorem 7.13 this shows in particular that giving an F-zip of type $\underline{n}$, or a GL_n-zip of type χ, or a GL_n-zip functor of type χ, are all equivalent. Also, combined with Proposition 3.11 it shows that the isomorphism classes of F-zips of type $\underline{n}$ over an algebraically closed field K containing $\mathbb{F}_q$ are in bijection with the $E_{\mathrm{GL}_n,\chi}(K)$-orbits on $\mathrm{GL}_n(K)$, which in turn have a combinatorial description in terms of the Weyl group of GL_n, as in Example 3.23. The analogous remarks apply to the cases treated in the rest of this section.

8B. ***F-zips with trivialized determinant versus*** $\mathbf{SL}_n$***-zips.*** Keeping the notations of the preceding subsection, we now assume that $\sum_i n_i i = 0$. Then the highest exterior power of any F-zip of type $\underline{n}$ is an F-zip of rank 1 whose filtrations are concentrated in degree 0. We call a pair $(\underline{\mathcal{M}}, \Delta)$ consisting of an F-zip $\underline{\mathcal{M}}$ of type $\underline{n}$ and an isomorphism $\Delta\colon \Lambda^n\underline{\mathcal{M}} \xrightarrow{\sim} \underline{\mathbb{1}}(0)$ an *F-zip of type $\underline{n}$ with trivialized determinant*. For the same reasons as before, the F-zips of type $\underline{n}$ with trivialized determinant, together with isomorphisms of such pairs, form a stack over k.

Let $\Delta_1\colon \Lambda^n(k^n) \xrightarrow{\sim} k$ denote the isomorphism induced by the determinant. With the basic F-zip from (8.1) the pair $(\underline{\mathcal{M}}_1, \Delta_1)$ is then an F-zip of type $\underline{n}$ with trivialized determinant over k. As in the preceding subsection, we compare arbitrary F-zips of type $\underline{n}$ with trivialized determinant with this basic one.

The relevant linear algebraic group is now $\mathrm{SL}_{n,k}$. Clearly χ factors through $\mathrm{SL}_{n,k}$, so that we can speak of SL_n-zips of type χ. The associated parabolics of $\mathrm{SL}_{n,k}$ are now $P_\pm \cap \mathrm{SL}_{n,k}$ with $P_\pm$ as in Section 8A.

Note that $\Delta_1\colon \Lambda^n(k^n) \xrightarrow{\sim} k$ is an isomorphism in the category SL_n-$\mathtt{Rep}$ if the target k is endowed with the trivial representation. The equivalence (8.6) below can be interpreted as saying that an F-zip with trivialized determinant is a partial SL_n-zip functor containing just enough information to possess a unique extension to a full SL_n-zip functor $\mathfrak{z}\colon\ \mathrm{SL}_n\text{-}\mathtt{Rep} \to F\text{-}\mathtt{Zip}(S)$.

Construction 8.5. Let $(\underline{\mathcal{M}}, \Delta)$ with $\underline{\mathcal{M}} = (\mathcal{M}, C^\bullet, D_\bullet, \varphi_\bullet)$ be an F-zip of type $\underline{n}$ with trivialized determinant over S. Let $\underline{I} := (I, I_+, I_-, \iota)$ be the GL_n-zip associated to $\underline{\mathcal{M}}$ by Construction 8.2. Let $I' \subset I$ be the subsheaf of all isomorphisms $u\colon (k^n)_{S'} \xrightarrow{\sim} \mathcal{M}_{S'}$ for which the composite

$$\Lambda^n(k^n)_{S'} \xrightarrow{\Lambda^n u} \Lambda^n \mathcal{M}_{S'} \xrightarrow{\Delta} (k)_{S'}$$

is equal to $\Delta_{1,S'}$. One easily checks that

(a) I' is a right SL_n-torsor over S,

(b) $I'_+ := I_+ \cap I'$ is a right $P_+ \cap \mathrm{SL}_{n,k}$-torsor over S,

(c) $I'_- := I_- \cap I'$ is a right $P_-^{(q)} \cap \mathrm{SL}_{n,k}$-torsor over S.

Moreover, the highest exterior power of a filtered locally free sheaf of finite rank is canonically isomorphic to the highest exterior power of the associated graded sheaf. Thus the isomorphism of F-zips Δ amounts to a commutative diagram of isomorphisms

$$\begin{tikzcd}
\Lambda^n(\mathrm{gr}^\bullet_C \mathcal{M})^{(q)} \arrow[r, equal, "\sim"] \arrow[d, "\Lambda^n\varphi_\bullet"', "\wr"] & \Lambda^n \mathcal{M}^{(q)} \arrow[r, "\Delta^{(q)}", "\sim"'] & (k)_S^{(q)} \arrow[d, equal] \\
\Lambda^n \mathrm{gr}^D_\bullet \mathcal{M} \arrow[r, equal, "\sim"] & \Lambda^n \mathcal{M} \arrow[r, "\Delta", "\sim"'] & (k)_S.
\end{tikzcd}$$

From this one easily deduces that the isomorphism $\iota\colon I_+^{(q)}/U_+^{(q)} \xrightarrow{\sim} I_-/U_-^{(q)}$ induces an isomorphism $\iota'\colon (I')_+^{(q)}/U_+^{(q)} \xrightarrow{\sim} I'_-/U_-^{(q)}$. Together the assembled data therefore defines an SL_n-zip $\underline{I}' := (I', I'_+, I'_-, \iota')$ of type χ over S. Clearly this construction is $\mathbb{F}_q$-linearly functorial in $(\underline{\mathcal{M}}, \Delta)$ and compatible with pullback. Thus it defines a morphism of stacks

(8.6) $$\big(\!\big(F\text{-zips of type } \underline{n} \text{ with trivialized determinant}\big)\!\big) \longrightarrow \mathrm{SL}_n\text{-}\mathtt{Zip}_k^\chi\,.$$

Proposition 8.7. *This morphism is an isomorphism.*

Proof. By the remarks in Section 4G and exactness, any zip functor $\mathfrak{z}\colon\ \mathrm{SL}_n\text{-}\mathtt{Rep} \to F\text{-}\mathtt{Zip}(S)$ commutes with alternating powers; hence it sends Δ_1 to an isomorphism $\mathfrak{z}(\Delta_1)\colon\ \Lambda^n(\mathfrak{z}(k^n)) \cong \mathfrak{z}(\Lambda^n(k^n)) \xrightarrow{\sim} \mathfrak{z}(k) = \underline{\mathbb{1}}(0)$. Therefore $\mathfrak{z} \mapsto (\mathfrak{z}(k^n), \mathfrak{z}(\Delta_1))$ defines a morphism from the stack of SL_n-zips of type χ to the stack of F-zips of type $\underline{n}$ with trivialized determinant. Composed with the inverse of the morphism (7.12) we thus obtain a morphism in the other direction. The careful reader will be able to check that this is a two-sided inverse of the morphism (8.6). $\square$

8C. ***Symplectic F-zips versus*** $\mathbf{Sp}_n$***-zips.*** We call a pair $(\underline{\mathcal{M}}, E)$ consisting of an F-zip $\underline{\mathcal{M}}$ of type $\underline{n}$ over S and an admissible epimorphism $E\colon \Lambda^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathbb{1}}(0)$, whose underlying alternating pairing $\mathcal{M} \times \mathcal{M} \to (k)_S$ is nondegenerate everywhere, a *symplectic F-zip of type $\underline{n}$ over S*. For the same reasons as before these pairs, together with compatible isomorphisms, form a stack over $k := \mathbb{F}_q$. For this stack to be nonempty we assume that $n := \sum_i n_i$ is even and that $n_i = n_{-i}$ for all i.

Fix a nondegenerate alternating pairing $E_1\colon k^n \times k^n \to k$, and let $\mathrm{Sp}_{n,k} \subset \mathrm{GL}_{n,k}$ denote the associated symplectic group. Then E_1 can be viewed as an equivariant epimorphism $\Lambda^2(k^n) \twoheadrightarrow k$, where $\mathrm{Sp}_{n,k}$ acts trivially on the target k. By the assumptions on $\underline{n}$ there exists a cocharacter $\chi\colon \mathbb{G}_{m,k} \to \mathrm{Sp}_{n,k}$, unique up to conjugation, whose weights on the standard representation k^n of $\mathrm{Sp}_{n,k}$ are i with multiplicity n_i for all i. Fixing such a cocharacter, we can thus speak of Sp_n-zips of type χ over any scheme S over k.

To any symplectic F-zip $(\underline{\mathcal{M}}, E)$ of type $\underline{n}$ over S we can associate an Sp_n-zip $\underline{I}' := (I', I'_+, I'_-, \iota')$ of type χ over S. Namely, if $\underline{I} := (I, I_+, I_-, \iota)$ denotes the GL_n-zip associated to $\underline{\mathcal{M}}$ by Construction 8.2, we let $I' \subset I$ be the subsheaf of isomorphisms $(k^n)_S \xrightarrow{\sim} \mathcal{M}_S$ which are compatible with E_1 and E. From the fact that any two nondegenerate alternating pairings on the sheaf $\mathcal{O}_S^{\oplus n}$ are conjugate under $\mathrm{GL}_n(S)$ one deduces that this is an $\mathrm{Sp}_{n,k}$-torsor. Also $I'_+, I'_- \subset I'$ are the subsheaves of isomorphisms preserving the filtrations $C^\bullet$, respectively $D_\bullet$, and ι' is constructed from the isomorphism $\varphi_\bullet$ in $\underline{\mathcal{M}}$. Together we obtain a morphism of stacks

$$\big(\!\big(\text{symplectic } F\text{-zips of type } \underline{n}\big)\!\big) \longrightarrow \mathrm{Sp}_n\text{-}\mathtt{Zip}_k^\chi . \tag{8.8}$$

Conversely, for any Sp_n-zip of type χ over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation k^n and the homomorphism E_1 in $\mathrm{Sp}_{n,k}$-$\mathtt{Rep}$, obtaining a symplectic F-zip $(\mathfrak{z}(k^n), \mathfrak{z}(E_1))$ of type $\underline{n}$ over S. By showing that this construction yields a two-sided inverse of the first one proves that the morphism (8.8) is an isomorphism.

8D. ***Twisted symplectic F-zips versus*** $\mathbf{CSp}_n$***-zips.*** We call a triple $(\underline{\mathcal{M}}, \underline{\mathcal{L}}, E)$ consisting of an F-zip $\underline{\mathcal{M}}$ of type $\underline{n}$ over S, an F-zip $\underline{\mathcal{L}}$ of rank 1 over S, and an admissible epimorphism $E\colon \Lambda^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathcal{L}}$, whose underlying alternating pairing $\mathcal{M} \times \mathcal{M} \to \mathcal{L}$ is nondegenerate everywhere on S, a *twisted symplectic F-zip of type $\underline{n}$ over S*. For the same reasons as before these triples, together with compatible isomorphisms, form a stack over $k := \mathbb{F}_q$. For this stack to be nonempty we assume that $n := \sum_i n_i$ is even and that there is an integer d satisfying $n_i = n_{d-i}$ for all i. This d is then unique, and the above $\underline{\mathcal{L}}$ must be of type d.

Fix a nondegenerate alternating pairing $E_1\colon k^n \times k^n \to k$, and let $\mathrm{CSp}_{n,k}$ denote the associated group of symplectic similitudes, that is, the group of all $g \in \mathrm{GL}_{n,k}$

satisfying $E_1 \circ (g \times g) = \mu(g) \cdot E_1$ for a scalar $\mu(g)$. Then E_1 can be viewed as a $\mathrm{CSp}_{n,k}$-equivariant epimorphism $\Lambda^2(k^n) \twoheadrightarrow k$, where $\mathrm{CSp}_{n,k}$ acts on the target k through the multiplier character $\mu\colon \mathrm{CSp}_{n,k} \twoheadrightarrow \mathbb{G}_{m,k}$. By the assumptions on $\underline{n}$ there exists a cocharacter $\chi : \mathbb{G}_{m,k} \to \mathrm{CSp}_{n,k}$, unique up to conjugation, whose weights on the standard representation k^n of $\mathrm{CSp}_{n,k}$ are i with multiplicity n_i for all i. Fixing such a cocharacter, we can thus speak of CSp_n-zips of type χ over any scheme S over k.

Using the same principles as before, to any twisted symplectic F-zip $(\underline{\mathcal{M}}, \underline{\mathcal{L}}, E)$ of type $\underline{n}$ over S we can associate a CSp_n-zip $\underline{I} := (I, I_+, I_-, \iota)$ of type χ over S. In the interest of brevity we only sketch the construction: Here I is the sheaf of pairs of isomorphisms $(k^n)_{S'} \xrightarrow{\sim} \mathcal{M}_{S'}$ and $(k)_{S'} \xrightarrow{\sim} \mathcal{L}_{S'}$ that are compatible with E_1 and E. That this is a $\mathrm{CSp}_{n,k}$-torsor again results from the fact that any two nondegenerate alternating pairings on the sheaf $\mathcal{O}_{S'}^{\oplus n}$ are conjugate under $\mathrm{GL}_n(S')$. Also $I_+, I_- \subset I$ are the subsheaves of isomorphisms preserving the filtrations $C^\bullet$, respectively $D_\bullet$, and ι is constructed from the isomorphisms $\varphi_\bullet$ in $\underline{\mathcal{M}}$ and $\underline{\mathcal{L}}$. Together we obtain a morphism of stacks

$$\text{(8.9)} \qquad \big(\!\big(\text{twisted symplectic } F\text{-zips of type } \underline{n}\big)\!\big) \longrightarrow \mathrm{CSp}_n\text{-}\mathtt{Zip}_k^\chi\,.$$

Conversely, for any CSp_n-zip of type χ over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation k^n, the representation k with the multiplier character μ, and the homomorphism E_1, obtaining a twisted symplectic F-zip $(\mathfrak{z}(k^n), \mathfrak{z}(k), \mathfrak{z}(E_1))$ of type $\underline{n}$ over S. By showing that this construction yields a two-sided inverse of the first one proves that the morphism (8.9) is an isomorphism. The details in these arguments follow those in the preceding subsections and are left to the conscientious reader.

8E. *Orthogonal F-zips versus* $\mathbf{O}_n$*-zips.* To avoid the usual idiosyncrasies of symmetric bilinear forms in characteristic 2 we assume that q is odd in this subsection and the next. We call a pair $(\underline{\mathcal{M}}, B)$ consisting of an F-zip $\underline{\mathcal{M}}$ of type $\underline{n}$ over S and an admissible epimorphism $E\colon \mathrm{S}^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathbb{1}}(0)$, whose underlying symmetric pairing $\mathcal{M} \times \mathcal{M} \to (k)_S$ is nondegenerate everywhere, an *orthogonal F-zip of type $\underline{n}$ over S*. For the same reasons as before these pairs, together with compatible isomorphisms, form a stack over $k := \mathbb{F}_q$. For this stack to be nonempty we assume that $n_i = n_{-i}$ for all i.

Fix a nondegenerate split symmetric bilinear form $B_1\colon k^n \times k^n \to k$, and let $\mathrm{O}_{n,k} \subset \mathrm{GL}_{n,k}$ denote the associated orthogonal group. Then B_1 can be viewed as an equivariant epimorphism $\mathrm{S}^2(k^n) \twoheadrightarrow k$, where $\mathrm{O}_{n,k}$ acts trivially on the target k. Note that $\mathrm{O}_{n,k}$ has two connected components and that its identity component is a split special orthogonal group $\mathrm{SO}_{n,k}$. By the assumptions on $\underline{n}$ there exists a cocharacter $\chi : \mathbb{G}_{m,k} \to \mathrm{O}_{n,k}$, unique up to conjugation, whose weights on the standard representation k^n are i with multiplicity n_i for all i. We fix such a cocharacter and set $\hat{L} := \mathrm{Cent}_{\mathrm{O}_{n,k}}(\chi)$ and $\Theta := \pi_0(\hat{L}) \subset \pi_0(\mathrm{O}_{n,k})$. A quick calculation shows that $\hat{L} \cong \mathrm{O}_{n_0,k} \times \prod_{i>0} \mathrm{GL}_{n_i,k}$; hence Θ is trivial if $n_0 = 0$, and equal to $\pi_0(\mathrm{O}_{n,k})$ if $n_0 > 0$. According to Definition 3.1 we can speak of O_n-zips of type (χ, Θ) over any scheme S over k.

The definition of $\hat{L}$ implies that the associated subgroups $\hat{P}_\pm$ from Section 3B are precisely the stabilizers of the descending and ascending filtrations of k^n induced by χ. Also, observe that any two nondegenerate symmetric pairings on the sheaf $\mathbb{O}_S^{\oplus n}$ are fpqc-locally conjugate under GL_n. Using these facts and the same construction as in Section 8C, to any orthogonal F-zip $(\underline{\mathcal{M}}, B)$ of type $\underline{n}$ over S we can associate an O_n-zip of type (χ, Θ) over S, obtaining a morphism of stacks

$$\big(\!\big(\text{orthogonal } F\text{-zips of type } \underline{n}\big)\!\big) \longrightarrow \mathrm{O}_n\text{-}\mathtt{Zip}_k^{\chi,\Theta}. \tag{8.10}$$

Conversely, for any O_n-zip of type (χ, Θ) over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation k^n and the homomorphism B_1 in $\mathrm{O}_{n,k}$-$\mathtt{Rep}$, obtaining an orthogonal F-zip $(\mathfrak{z}(k^n), \mathfrak{z}(B_1))$ of type $\underline{n}$ over S. This construction yields a two-sided inverse of the first and thereby proves that the morphism (8.10) is an isomorphism.

8F. ***Twisted orthogonal F-zips versus*** $\mathbf{CO}_n$***-zips.*** Again we assume that q is odd. We call a triple $(\underline{\mathcal{M}}, \underline{\mathcal{L}}, B)$ consisting of an F-zip $\underline{\mathcal{M}}$ of type $\underline{n}$ over S, an F-zip $\underline{\mathcal{L}}$ of rank 1 over S, and an admissible epimorphism $B\colon \mathrm{S}^2\underline{\mathcal{M}} \twoheadrightarrow \underline{\mathcal{L}}$, whose underlying symmetric pairing $\mathcal{M} \times \mathcal{M} \to \mathcal{L}$ is nondegenerate everywhere on S, a *twisted orthogonal F-zip of type $\underline{n}$ over S*. For the same reasons as before these triples, together with compatible isomorphisms, form a stack over $k := \mathbb{F}_q$. For this stack to be nonempty we assume that there is an integer d satisfying $n_i = n_{d-i}$ for all i. This d is then unique, and the above $\underline{\mathcal{L}}$ must be of type d.

Fix a nondegenerate split symmetric bilinear form $B_1\colon k^n \times k^n \to k$. Let $\mathrm{CO}_{n,k}$ denote the associated group of orthogonal similitudes, that is, the group of all $g \in \mathrm{GL}_{n,k}$ satisfying $B_1 \circ (g \times g) = \mu(g) \cdot B_1$ for a scalar $\mu(g)$. Then B_1 can be viewed as a $\mathrm{CO}_{n,k}$-equivariant epimorphism $\mathrm{S}^2(k^n) \twoheadrightarrow k$, where $\mathrm{CO}_{n,k}$ acts on the target k through the character $\mu\colon \mathrm{CO}_{n,k} \twoheadrightarrow \mathbb{G}_{m,k}$. If n is odd, then $\mathrm{CO}_{n,k}$ is connected with a root system of type $B_{(n-1)/2}$ and is therefore split. If n is even, then $\mathrm{CO}_{n,k}$ has two connected components and a root system of type $D_{n/2}$. In both cases the identity component of $\mathrm{CO}_{n,k}$ is split, because B_1 is split. Thus there exists a cocharacter $\chi : \mathbb{G}_{m,k} \to \mathrm{CO}_{n,k}$, unique up to conjugation, whose weights on the standard representation k^n of $\mathrm{CO}_{n,k}$ are i with multiplicity n_i for all i. We fix such a cocharacter and set $\hat{L} := \mathrm{Cent}_{\mathrm{CO}_{n,k}}(\chi)$ and $\Theta := \pi_0(\hat{L}) \subset \pi_0(\mathrm{CO}_{n,k})$. A quick calculation shows that $\hat{L} \cong \mathrm{CO}_{n_{d/2},k} \times \prod_{i>d/2} \mathrm{GL}_{n_i,k}$ if d is even and $n_{d/2} > 0$, and $\hat{L} \cong \mathbb{G}_{m,k} \times \prod_{i>d/2} \mathrm{GL}_{n_i,k}$ otherwise. Thus Θ is trivial unless d is even and $n_{d/2}$

is even and positive, in which case $\Theta = \pi_0(\mathrm{CO}_{n,k})$ of order 2. In either case we can speak of CO_n-zips of type (χ, Θ) over any scheme S over k.

In the same way as in Section 8D, to any twisted orthogonal F-zip $(\underline{\mathcal{M}}, \underline{\mathcal{L}}, B)$ of type $\underline{n}$ over S we can associate a CO_n-zip of type (χ, Θ) over S, obtaining a morphism of stacks

$$\text{(8.11)} \qquad \big(\!\big(\text{twisted orthogonal } F\text{-zips of type } \underline{n}\big)\!\big) \longrightarrow \mathrm{CO}_n\text{-}\mathtt{Zip}_k^{\chi,\Theta}.$$

Conversely, for any CO_n-zip of type (χ, Θ) over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation k^n, the representation k with the multiplier character μ, and the homomorphism B_1, obtaining a twisted orthogonal F-zip $(\mathfrak{z}(k^n), \mathfrak{z}(k), \mathfrak{z}(B_1))$ of type $\underline{n}$ over S. By showing that this construction yields a two-sided inverse of the first one proves that the morphism (8.11) is an isomorphism.

8G. *Unitary F-zips versus* $\mathbf{U}_n$*-zips.* Let $\mathbb{F}_{q^2}$ denote a fixed quadratic extension of $\mathbb{F}_q$, and let σ denote its nontrivial automorphism $x \mapsto x^q$ over $\mathbb{F}_q$. Let S be a scheme over $\mathbb{F}_q$. We call a triple $(\underline{\mathcal{M}}, \rho, H)$ consisting of an F-zip $\underline{\mathcal{M}}$ over S, an $\mathbb{F}_q$-algebra homomorphism $\rho\colon \mathbb{F}_{q^2} \to \mathrm{End}(\underline{\mathcal{M}})$, and an admissible epimorphism $H\colon \underline{\mathcal{M}} \otimes \underline{\mathcal{M}} \to \mathbb{F}_{q^2} \otimes_{\mathbb{F}_q} \underline{\mathbb{1}}(0)$, which satisfies

(a) $H \circ (\rho(\alpha) \otimes \rho(\beta)) = (\alpha^q \beta \otimes 1) \circ H$ for all $\alpha, \beta \in \mathbb{F}_{q^2}$, and

(b) $H(m_2, m_1) = (\sigma \otimes 1) \circ H(m_1, m_2)$ for all local sections m_1, m_2 of $\mathcal{M}$,

and whose hermitian pairing on the underlying sheaf $\mathcal{M} \times \mathcal{M} \to \mathbb{F}_{q^2} \otimes_{\mathbb{F}_q} \mathcal{O}_S$ is nondegenerate everywhere, a *unitary F-zip over S*. To classify such objects we use base change fro $\mathbb{F}_q$ to $\mathbb{F}_{q^2}$:

Let $\tilde{S}$ be a scheme over $\mathbb{F}_{q^2}$ and $(\underline{\tilde{\mathcal{M}}}, \tilde{\rho}, \tilde{H})$ a unitary F-zip over $\tilde{S}$. Then we have a unique decomposition $\tilde{\mathcal{M}} = \tilde{\mathcal{N}} \oplus \tilde{\mathcal{N}}'$, where $\tilde{\rho}(\alpha)$ acts on $\tilde{\mathcal{N}}$ through multiplication by α and on $\tilde{\mathcal{N}}'$ through multiplication by α^q, and the hermitian pairing $\tilde{H}$ amounts to an isomorphism $\tilde{\mathcal{N}}' \xrightarrow{\sim} \tilde{\mathcal{N}}^\vee$. Working out the rest of the data we find that giving a unitary F-zip over S is equivalent to giving a quadruple $(\tilde{\mathcal{N}}, C^\bullet, D_\bullet, \psi_\bullet)$ consisting of a locally free sheaf of $\mathcal{O}_S$-modules of finite rank $\tilde{\mathcal{N}}$ on S, a descending filtration $C^\bullet$ and an ascending filtration $D_\bullet$ of $\tilde{\mathcal{N}}$, and an $\mathcal{O}_S$-linear isomorphism $\psi_i\colon (\mathrm{gr}^i_C \tilde{\mathcal{N}})^{(q)} \xrightarrow{\sim} (\mathrm{gr}^D_{-i} \tilde{\mathcal{N}})^\vee$ for every $i \in \mathbb{Z}$. We call $(\underline{\tilde{\mathcal{M}}}, \tilde{\rho}, \tilde{H})$ *of type* $\underline{n}$ if the associated $\mathrm{gr}^i_C \tilde{\mathcal{N}}$ is locally free of constant rank n_i for all i. For the same reasons as before the unitary F-zips of type $\underline{n}$, together with compatible isomorphisms, form a stack over $\mathbb{F}_{q^2}$. is no further condition on $\underline{n}$ in this case.

Let as above S be a scheme over $\mathbb{F}_q$ and $(\underline{\mathcal{M}}, \rho, H)$ a unitary F-zip over S. We have $\mathcal{M} = \mathrm{pr}_{2*} \widetilde{\mathcal{M}}$ for a locally free sheaf of $\mathcal{O}_{\tilde{S}}$-modules $\widetilde{\mathcal{M}}$ on $\tilde{S} := \mathrm{Spec}\, \mathbb{F}_{q^2} \times_{\mathrm{Spec}\, \mathbb{F}_q} S$, such that the action ρ of $\mathbb{F}_{q^2}$ is induced from the first factor. The $\mathbb{F}_{q^2}$-invariant filtration $C^\bullet$ on $\mathcal{M}$ then comes from a filtration of $\widetilde{\mathcal{M}}$. In this case we call a unitary F-zip *of type* $\underline{n}$ if the associated $\mathrm{gr}^i_C \widetilde{\mathcal{M}}$ is locally free of constant rank n_i for all i.

Since the hermitian pairing H induces isomorphisms $(\sigma\times\mathrm{id})^*\mathrm{gr}^i_C\widetilde{\mathcal{M}} \xrightarrow{\sim} (\mathrm{gr}^{-i}_C\widetilde{\mathcal{M}})^\vee$, this condition can be satisfied nontrivially only if $n_i = n_{-i}$ for all i. Under this assumption the unitary F-zips of type $\underline{n}$, together with compatible isomorphisms, form a stack over $\mathbb{F}_q$. Moreover, a unitary F-zip is of type $\underline{n}$ in this sense if and only if its pullback to $\tilde{S}$ is of type $\underline{n}$ in the previous sense; hence the stack over $\mathbb{F}_{q^2}$ described before is just the base change of the present stack over $\mathbb{F}_q$.

In summary, set $k := \mathbb{F}_q$ if $n_i = n_{-i}$ for all i, respectively $k := \mathbb{F}_{q^2}$ if not; the unitary F-zips of type $\underline{n}$ then form a natural stack over k.

Fix a nondegenerate σ-hermitian form $H_1\colon \mathbb{F}^n_{q^2}\times\mathbb{F}^n_{q^2}\to\mathbb{F}_{q^2}$, and let $\mathrm{U}_{n,\mathbb{F}_q}\subset \mathcal{R}_{\mathbb{F}_{q^2}/\mathbb{F}_q}\mathrm{GL}_{n,\mathbb{F}_{q^2}}$ denote the associated unitary group. The assumptions on $\underline{n}$ and k imply that there exists a cocharacter $\chi\colon\mathbb{G}_{m,k}\to\mathrm{U}_{n,k}$, unique up to conjugation, whose weights on the standard representation $\mathbb{F}^n_{q^2}$ of $\mathrm{U}_{n,\mathbb{F}_{q^2}}$ are i with multiplicity n_i for all i. Fixing such a cocharacter, we can thus speak of U_n-zips of type χ over any scheme S over k.

Also, set $M_1 := \mathbb{F}^n_{q^2}\otimes_{\mathbb{F}_q} k$ with the descending filtration $C^\bullet$ associated to χ and the ascending filtration $D_\bullet$ associated to the Frobenius twist $\chi^{(q)}$, so that there are natural $\mathbb{F}_{q^2}\otimes_{\mathbb{F}_q} k$-linear isomorphisms $\varphi_i\colon(\mathrm{gr}^i_C M_1)^{(q)}\xrightarrow{\sim}\mathrm{gr}^D_i M_1$ for all $i\in\mathbb{Z}$. Then $\underline{\mathcal{M}}_1 := (M_1, C^\bullet, D_\bullet, \varphi_\bullet)$ together with the evident action of $\mathbb{F}_{q^2}$ and the pairing H_1 is a unitary F-zip of type $\underline{n}$ over k.

Using the same principles as in the preceding subsections, to any unitary F-zip $(\underline{\mathcal{M}}, \rho, H)$ of type $\underline{n}$ over S we can now associate a U_n-zip $\underline{I} := (I, I_+, I_-, \iota)$ of type χ over S. Here I is the sheaf of all $\mathbb{F}_{q^2}\otimes_{\mathbb{F}_q}\mathcal{O}_{S'}$-linear isomorphisms $M_{1,S'}\xrightarrow{\sim}\mathcal{M}_{S'}$ which are compatible with H_1 and H, for all morphisms $S'\to S$, and $I_\pm$ are the subsheaves of isomorphisms which are in addition compatible with the filtrations $C^\bullet$, respectively $D_\bullet$, and ι is obtained from the graded isomorphisms $\varphi_\bullet$. Together this yields a morphism of stacks

$$\big(\!\big(\text{unitary } F\text{-zips of type } \underline{n}\big)\!\big) \longrightarrow \mathrm{U}_n\text{-}\mathtt{Zip}^\chi_k\,. \tag{8.12}$$

Conversely, for any U_n-zip of type χ over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation $\mathbb{F}^n_{q^2}$, the obvious homomorphism $\mathbb{F}_{q^2}\to\mathrm{End}_{\mathrm{U}_{n,\mathbb{F}_q}}(\mathbb{F}^n_{q^2})$, and the hermitian pairing H_1 (all of which are objects and morphisms in U_n-$\mathtt{Rep}$), obtaining a unitary F-zip of type $\underline{n}$ over S. By showing that this construction yields a two-sided inverse of the first one proves that the morphism (8.12) is an isomorphism.

8H. ***Twisted unitary F-zips versus*** $\mathbf{CU}_n$***-zips.*** Again let $\mathbb{F}_{q^2}$ denote a fixed quadratic extension of $\mathbb{F}_q$, and let σ denote its nontrivial automorphism $x\mapsto x^q$ over $\mathbb{F}_q$. We call a quadruple $(\underline{\mathcal{M}}, \rho, \underline{\mathcal{L}}, H)$ consisting of an F-zip $\underline{\mathcal{M}}$ over S, an $\mathbb{F}_q$-algebra homomorphism $\rho\colon\mathbb{F}_{q^2}\to\mathrm{End}(\underline{\mathcal{M}})$, an F-zip $\underline{\mathcal{L}}$ of rank 1 over S, and an admissible epimorphism $H\colon\underline{\mathcal{M}}\otimes\underline{\mathcal{M}}\to\mathbb{F}_{q^2}\otimes_{\mathbb{F}_q}\underline{\mathcal{L}}$, which satisfies the same conditions (a) and (b) as in Section 8G and whose hermitian pairing on the underlying sheaf is nondegenerate everywhere, a *twisted unitary F-zip over S*.

If S is a scheme over $\mathbb{F}_{q^2}$, for any twisted unitary F-zip over S there is a unique decomposition $\mathcal{M} = \mathcal{N} \oplus \mathcal{N}'$, where $\rho(\alpha)$ acts on $\mathcal{N}$ through multiplication by α and on $\mathcal{N}'$ through multiplication by α^q, and it is compatible with the filtration $C^\bullet$. In this case we call a twisted unitary F-zip over S *of type* $(\underline{n}, d)$ if the associated $\mathrm{gr}^i_C \mathcal{N}$ is locally free of constant rank n_i for all i and $\underline{\mathcal{L}}$ is of type d. For the same reasons as before the twisted unitary F-zips of type $(\underline{n}, d)$, together with compatible isomorphisms, form a stack over $\mathbb{F}_{q^2}$. There is no further condition on $(\underline{n}, d)$ in this case.

If S is only a scheme over $\mathbb{F}_q$, we still have $\mathcal{M} = \mathrm{pr}_{2*} \widetilde{\mathcal{M}}$ for a locally free sheaf of $\mathcal{O}_{\tilde{S}}$-modules $\widetilde{\mathcal{M}}$ on $\tilde{S} := \operatorname{Spec} \mathbb{F}_{q^2} \times_{\operatorname{Spec} \mathbb{F}_q} S$, such that the action ρ of $\mathbb{F}_{q^2}$ is induced by the first factor. In this case we call a twisted unitary F-zip *of type* $(\underline{n}, d)$ if the associated $\mathrm{gr}^i_C \widetilde{\mathcal{M}}$ is locally free of constant rank n_i for all i and $\underline{\mathcal{L}}$ is of type d. Since the hermitian pairing H must induce isomorphisms $(\sigma \times \mathrm{id})^* \mathrm{gr}^i_C \widetilde{\mathcal{M}} \xrightarrow{\sim} (\mathrm{gr}^{d-i}_C \widetilde{\mathcal{M}})^\vee \otimes \mathrm{pr}_2^* \mathcal{L}$, this condition can be satisfied nontrivially only if $n_i = n_{d-i}$ for all i. Under this assumption the twisted unitary F-zips of type $(\underline{n}, d)$, together with compatible isomorphisms, form a stack over $\mathbb{F}_q$. Moreover, a twisted unitary F-zip is of type $(\underline{n}, d)$ in this sense if and only if its pullback to $\tilde{S}$ is of type $(\underline{n}, d)$ in the previous sense; hence the stack over $\mathbb{F}_{q^2}$ described before is just the base change of the present stack over $\mathbb{F}_q$.

In summary, set $k := \mathbb{F}_q$ if $n_i = n_{d-i}$ for all i, respectively $k := \mathbb{F}_{q^2}$ if not; the twisted unitary F-zips of type $(\underline{n}, d)$ then form a natural stack over k.

Fix a nondegenerate σ-hermitian form $H_1 : \mathbb{F}^n_{q^2} \times \mathbb{F}^n_{q^2} \to \mathbb{F}_{q^2}$, and let $\mathrm{CU}_{n,\mathbb{F}_q} \subset \mathcal{R}_{\mathbb{F}_{q^2}/\mathbb{F}_q} \mathrm{GL}_{n,\mathbb{F}_{q^2}}$ denote the associated group of unitary similitudes, that is, of sections g that satisfy $H_1 \circ (g \times g) = \mu(g) \cdot H_1$ for a scalar $\mu(g)$ in $\mathbb{G}_{m,\mathbb{F}_q}$. The assumptions on $(\underline{n}, d)$ and k imply that there exists a cocharacter $\chi : \mathbb{G}_{m,k} \to \mathrm{CU}_{n,k}$, unique up to conjugation, whose weights on the standard representation $\mathbb{F}^n_{q^2}$ of $\mathrm{U}_{n,\mathbb{F}_{q^2}}$ are i with multiplicity n_i for all i, and whose weight under the multiplier character μ is d. Fixing such a cocharacter, we can thus speak of CU_n-zips of type χ over any scheme S over k.

By the same procedure as before, to any twisted unitary F-zip $(\underline{\mathcal{M}}, \rho, H)$ of type $(\underline{n}, d)$ over S we can associate a CU_n-zip of type χ over S, obtaining a morphism of stacks

$$\text{(8.13)} \qquad \big(\!\big(\text{twisted unitary } F\text{-zips of type } (\underline{n}, d)\big)\!\big) \longrightarrow \mathrm{CU}_n\text{-}\mathtt{Zip}^\chi_k .$$

Conversely, for any CU_n-zip of type χ over S we evaluate the associated zip functor $\mathfrak{z}$ on the standard representation $\mathbb{F}^n_{q^2}$, the obvious homomorphism $\mathbb{F}_{q^2} \to \mathrm{End}_{\mathrm{CU}_{n,\mathbb{F}_q}}(\mathbb{F}^n_{q^2})$, the multiplier representation on $\mathbb{F}_q$, and the hermitian pairing H_1 (all of which are objects and morphisms in CU_n-$\mathtt{Rep}$), obtaining a twisted unitary

F-zip of type $(\underline{n}, d)$ over S. By showing that this construction yields a two-sided inverse of the first one proves that the morphism (8.13) is an isomorphism.

8I. *Other groups.* For each of the groups $\hat{G}$ above, we have identified a finite subcategory $\mathscr{C}$ of $\hat{G}$-$\mathtt{Rep}$ and have shown that any suitable functor $\mathscr{C} \to F$-$\mathtt{Zip}(S)$ extends to a $\hat{G}$-zip functor $\hat{G}$-$\mathtt{Rep} \to F$-$\mathtt{Zip}(S)$. Surely it must be possible to apply the same principle to an arbitrary reductive linear algebraic group $\hat{G}$ over k. However, identifying a suitable subcategory $\mathscr{C}$ becomes tiresome very quickly.

For instance, it should be possible to describe SO_n-zip functors in terms of triples $(\underline{\mathcal{M}}, B, \Delta)$ consisting of an F-zip $\underline{\mathcal{M}}$ of rank n, an everywhere nondegenerate symmetric pairing B on $\underline{\mathcal{M}}$, and a trivialization Δ of the highest exterior power of $\underline{\mathcal{M}}$. However, to guarantee the extendability to $\mathrm{SO}_{n,k}$-$\mathtt{Rep}$ one must also impose a certain relation between B and Δ which is more complicated to describe. Moreover, the identification of the type of an SO_n-zip functor might require some extra data, because SO_n may possess nonconjugate cocharacters which are conjugate under GL_n. A similar situation arises for the group $\mathrm{SU}_{n,k}$.

9. Applications

9A. *Zip strata attached to smooth proper morphism with degenerating Hodge spectral sequence.* In the following we give a generalization of a construction from [Moonen and Wedhorn 2004]. Let S be a scheme over $\mathbb{F}_p$, let $\mathscr{X}$ be a Deligne–Mumford stack and let $f : \mathscr{X} \to S$ be a morphism of finite type. For every étale morphism $U \to \mathscr{X}$, where U is a scheme, we set $\Omega^\bullet_{\mathscr{X}/S}|U := \Omega^\bullet_{U/S}$, where $\Omega^\bullet_{U/S}$ is the de Rham complex of U over S. As the formation of the de Rham complex $\Omega^\bullet_{U/S}$ commutes with étale localization on U, this defines a complex of quasicoherent sheaves of $\mathscr{O}_{\mathscr{X}}$-modules of finite type on the étale site on $\mathscr{X}$ whose differentials are $f^{-1}\mathscr{O}_S$-linear.

Attached to the naive and the canonical filtration of the de Rham complex $\Omega^\bullet_{\mathscr{X}/S}$ we obtain two spectral sequences converging to the de Rham cohomology $H^\bullet_{\mathrm{DR}}(\mathscr{X}/S) = \mathbb{R}^\bullet f_*(\Omega^\bullet_{\mathscr{X}/S})$, namely the Hodge–de Rham spectral sequence

$$ {}_HE_1^{ab} = R^b f_*(\Omega^a_{\mathscr{X}/S}) \Longrightarrow H^{a+b}_{\mathrm{DR}}(\mathscr{X}/S)$$

and the conjugate spectral sequence

$$ {}_{\mathrm{conj}}E_2^{ab} = R^a f_*(\mathscr{H}^b(\Omega^\bullet_{\mathscr{X}/S})) \Longrightarrow H^{a+b}_{\mathrm{DR}}(\mathscr{X}/S).$$

In particular these spectral sequences endow $H^d_{\mathrm{DR}}(\mathscr{X}/S)$ for $d \geq 0$ with two descending filtrations $({}_HF^i H^d_{\mathrm{DR}}(\mathscr{X}/S))_{i\in\mathbb{Z}}$ and $({}_{\mathrm{conj}}F^i H^d_{\mathrm{DR}}(\mathscr{X}/S))_{i\in\mathbb{Z}}$ by sheaves of $\mathscr{O}_S$-submodules which are called the *Hodge filtration* and the *conjugate filtration*.

We denote by $F\colon \mathscr{X} \to \mathscr{X}^{(p)}$ the relative Frobenius of $\mathscr{X}$ over S. For étale morphisms $g\colon U \to \mathscr{X}$ the diagram

$$\begin{tikzcd} U \arrow[r, "\mathrm{Frob}"] \arrow[d, "g"'] & U \arrow[d, "g"] \\ \mathscr{X} \arrow[r, "\mathrm{Frob}"] & \mathscr{X}, \end{tikzcd}$$

where the horizontal morphisms are the absolute Frobenii, is cartesian. This shows that the formation of the relative Frobenius also commutes with étale base change. In particular, F is representable and finite.

If f is smooth, there is a unique isomorphism of graded sheaves of $\mathcal{O}_{\mathscr{X}^{(p)}}$-modules

$$\mathscr{C}^{-1}\colon \bigoplus_{b\geq 0} \Omega^b_{\mathscr{X}^{(p)}/S} \xrightarrow{\sim} \bigoplus_{b\geq 0} \mathscr{H}^b\big(F_*(\Omega^\bullet_{\mathscr{X}/S})\big),$$

the (inverse) Cartier isomorphism, which satisfies

$$\begin{aligned} \mathscr{C}^{-1}(1) &= 1 \\ \mathscr{C}^{-1}(d\sigma^{-1}(x)) &= \text{class of } x^{p-1}dx \\ \mathscr{C}^{-1}(\omega\wedge\omega') &= \mathscr{C}^{-1}(\omega)\wedge\mathscr{C}^{-1}(\omega'). \end{aligned}$$

To see this, we remark that because of the uniqueness assertion one may work locally for étale topology on $\mathscr{X}^{(p)}$. As the formation of differentials, of $\mathscr{H}^i(\ )$, and of Frobenius is compatible with étale base change $U \to \mathscr{X}$, the unique existence of $\mathscr{C}^{-1}$ for Deligne–Mumford stacks follows from the analogous result for smooth morphisms of schemes.

From now on we assume that f is smooth and proper. We fix an integer $d \geq 0$. We assume that f and d satisfy the following two conditions.

(D1) The sheaves of $\mathcal{O}_S$-modules $R^b f_*(\Omega^a_{\mathscr{X}/S})$ are locally free of finite rank for all $a, b \geq 0$ with $a + b \leq d$.

(D2) The Hodge–de Rham spectral sequence ${}_HE_1^{ab} = R^b f_*(\Omega^a_{\mathscr{X}/S}) \Longrightarrow H^{a+b}_{\mathrm{DR}}(\mathscr{X}/S)$ degenerates for $a+b \leq d$ (that is, for all $r \geq 1$ and a, b with $a+b \leq d$ the differentials from and to ${}_HE_r^{ab}$ vanish).

Then the formation of the Hodge–de Rham spectral sequence for $a+b\leq d$ commutes with base change $S' \to S$, and $H^e_{\mathrm{DR}}(\mathscr{X}/S)$ is locally free of finite rank for $e \leq d$.

Now one has $R^a f^{(p)}_* \circ F_* = R^a f_*$ because F is affine. Hence applying the functor $R^a f^{(p)}_*$ to the Cartier isomorphism we obtain an isomorphism

$$R^a f^{(p)}_*(\Omega^b_{\mathscr{X}^{(p)}/S}) \xrightarrow{\sim} R^a f_*\big(\mathscr{H}^b(\Omega^\bullet_{\mathscr{X}/S})\big).$$

Because of Condition (D1) the $\mathcal{O}_S$-modules $R^a f_*(\Omega^b_{X/S})$ are flat for all $a, b \geq 0$ with $a + b \leq d$ and we obtain isomorphisms

$$\text{(9.1)} \qquad \varphi^{ab}\colon R^a f_*(\Omega^b_{\mathcal{X}/S})^{(p)} = ({}_H E^{ba})^{(p)} \xrightarrow{\sim} {}_{\mathrm{conj}}E_2^{ab} = R^a f_*\big(\mathcal{H}^b(\Omega^\bullet_{\mathcal{X}/S})\big).$$

This implies that the conjugate spectral sequence also degenerates for $a + b \leq d$ and that its formation commutes with arbitrary base change for $a + b \leq d$ (see [Katz 1972] 2 if $\mathcal{X}$ is a scheme; the arguments for Deligne–Mumford stacks $\mathcal{X}$ are verbatim the same).

Remark 9.2. We list some examples of morphisms f and integers d that satisfy conditions (D1) and (D2).

(a) By [Moonen and Wedhorn 2004], conditions (D1) and (D2) are satisfied for all d in case $\mathcal{X}$ is a smooth proper relative curve over S, in case $\mathcal{X}$ is an abelian scheme over S, in case $\mathcal{X}$ is a smooth toric scheme over S and in case $\mathcal{X}$ is a relative K3-surface over S.

(b) Conditions (D1) and (D2) are satisfied for all $d \leq p - 1$ if there exists a flat scheme $\tilde{S}$ over $\mathbb{Z}/p^2\mathbb{Z}$ satisfying $\tilde{S} \otimes_{\mathbb{Z}/p^2\mathbb{Z}} \mathbb{F}_p \cong S$ and a smooth proper lift of $\mathcal{X}$ to $\tilde{S}$.

This is shown in [Deligne and Illusie 1987] if $\mathcal{X}$ is a scheme, and the proof carries over verbatim to the case of Deligne–Mumford stacks because the formation of the de Rham complex and the relative Frobenius is compatible with pull back via étale morphisms $X' \to X$ (see also [Satriano 2012, Theorem 3.7] for a generalization to tame Artin stacks; note that Satriano formulates only the case where $S = \operatorname{Spec} k$ for a perfect field k but combining his proof with the proof over a general base scheme in [Deligne and Illusie 1987] also shows the general case).

(c) Let $S = \operatorname{Spec} k$ for a perfect field k. Let X be a smooth proper scheme over k and let $D \in \operatorname{Div}(X) \otimes \mathbb{Q}$ be a $\mathbb{Q}$-divisor whose support has only normal crossings and such that exists an integer b prime to $\operatorname{char}(k)$ such that bD is integral. Then in [Matsuki and Olsson 2005] there is attached a morphism $\mathcal{X} \to X$, where $\mathcal{X}$ is a smooth proper Deligne–Mumford stack which is the "minimal covering" of X such that D becomes integral. Moreover the authors show that each lift of (X, D) over $W_2(k)$ yields also a smooth proper lift of $\mathcal{X}$ making it possible to apply (b). We refer to [loc. cit.,Theorem 4.1] for details.

We associate to f and d an F-zip $(\mathcal{M}, C^\bullet, D_\bullet, \varphi_\bullet)$ over S as follows: Set $\mathcal{M} = H^d_{\mathrm{DR}}(\mathcal{X}/S)$. Let $C^\bullet$ be the Hodge filtration on $\mathcal{M}$, and define the filtration $D_\bullet$ by $D_i = {}_{\mathrm{conj}}F^{d-i} H^d_{\mathrm{DR}}(\mathcal{X}/S)$. As the formation of both spectral sequences commutes with arbitrary base change, $C^\bullet$ and $D_\bullet$ are filtrations by locally direct summands, that is they are filtrations in the sense of Sections 4C and 4D. The assumption of the degeneracy of the Hodge spectral sequence and hence of the conjugate spectral sequence shows that one has functorial isomorphisms

$$\begin{aligned} \mathrm{gr}^i_C H^d_{\mathrm{DR}}(\mathscr{X}/S) &\cong R^{d-i} f_*(\Omega^i_{\mathscr{X}/S}) \\ \mathrm{gr}^D_i H^d_{\mathrm{DR}}(\mathscr{X}/S) &\cong R^{d-i} f_*(\mathscr{H}^i(\Omega^\bullet_{\mathscr{X}/S})) \end{aligned} \tag{9.3}$$

Finally, let

$$\varphi_i := \varphi^{d-i,i} \colon (\mathrm{gr}^i_C)^{(p)} = R^{d-i} f_*(\Omega^i_{\mathscr{X}/S})^{(p)} \xrightarrow{\sim} \mathrm{gr}^D_i = R^{d-i} f_*(\mathscr{H}^i(\Omega^\bullet_{\mathscr{X}/S})),$$

where $\varphi^{d-i,i}$ is the isomorphism defined in (9.1). We denote this F-zip by $\underline{H}^d_{\mathrm{DR}}(\mathscr{X}/S)$. For $i \in \mathbb{Z}$ set

$$n_i := \begin{cases} h^{d-i,i} = \mathrm{rk}(R^{d-i} f_*(\Omega^i_{X/S})), & \text{for } 0 \le i \le d; \\ 0, & \text{otherwise.} \end{cases}$$

This is a locally constant function on S. If n_i is constant for all i, which is automatic if S is connected, then $\underline{n} = (n_i)$ is the type of the F-zip $\underline{H}^d_{\mathrm{DR}}(\mathscr{X}/S)$. Thus for $n := \sum_i n_i$ the isomorphism (8.3) yields a GL_n-zip $\underline{I}$ of type χ over S, where χ is the cocharacter of GL_n associated to $\underline{n}$ as in Section 8A. By (3.28) we obtain a decomposition into locally closed subschemes

$$S = \bigcup_{w \in {}^I W} S^w_{\underline{I}}$$

indexed by

$${}^I W = \Big\{ w \in S_h : \forall i \in \mathbb{Z} : w^{-1}\Big(\sum_{j<i} n_j + 1\Big) < \cdots < w^{-1}\Big(\sum_{j<i} n_j + n_i\Big)\Big\}.$$

Let $\preceq$ be the partial order on ${}^I W$ given by (3.16). By the inclusion (3.29) and Proposition 3.30 one has

$$\overline{S^w_{\underline{I}}} \subseteq \bigcup_{w' \preceq w} S^{w'}_{\underline{I}}$$

with equality if the classifying morphism $S \to \mathrm{GL}_n\text{-}\mathtt{Zip}^\chi_{\mathbb{F}_p}$ of the GL_n-zip $\underline{I}$ is generizing.

9B. ***Cup product and duality.*** The cup product in de Rham cohomology yields a bilinear map of F-zips, as follows. As the cup product has not yet been worked out for Deligne–Mumford stacks (as far as we know), we restrict ourself to the case that $f \colon X \to S$ is a smooth and proper morphisms of *schemes* over $\mathbb{F}_p$ satisfying conditions (D1) and (D2) for all d. Using the Künneth formula for hypercohomology of complexes ([Grothendieck 1963, Section 6.7.8], applicable because $\Omega^a_{X/S}$ and $H^d_{\mathrm{DR}}(X/S)$ are S-flat for all a and d) one sees that the wedge product

$$\Omega^\bullet_{X/S} \otimes_{f^{-1}\mathscr{O}_S} \Omega^\bullet_{X/S} \to \Omega^\bullet_{X/S} \tag{9.4}$$

induces a homomorphism of locally free graded $\mathscr{O}_S$-modules of finite rank

$$\cup \colon H^\bullet_{\mathrm{DR}}(X/S) \otimes_{\mathscr{O}_S} H^\bullet_{\mathrm{DR}}(X/S) \longrightarrow H^\bullet_{\mathrm{DR}}(X/S), \tag{9.5}$$

the cup product. This makes $H^\bullet(X/S)$ into a graded anticommutative $\mathcal{O}_S$-algebra. It is easily checked that the wedge product sends the tensor product of the naive (resp. canonical) filtrations to the naive (resp. canonical) filtration. Thus by functoriality of the spectral sequence associated to a filtered complex the cup product induces for all $d, e \geq 0$ a morphism of filtered locally free modules of finite rank

$$\cup\colon H^d_{\mathrm{DR}}(X/S) \otimes_{\mathcal{O}_S} H^e_{\mathrm{DR}}(X/S) \longrightarrow H^{d+e}_{\mathrm{DR}}(X/S). \tag{9.6}$$

In particular we obtain induced pairings on the associated graded pieces. Moreover, using the defining properties of the Cartier isomorphism, one sees that there is a commutative diagram

$$\begin{tikzcd} \mathrm{gr}^i_C H^d_{\mathrm{DR}}(X/S)^{(p)} \otimes \mathrm{gr}^j_C H^e_{\mathrm{DR}}(X/S)^{(p)} \arrow[r] \arrow[d, "\varphi^{d-i,i}\otimes\varphi^{e-j,j}"'] & \mathrm{gr}^{i+j}_C H^{d+e}_{\mathrm{DR}}(X/S)^{(p)} \arrow[d, "\varphi^{d+e-i-j,i+j}"] \\ \mathrm{gr}^D_i H^d_{\mathrm{DR}}(X/S) \otimes \mathrm{gr}^D_j H^e_{\mathrm{DR}}(X/S) \arrow[r] & \mathrm{gr}^D_{i+j} H^{d+e}_{\mathrm{DR}}(X/S). \end{tikzcd} \tag{9.7}$$

Hence we obtain a morphism of F-zips over S

$$\cup\colon \underline{H}^d_{\mathrm{DR}}(X/S) \otimes \underline{H}^e_{\mathrm{DR}}(X/S) \to \underline{H}^{d+e}_{\mathrm{DR}}(X/S), \tag{9.8}$$

Example 9.9. Let $A \to S$ be an abelian scheme. Then the cup product yields an isomorphism of graded anticommutative algebras

$$\Lambda^\bullet H^1_{\mathrm{DR}}(A/S) \xrightarrow{\sim} H^\bullet_{\mathrm{DR}}(A/S)$$

(see for example [Berthelot et al. 1982, Proposition 2.5.2]). The above arguments show that this is in fact an isomorphism of F-zips.

Now assume in addition that f has geometrically connected fibers of fixed dimension n. Then we have a trace isomorphism

$$R^n f_* \Omega^n_{X/S} \cong H^{2n}_{\mathrm{DR}}(X/S) \overset{\mathrm{tr}}{\xrightarrow{\sim}} \mathcal{O}_S. \tag{9.10}$$

In other words, we obtain an isomorphism of F-zips

$$\underline{H}^{2n}_{\mathrm{DR}}(X/S) \overset{\mathrm{tr}}{\xrightarrow{\sim}} \underline{\mathbb{1}}(n). \tag{9.11}$$

The cup-product pairings of F-zips

$$\underline{H}^d_{\mathrm{DR}}(X/S) \otimes_{\mathcal{O}_S} \underline{H}^{2n-d}_{\mathrm{DR}}(X/S) \longrightarrow \underline{H}^{2n}_{\mathrm{DR}}(X/S) = \underline{\mathbb{1}}(n) \tag{9.12}$$

are perfect dualities [Katz 1972, (2.3.5.1)], that is they yield isomorphisms of F-zips

$$\underline{H}^d_{\mathrm{DR}}(X/S) \xrightarrow{\sim} \underline{H}^{2n-d}_{\mathrm{DR}}(X/S)^\vee(n). \tag{9.13}$$

For $d = n$ the morphism (9.12) factors through

$$\begin{aligned} v\colon \Lambda^2 \underline{H}^n_{\mathrm{DR}}(X/S) &\to \underline{\mathbb{1}}(n), \qquad \text{if } n \text{ is odd;} \\ b\colon S^2 \underline{H}^n_{\mathrm{DR}}(X/S) &\to \underline{\mathbb{1}}(n), \qquad \text{if } n \text{ is even.} \end{aligned} \tag{9.14}$$

In other words the pairing is symplectic if n is odd and it is symmetric if n is even.

For n odd we hence obtain a twisted symplectic F-zip $(\underline{H}^n_{\mathrm{DR}}(X/S), \underline{\mathbb{1}}(n), v)$. For n even (and $p > 2$) we obtain a twisted orthogonal F-zip $(\underline{H}^n_{\mathrm{DR}}(X/S), \underline{\mathbb{1}}(n), b)$.

9C. ***Zip strata attached to truncated Barsotti–Tate groups of level* 1.** Let S be a scheme over $\mathbb{F}_p$ and let X be a truncated Barsotti–Tate group of level 1 over S. We denote by $X^\vee$ its Cartier dual. Let $\mathbb{D}(X)$ be its *covariant* Dieudonné crystal and let $\mathcal{M}(X)$ be its evaluation at the trivial PD-thickening $(S, S, 0)$. Then there is an exact sequence, functorial in X and compatible with base change $S' \to S$

$$0 \to \omega_{X^\vee} \to \mathcal{M}(X) \to \mathrm{Lie}(X) \to 0, \tag{9.15}$$

where $\omega_{X^\vee} = e^*\Omega_{X^\vee/S}$ is the sheaf of $\mathcal{O}_S$-modules of invariant differentials of $X^\vee$ (see [Berthelot et al. 1982, Corollary 3.2]). In particular the relative Frobenius $F\colon X \to X^{(p)}$ and the Verschiebung $V\colon X^{(p)} \to X$ induce $\mathcal{O}_S$-linear homomorphisms

$$\mathcal{F} := \mathcal{M}(V)\colon \mathcal{M}(X)^{(p)} \to \mathcal{M}(X), \qquad \mathcal{V} := \mathcal{M}(F)\colon \mathcal{M}(X) \to \mathcal{M}(X)^{(p)}.$$

Note that the roles of F and V are switched as we are considering covariant Dieudonné theory. Moreover

$$(\omega_{X^\vee})^{(p)} = \ker(\mathcal{F}) = \mathrm{Im}(\mathcal{V}), \qquad \ker(\mathcal{V}) = \mathrm{Im}(\mathcal{F})$$

are locally direct summands of $\mathcal{M}(X)^{(p)}$ and of $\mathcal{M}(X)$, respectively.

We attach an F-zips $\underline{\mathcal{M}}(X) := (\mathcal{M}(X), C^\bullet, D_\bullet, \varphi_\bullet)$ as follows. Set

$$\begin{aligned} C^0 &:= \mathcal{M}(X), & C^1 &:= \omega_{X^\vee}, & C^2 &:= 0 \\ D_{-1} &:= 0, & D_0 &:= \ker(\mathcal{V}), & D_1 &:= \mathcal{M}(X) \end{aligned}$$

and let

$$\varphi_0\colon \mathcal{M}(X)^{(p)}/(C^1)^{(p)} \to D_0, \qquad \varphi_1\colon (C^1)^{(p)} \to \mathcal{M}(X)/D_0$$

be the $\mathcal{O}_S$-linear isomorphisms induced by $\mathcal{F}$ and $\mathcal{V}^{-1}$, respectively.

Altogether we obtain a functor $X \mapsto \underline{\mathcal{M}}(X)$ from the category of truncated Barsotti–Tate groups of level 1 over S to the category of F-zips over S. Moreover it follows from [Berthelot et al. 1982, Proposition 5.2] that there is an isomorphism of F-zips

$$\underline{\mathcal{M}}(X^\vee) \cong \underline{\mathrm{Hom}}(\underline{\mathcal{M}}(X), \underline{\mathbb{1}}(1)), \tag{9.16}$$

which is functorial in X and compatible with base change. If $d := \mathrm{rk}_{\mathcal{O}_S}(\mathrm{Lie}\, X)$ is the dimension of X and n its height, then the type of the F-zip $\underline{\mathcal{M}}$ is $(n_i)_i$ with

$$n_0 = d, \qquad n_1 = n - d, \qquad n_i = 0 \quad \text{for } i \neq 0, 1. \tag{9.17}$$

Truncated Barsotti–Tate groups of level 1 of height n and dimension d over schemes over $\mathbb{F}_p$ form a smooth algebraic stack $\mathtt{BT}_1^{n,d}$ of finite type over $\mathbb{F}_p$ [Wedhorn 2001, Proposition 1.8] and the above construction yields a morphism of algebraic stacks

$$\Phi\colon \mathtt{BT}_1^{n,d} \to F\text{-}\mathtt{Zip}_{\mathbb{F}_p}^{\underline{n}}, \tag{9.18}$$

where $\underline{n}$ is given by (9.17). By Dieudonné theory this functor is an equivalence on points with values in a perfect field. In particular, for every algebraically closed field K of characteristic p we obtain a bijection

$$\begin{Bmatrix} \text{isomorphism classes of truncated Barsotti–Tate groups} \\ \text{over } K \text{ of level 1, height } n, \text{ and dimension } d \end{Bmatrix} \tag{9.19}$$
$$\leftrightarrow \{w \in S_n : w^{-1}(1) < \cdots < w^{-1}(d),\, w^{-1}(d+1) < \cdots < w^{-1}(n)\}$$

This was first proved by Moonen [2001].

The following results (all due to Eike Lau) show that Φ (9.18) is a smooth (nonrepresentable) morphism.

Remark 9.20. Let R be an $\mathbb{F}_p$-algebra. Let σ be the ring endomorphism $x \mapsto x^p$, let $I := R_{(\sigma)}$ be the restrictions of scalars of the R-module R under σ, and let $\sigma_1\colon I \to R$ be the σ-linear map given by the identity of R. Lau [2013] has defined the notion of a *display of level* 1 over R. Recall that this is a tuple $\mathcal{D} = (P, Q, \iota, \varepsilon, F, F_1)$ consisting of R-modules P and Q together with R-linear maps $I \otimes P \xrightarrow{\varepsilon} Q \xrightarrow{\iota} P$ such that P and $\mathrm{coker}(\iota)$ are finitely generated and projective and such that the following sequence is exact

$$0 \longrightarrow I \otimes \mathrm{coker}(\iota) \xrightarrow{\varepsilon} Q \xrightarrow{\iota} P \longrightarrow \mathrm{coker}(\iota) \longrightarrow 0.$$

Finally, $F\colon P \to P$ and $F_1\colon Q \to P$ are σ-linear maps such that $F_1(Q)$ generates P and such that $F_1 \circ \varepsilon = \sigma_1 \otimes F$. The rank of P is called the *height of* $\mathcal{D}$ and the rank of $\mathrm{coker}(\iota)$ is called the *dimension of* $\mathcal{D}$. One has the obvious notion of an isomorphism of displays of level 1 and of base change for a ring homomorphism $R \to R'$. One obtains the category $\mathtt{Disp}_1^{n,d}$ of level 1 displays of height n and dimension d fibered over the category of $\mathbb{F}_p$-algebras. This is a smooth algebraic stack over $\mathbb{F}_p$ [Lau 2013, Proposition 3.15].

To every display $\mathcal{D} = (P, Q, \iota, \varepsilon, F, F_1)$ of level 1 of height n and dimension d over R one can attach an F-zip of type $\underline{n}$ with $\underline{n}$ as in (9.17) as follows. We set $M = P$, $C^1 = \mathrm{im}(\iota\colon Q \to P)$, $D_0 := \mathrm{im}(F^\sharp\colon P^{(\sigma)} \to P)$, where $F^\sharp$ denotes the linearization of F, and we define φ_0, φ_1 as the linearizations of the σ-linear maps $\varphi_0^\flat$, $\varphi_1^\flat$ defined by the following commutative diagrams

$$\begin{tikzcd} P \arrow[r, "F"] \arrow[d] & P \arrow[d] \\ P/C^1 \arrow[r, "\varphi_0^\flat"] & D_0 \end{tikzcd} \qquad \begin{tikzcd} Q \arrow[r, "F_1"] \arrow[d] & P \arrow[d] \\ C^1 \arrow[r, "\varphi_1^\flat"] & P/D_0. \end{tikzcd}$$

Then it is straight forward to check (by choosing a normal decomposition, see [Lau 2013, Section 3.2]) that this contruction defines an equivalence of the category of displays of level 1 of height n and dimension d over R with the category of F-zips of type $\underline{n}$ over R. We obtain an equivalence of algebraic stacks $\Theta\colon \mathtt{Disp}_1^{n,d} \xrightarrow{\sim} F\text{-}\mathtt{Zip}_{\mathbb{F}_p}^{\underline{n}}$.

Lau has defined a morphism $\Psi\colon \mathtt{BT}_1^{n,d} \to \mathtt{Disp}_1^{n,d}$ of algebraic stacks such that the composition with the equivalence Θ is the morphism Φ [Lau 2013, Section 4]. Moreover he proves that Ψ is a smooth morphism (Theorem A of loc. cit.) which shows the smoothness of Φ.

Example 9.21. Let $\mathbb{X}$ be a p-divisible group of height n and dimension d over a finite field $\mathbb{F}_q$ (where q is a power of p). Let $\mathcal{N}$ be the attached Rapoport–Zink space over $\mathbb{F}_q$ [Rapoport and Zink 1996], that is, $\mathcal{N}(S)$ consists for every scheme S over $\mathbb{F}_q$ of isomorphism classes of pairs (X, ρ), where X is a p-divisible group over S and $\rho\colon \mathbb{X}_S \to X$ is a quasiisogeny. Then $\mathcal{N}$ is representable by a formal scheme locally formally of finite type over $\mathbb{F}_q$. Attaching to X its p-torsion $X[p]$ defines a morphism

$$\varepsilon\colon \mathcal{N} \to (\mathtt{BT}_1^{n,d}) \otimes \mathbb{F}_q$$

which is formally smooth by the main result of [Illusie 1985] and by Drinfeld's result that quasiisogenies between p-divisible groups can always be deformed uniquely. Composing ε with the smooth morphism Φ from (9.18) we obtain a formally smooth (and hence generizing) morphism $\mathcal{N} \to F\text{-}\mathtt{Zip}_{\mathbb{F}_q}^{\underline{n}}$ and hence locally closed formal subschemes $\mathcal{N}^w$ as in (3.28) with

$$\bar{\mathcal{N}}^w = \bigcup_{w' \preceq w} \mathcal{N}^{w'}.$$

This can be generalized to other Rapoport–Zink spaces.

For every truncated Barsotti–Tate group X of level 1 over a scheme S over $\mathbb{F}_p$ the morphism $\Phi\colon \mathtt{BT}_1^{n,d} \to F\text{-}\mathtt{Zip}_{\mathbb{F}_p}^{\underline{n}}$ from (9.18) yields a homomorphism of the automorphism group schemes

$$\alpha\colon\ \underline{\mathrm{Aut}}(X) \to \underline{\mathrm{Aut}}(\underline{\mathcal{M}}(X)).$$

As both stacks are quotient stacks of a linear group acting on a scheme of finite type over $\mathbb{F}_p$, the two group schemes $\underline{\mathrm{Aut}}(X)$ and $\underline{\mathrm{Aut}}(\underline{\mathcal{M}}(X))$ are affine and of finite type over S (Proposition 2.5). If $S = \mathrm{Spec}\, K$ for an algebraically closed field K, it is shown in [Wedhorn 2001, Section (5.7)], that α induces a homeomorphism of the reduced subgroup schemes

$$\underline{\mathrm{Aut}}(X)_{\mathrm{red}} \xrightarrow{\sim} \underline{\mathrm{Aut}}(\underline{\mathcal{M}}(X))_{\mathrm{red}}.$$

Hence we can use Proposition 3.34 to describe $\underline{\mathrm{Aut}}(X)$.

Proposition 9.22. *Let X be a truncated Barsotti–Tate group of level 1 over an algebraically closed field K of characteristic p. Let n be its height and d the dimension of its Lie algebra. Let w be the permutation corresponding to the isomorphism class of X via the bijection (9.19).*

(a) *The reduced subgroup scheme of the identity component $\underline{\mathrm{Aut}}(X)$ is a unipotent linear algebraic group of dimension $d(n-d) - \ell(w)$. In particular*

$$\dim(\underline{\mathrm{Aut}}(X)) = d(n-d) - \ell(w).$$

(b) *The group of connected components of $\underline{\mathrm{Aut}}(X)$ is isomorphic to the group Π defined in Proposition 3.34.*

Note that for a permutation $w \in S_n$ the length can be easily computed by

$$\ell(w) = \#\{(i, j) : 1 \leq i < j \leq n,\ w(i) > w(j)\}.$$

Acknowledgments

We thank Chao Zhang for pointing out a mistake in an earlier version of the proof of Lemma 3.5 and Eike Lau for explaining us Remark 9.20. We are grateful to the referee for helpful remarks.

References

[Berthelot et al. 1982] P. Berthelot, L. Breen, and W. Messing, *Théorie de Dieudonné cristalline, II*, Lecture Notes in Math. **930**, Springer, Berlin, 1982. MR 85k:14023 Zbl 0516.14015

[Bühler 2010] T. Bühler, "Exact categories", *Expo. Math.* **28**:1 (2010), 1–69. MR 2606234 Zbl 1192.18007

[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

[Deligne and Illusie 1987] P. Deligne and L. Illusie, "Relèvements modulo p^2 et décomposition du complexe de de Rham", *Invent. Math.* **89**:2 (1987), 247–270. MR 88j:14029 Zbl 0632.14017

[Grothendieck 1963] A. Grothendieck, "Éléments de géométrie algébrique, III: Étude cohomologique des faisceaux cohérents (2e partie)", *Inst. Hautes Études Sci. Publ. Math.* **17** (1963), 5–91. MR 29 #1210 Zbl 0122.16102

[Grothendieck 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, (4e partie)", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR 39 #220 Zbl 0153.22301

[He 2007] X. He, "The G-stable pieces of the wonderful compactification", *Trans. Amer. Math. Soc.* **359**:7 (2007), 3005–3024. MR 2008c:14063 Zbl 1124.20033

[Illusie 1985] L. Illusie, "Déformations de groupes de Barsotti–Tate (d'après A. Grothendieck)", pp. 151–198 in *Séminaire sur les pinceaux arithmétiques: la conjecture de Mordell* (Paris, 1983–1984), edited by L. Szpiro, Astérisque **127**, Sociéte Mathématique de France, Paris, 1985. MR 801922 Zbl 1182.14050

[Katz 1972] N. M. Katz, "Algebraic solutions of differential equations (p-curvature and the Hodge filtration)", *Invent. Math.* **18** (1972), 1–118. MR 49 #2728 Zbl 0278.14004

[Lau 2013] E. Lau, "Smoothness of the truncated display functor", *J. Amer. Math. Soc.* **26**:1 (2013), 129–165. MR 2983008 Zbl 1273.14040

[Laumon and Moret-Bailly 2000] G. Laumon and L. Moret-Bailly, *Champs algébriques*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **39**, Springer, Berlin, 2000. MR 2001f:14006 Zbl 0945.14005

[Lusztig 2004a] G. Lusztig, "Parabolic character sheaves, I", *Mosc. Math. J.* **4**:1 (2004), 153–179. MR 2006d:20091a Zbl 1102.20030

[Lusztig 2004b] G. Lusztig, "Parabolic character sheaves, II", *Mosc. Math. J.* **4**:4 (2004), 869–896. MR 2006d:20091b Zbl 1103.20041

[Matsuki and Olsson 2005] K. Matsuki and M. Olsson, "Kawamata–Viehweg vanishing as Kodaira vanishing for stacks", *Math. Res. Lett.* **12**:2-3 (2005), 207–217. MR 2006c:14023 Zbl 1080.14023

[Moonen 2001] B. Moonen, "Group schemes with additional structures and Weyl group cosets", pp. 255–298 in *Moduli of abelian varieties* (Texel Island, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR 2002c:14074 Zbl 1084.14523

[Moonen and Wedhorn 2004] B. Moonen and T. Wedhorn, "Discrete invariants of varieties in positive characteristic", *Int. Math. Res. Not.* **2004**:72 (2004), 3855–3903. MR 2005k:14040 Zbl 1084.14023

[Nori 1976] M. V. Nori, "On the representations of the fundamental group", *Compositio Math.* **33**:1 (1976), 29–41. MR 54 #5237 Zbl 0337.14016

[Pink et al. 2011] R. Pink, T. Wedhorn, and P. Ziegler, "Algebraic zip data", *Doc. Math.* **16** (2011), 253–300. MR 2012j:14065 Zbl 1230.14070 arXiv 1010.0811

[Rapoport and Richartz 1996] M. Rapoport and M. Richartz, "On the classification and specialization of F-isocrystals with additional structure", *Compositio Math.* **103**:2 (1996), 153–181. MR 98c:14015 Zbl 0874.14008

[Rapoport and Zink 1996] M. Rapoport and T. Zink, *Period spaces for p-divisible groups*, Annals of Mathematics Studies **141**, Princeton University Press, 1996. MR 97f:14023 Zbl 0873.14039

[Saavedra Rivano 1972] N. Saavedra Rivano, *Catégories Tannakiennes*, Lecture Notes in Math. **265**, Springer, Berlin, 1972. MR 49 #2769 Zbl 0241.14008

[Satriano 2012] M. Satriano, "De Rham theory for tame stacks and schemes with linearly reductive singularities", *Ann. Inst. Fourier* (*Grenoble*) **62**:6 (2012), 2013–2051. MR 3060750 Zbl 06159904 arXiv 0911.2056

[Wedhorn 2001] T. Wedhorn, "The dimension of Oort strata of Shimura varieties of PEL-type", pp. 441–471 in *Moduli of abelian varieties* (Texel Island, 1999), edited by C. Faber et al., Progr. Math. **195**, Birkhäuser, Basel, 2001. MR 2002b:14029 Zbl 1052.14026

[Wedhorn and Yatsyshyn 2014] T. Wedhorn and Y. Yatsyshyn, "Purity of zip strata with applications to Ekedahl–Oort strata of Shimura varieties of Hodge type", preprint, 2014. arXiv 1404.0577

[Wortmann 2013] D. Wortmann, "The μ-ordinary locus for Shimura varieties of Hodge type", preprint, 2013. arXiv 1310.6444

[Zhang 2013] C. Zhang, "Ekedahl–Oort strata for good reductions of Shimura varieties of Hodge type", preprint, 2013. arXiv 1312.4869

[Ziegler 2011] P. Ziegler, "Graded and filtered fiber functors on Tannakian categories", preprint, 2011. To appear in J. Inst. Math. Jussieu. arXiv 1111.1981

Received September 5, 2013. Revised April 3, 2014.

RICHARD PINK
DEPARTMENT OF MATHEMATICS
ETH ZÜRICH
CH-8092 ZÜRICH
SWITZERLAND

pink@math.ethz.ch

TORSTEN WEDHORN
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT PADERBORN
WARBURGER STRASSE 100
D-33098 PADERBORN
GERMANY

wedhorn@math.uni-paderborn.de

PAUL ZIEGLER
DEPARTMENT OF MATHEMATICS
ETH ZÜRICH
CH-8092 ZÜRICH
SWITZERLAND

paul.ziegler@math.ethz.ch

PACIFIC JOURNAL OF MATHEMATICS
Vol. 274, No. 1, 2015
dx.doi.org/10.2140/pjm.2015.274.237

MEAN VALUES OF L-FUNCTIONS OVER FUNCTION FIELDS

JEFFREY LIN THUNDER

For a fixed global function field, positive integer and complex number, we prove estimates for mean values of L-functions evaluated at the given complex number, where the averaging is done over quadratic extensions of the given function field with genus equal to the given positive integer. To accomplish this we utilize our previous results on certain quadratic character sums over function fields.

1. Introduction

In modern number theory L-series play a prominent role. They encode many deep properties of number fields and primes and are objects of intense interest. The analogous L-functions over global function fields play an equally prominent role. Here we will prove estimates for mean values of such L-functions, where the averaging is done over quadratic extensions of a fixed global function field. Our estimates cover a much wider range of cases than the similar estimates of Hoffstein and Rosen [1992] and those of Andrade and Keating (for values on the critical line) [2012]. Our methods are akin to those used by Siegel [1944], where he estimates the average number of quadratic forms with given discriminant and signature.

For a prime p, let $\mathbb{F}_p$ denote the finite field with p elements and let X be transcendental over $\mathbb{F}_p$, so that $\mathbb{F}_p(X)$ is a field of rational functions. Fix algebraic closures $\overline{\mathbb{F}_p}$ of $\mathbb{F}_p$ and $\overline{\mathbb{F}_p(X)} \supset \overline{\mathbb{F}_p}$ of $\mathbb{F}_p(X)$. In what follows, by *global function field* (or simply *function field*) we mean a finite algebraic extension $K \supseteq \mathbb{F}_p(X)$ contained in $\overline{\mathbb{F}_p(X)}$. For such a field K we have $K \cap \overline{\mathbb{F}_p} = \mathbb{F}_{q_K}$ for some finite field $\mathbb{F}_{q_K}$ with q_K elements; this field is called the field of constants of K. We write g_K for the genus of K and J_K for the number of divisor classes of degree 0. We denote the set of places of K by $M(K)$ and the divisor group (i.e., the free abelian group generated by the places) by $\operatorname{Div} K$. The reader can refer to Chapters I and V of [Stichtenoth 1993] for a thorough background on these notions. We will use capital script German letters to denote divisors $\mathfrak{A}$, $\mathfrak{B}$, etc., with the sole exception of the zero divisor 0. For any divisor $\mathfrak{A} \in \operatorname{Div} K$ we write $\mathfrak{A} = \sum \operatorname{ord}_v(\mathfrak{A}) \cdot v$, where the

MSC2010: 11M41.
Keywords: function field, L-functions.

sum is over all places $v \in M(K)$. The support $\operatorname{Supp}\mathfrak{A}$ of a divisor $\mathfrak{A} \in \operatorname{Div} K$ is the finite (possibly empty) set of places $v \in M(K)$ where $\operatorname{ord}_v(\mathfrak{A}) \neq 0$. We say $\mathfrak{A}$ is effective if $\mathfrak{A} \geq 0$, i.e., $\operatorname{ord}_v(\mathfrak{A}) \geq 0$ for all places $v \in M(K)$. The degree map on $\operatorname{Div} K$, normalized to have image $\mathbb{Z}$ (see Chapter V of [Stichtenoth 1993] again) will be denoted deg.

With the above notation, the zeta function ζ_K is given by

$$\zeta_K(s) = \sum_{\substack{\mathfrak{A} \in \operatorname{Div} K \\ \mathfrak{A} \geq 0}} q^{-s \deg \mathfrak{A}}.$$

Since

$$\sum_{\substack{\mathfrak{A} \in \operatorname{Div} K \\ \mathfrak{A} \geq 0,\, \deg \mathfrak{A} = j}} 1 = \frac{J_K}{q_K - 1}(q_K^{j+1-g_K} - 1) \tag{0}$$

for all integers $j \geq 2g_K - 1$ by the Riemann–Roch Theorem (see [Stichtenoth 1993, Lemma V.1.4], for example), the series defining $\zeta_K(s)$ converges for all $s \in \mathbb{C}$ with $\Re(s) > 1$. The L-function L_K is given by

$$L_K(q_K^{-s}) = (1 - q_K^{-s})(1 - q_K^{1-s})\zeta_K(s) = \frac{\zeta_K(s)}{\zeta_{\mathbb{F}_{q_K}(X)}(s)}.$$

It is well known that L_K is a polynomial of degree $2g_K$ in q_K^{-s} and all its zeros have $\Re(s) = \frac{1}{2}$ (see [Stichtenoth 1993, Chapter V], for example).

For a fixed function field K and integer $m \geq 0$, we will be concerned with sums over quadratic extensions of K with genus m and the same field of constants $\mathbb{F}_{q_K}$. We first denote the number of such quadratic extensions

$$N_K(m) = \sum_{\substack{[F:K]=2 \\ g_F = m,\, q_F = q_K}} 1.$$

We note that $N_K(m)$ is asymptotically $q_K^{2m} 2 J_K q_K^{3-5g_K} / \zeta_K(2)(q_K - 1)$ as $m \to \infty$ (see Proposition 1 below). We will investigate the arithmetic mean of the set of values $L_F(q^{-s})$ over the quadratic extensions $F \supset K$ of genus m, and higher moments as well. It will prove convenient to multiply these means by the L-function value of the ground field K, Thus, for a fixed $s \in \mathbb{C}$ and integer $n \geq 1$, we set

$$M_K(s, m, n) = (N_K(m))^{-1} \sum_{\substack{[F:K]=2 \\ g_F = m,\, q_F = q_K}} L_F(q^{-s})^n L_K(q^{-s})^{-n}$$

provided $N_K(m) > 0$, and set $M_K(s, m, n) = 0$ otherwise. We will prove the following estimates:

Theorem 1. *Let K be a function field with field of constants $\mathbb{F}_q$. For all positive integers n and all $s \in \mathbb{C}$ with $\Re(s) > \frac{1}{2}$, set*

$$\sigma_K(s,n) = \prod_{v \in M(K)} \left(1 + \frac{P_n(q^{-s\deg v})}{(n-1)!(1+q^{-\deg v})(1-q^{-2s\deg v})^n}\right),$$

where $P_n(X) \in \mathbb{Z}[X]$ is given by

$$\frac{d^{n-1}X^{n+1}/(1-X^2)}{dX^{n-1}} = \frac{P_n(X)}{(1-X^2)^n}.$$

Then for all integers $m \geq 0$, all $\epsilon > 0$ and all $s \in \mathbb{C}$ with $\Re(s) > 1 + (n-1)\epsilon$ if q is odd, or $\Re(s) > 1 + n\epsilon - 1/2n$ if q is even, we have

$$|M_K(s,m,n) - \sigma_K(s,n)| \leq \begin{cases} c(\epsilon)^n(q^{-m} + q^{-4m(\Re(s)-1-(n-1)\epsilon)}) & \text{if } q \text{ is odd,} \\ c(\epsilon)^n(q^{-m} + q^{-2mn(\Re(s)-1-\epsilon+1/2n)}) & \text{if } q \text{ is even,} \end{cases}$$

where the constant $c(\epsilon) > 0$ depends only on K and ϵ.

We obtain stronger estimates (i.e., better error terms) when we consider the case $n = 1$:

Theorem 2. *Let K be a function field with field of constants $\mathbb{F}_q$ and let m be an integer with $m > g_K$. Then $M_K(s,m,1)$ is a polynomial of degree $2(m-g_K)$ in q^{-s} satisfying the same functional equation as the L-function,*

$$M_K(s,m,1) = q^{m-g_K} q^{-s2(m-g_K)} M_K(1-s,m,1).$$

Further, $M_K(s,m,1)$ is an even function in q^{-s} with

$$M_K(s,m,1) = 1 + a_2 q^{-2s} + \cdots + a_{2(m-g_K)} q^{-2s(m-g_K)}.$$

We have $a_{2(m-g_K-j)} = q^{m-g_K-2j} a_{2j}$ for all $j = 0, \ldots, m-g_K$, and, for all $\epsilon > 0$,

$$a_{2j} = \sum_{\substack{\mathfrak{C} \geq 0 \\ \deg \mathfrak{C} = j}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1+q^{-\deg v})^{-1} + \begin{cases} O(q^{-m} q^{2j(5/4+\epsilon)}) & \text{if } q \text{ is odd,} \\ O(q^{-m} q^{2j(1+\epsilon)}) & \text{if } q \text{ is even,} \end{cases}$$

where the implicit constants depend only on K and ϵ. Finally,

$$\sum_{\substack{\mathfrak{C} \geq 0 \\ \deg \mathfrak{C} = j}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1+q^{-\deg v})^{-1} = q^j \frac{J_K q^{1-g_K} \zeta_K(2)}{q-1} \prod_{v \in M(K)} (1 - 2q^{-2\deg v} + q^{-3\deg v}) + O(q^{\epsilon j})$$

for all $j \geq 0$, where the implicit constant depends only on K and ϵ.

Corollary 1. *Let K be a function field with field of constants $\mathbb{F}_q$ and let m be a positive integer. Then for all $\epsilon > 0$ and all $s \in \mathbb{C}$ with $\Re(s) > \frac{1}{2}$*

$$M_K(s, m, 1) = \sigma_K(s, 1) + \begin{cases} O\left(\dfrac{q^{-m(2/3-\epsilon)(2\Re(s)-1)}}{1-q^{1-2\Re(s)}}\right) & \text{if } q \text{ is odd,} \\ O\left(\dfrac{q^{-m(1-\epsilon)(2\Re(s)-1)}}{1-q^{1-2\Re(s)}}\right) & \text{if } q \text{ is even,} \end{cases}$$

where the implicit constants depend only on K and ϵ. In particular, setting $s = 1$ we have

$$\begin{aligned}(N_K(m))^{-1} \sum_{\substack{[F:K]=2 \\ g_F=m,\ q_F=q}} J_F q^{-g_F} &= J_K q^{-g_K} (\zeta_K(2))^2 \prod_{v \in M(K)} (1 - q^{-2\deg v} - q^{-3\deg v} + q^{-4\deg v}) \\ &\qquad + \begin{cases} O(q^{-m(2/3-\epsilon)}) & \text{if } q \text{ is odd,} \\ O(q^{-m(1-\epsilon)}) & \text{if } q \text{ is even.} \end{cases}\end{aligned}$$

Mean values similar to those in Corollary 1 were previously considered by Hoffstein and Rosen [1992], but only in the case where the field K is a field of rational functions and only in odd characteristic. More general cases were considered by Fisher and Friedberg [2004], with further refinements by Chinta, Friedberg and Hoffstein [Chinta et al. 2006], but again only in odd characteristic. The higher moments in Theorem 1 have not been previously estimated to our knowledge. Our approach differs from those previous by utilizing more general estimates for quadratic characters over function fields, including estimates in characteristic 2 (which is clearly special when considering quadratic extensions).

Theorem 2 can also be used to estimate the "average" of $L_F(q^{-1/2})$ (cf. [Goldfeld and Hoffstein 1985, Theorem 1] for the case where K is replaced by $\mathbb{Q}$). As alluded to above, such a result is proven in [Andrade and Keating 2012], though again only in certain special cases where the ground field is a field of rational functions (specifically, for q congruent to 1 modulo 4).

Corollary 2. *Let K be a function field with field of constants $\mathbb{F}_q$ and let m be an integer with $m > g_K$. Set $m' = m - g_K$ and*

$$C(K) = \frac{J_K q^{1-g_K} \zeta_K(2)}{q-1} \prod_{v \in M(K)} (1 - 2q^{-2\deg v} + q^{-3\deg v}).$$

Then the series

$$C'(K) = \sum_{j=0}^{\infty} \Bigg(\sum_{\substack{\mathfrak{C} \geq 0 \\ \deg \mathfrak{C} = j}} q^{-\deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} \Bigg) - C(K)$$

converges and, for all $\epsilon > 0$,

$$M_K\big(\tfrac{1}{2}, m, 1\big) = (m'+1)C(K) + 2C'(K) + \begin{cases} O(q^{-m(1/4-\epsilon)}) & \textit{if } q \textit{ is odd,} \\ O(q^{-m(1/2-\epsilon)}) & \textit{if } q \textit{ is even,} \end{cases}$$

where the implicit constants depend only on K *and* ϵ.

Finally, we note that Theorem 2 can also be used to give the "average" number of places of degree 1.

Corollary 3. *Let* K *be a function field with field of constants* $\mathbb{F}_q$. *Then* (*assuming* $N_K(m) \neq 0$)

$$(N_K(m))^{-1} \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} \#\{w \in M(F) : \deg w = 1\} = \#\{v \in M(K) : \deg v = 1\}.$$

One can compare this with the famous estimate due to Drinfeld and Vladut [Stichtenoth 1993, Theorem V.3.5],

$$\limsup_{m\to\infty} \max_{\substack{F\supset K \\ g_F=m,\, q_F=q}} \frac{\#\{w \in M(F) : \deg w = 1\}}{m} \leq q^{1/2} - 1.$$

2. Preparatory Results

We briefly discuss separability issues before proceeding further. If K is a function field and $F \supset K$ is a quadratic extension, then F is clearly a separable extension if q_K is odd. If q_K is even this is not necessarily the case. However, it turns out that there is exactly one inseparable quadratic extension $F \supset K$ with $q_F = q_K$ when q_K is even; it satisfies $K = \{\alpha^2 : \alpha \in F\}$ and $g_F = g_K$ by [Stichtenoth 1993, Proposition III.9.2]. Therefore we can safely ignore this inseparable extension and tacitly assume in what follows that all quadratic extensions that appear are separable extensions.

If K is a function field, $v \in M(K)$ and F is a quadratic extension of K with $q_F = q_K$, we set

$$\chi(F/v) = \begin{cases} 0 & \text{if } v \text{ ramifies in } F, \\ 1 & \text{if } v \text{ is inert in } F, \\ -1 & \text{if } v \text{ splits in } F. \end{cases}$$

This is extended to effective divisors $\mathfrak{A} \in \operatorname{Div} K$ by

$$\chi(F/\mathfrak{A}) = \prod_{v \in \operatorname{Supp} \mathfrak{A}} \big(\chi(F/v)\big)^{\operatorname{ord}(\mathfrak{A})}.$$

The following is shown in [Thunder 2013, §1]:

Lemma 1. *Let K be a function field with field with field of constants $\mathbb{F}_q$ and $F \supset K$ be a quadratic extension with $q_F = q$. Then*

$$L_F(q^{-s}) = L_K(q^{-s}) \sum_{\substack{\mathfrak{A}\in \operatorname{Div} K\\ \mathfrak{A}\geq 0}} \chi(F/\mathfrak{A}) q^{-s\deg \mathfrak{A}},$$

so that

$$N_K(m) M_K(s,m,n) = \sum_{\substack{\mathfrak{C}\in \operatorname{Div} K\\ \mathfrak{C}\geq 0}} \sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C}}} q^{-s\deg \mathfrak{C}} \sum_{\substack{[F:K]=2\\ g_F=m, q_F=q_K}} \chi(F/\mathfrak{C}).$$

It turns out that the sums in Lemma 1 where $\deg \mathfrak{C}$ is odd vanish entirely and, when $\deg \mathfrak{C}$ is even, the $\mathfrak{C} \in 2\operatorname{Div} K$ dominate.

Lemma 2 [Thunder 2013, Lemma 9]. *Suppose K is a function field with field of constants $\mathbb{F}_q$ and m is a nonnegative integer. Then for all effective divisors $\mathfrak{C} \in \operatorname{Div} K$ of odd degree,*

$$\sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}) = 0.$$

Proposition 1 [Thunder 2013, Proposition 7]. *Let K be a function field with field of constants $\mathbb{F}_q$ and m be a nonnegative integer. Set*

$$N'_K(m) = q^{2m} \frac{2J_K q^{3-5g_K}}{\zeta_K(2)(q-1)}.$$

For all effective divisors $\mathfrak{C} \in \operatorname{Div} K$ and all $\epsilon > 0$

$$\left| \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/2\mathfrak{C}) - N'_K(m) \prod_{v\in \operatorname{Supp}\mathfrak{C}} (1+q^{-\deg v})^{-1} \right| \leq \begin{cases} c'(\epsilon) q^{(1/2+\epsilon)m} q^{\epsilon \deg \mathfrak{C}} & \text{if } q \text{ is odd,} \\ c'(\epsilon)\big(q^{\epsilon m} q^{\epsilon \deg \mathfrak{C}} + q^m\big) & \text{if } q \text{ is even,} \end{cases}$$

where $c'(\epsilon) > 0$ depends only on K and ϵ. In particular,

$$|N_K(m) - N'_K(m)| \leq \begin{cases} c'(\epsilon) q^{(1/2+\epsilon)m} & \text{if } q \text{ is odd,} \\ c'(1) q^m & \text{if } q \text{ is even.} \end{cases}$$

Proposition 2 [Thunder 2013, Proposition 5]. *Suppose K is a function field with $q_K = q$ odd and let $\mathfrak{C} \in \operatorname{Div} K$ be an effective divisor with $\mathfrak{C} \notin 2\operatorname{Div} K$. Then, for all nonnegative integers m and all $\epsilon > 0$, we have*

$$\left| \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}) \right| \leq c''(\epsilon) q^m q^{(\epsilon+1/4)\deg \mathfrak{C}},$$

where the constant $c''(\epsilon) > 0$ depends only on K and ϵ.

Proposition 3 [Thunder 2013, Proposition 6]. *Suppose K is a function field with $q_K = q$ even and let $\mathfrak{C} \in \operatorname{Div} K$ be an effective divisor with $\mathfrak{C} \notin 2 \operatorname{Div} K$. Then, for all nonnegative integers m and all $\epsilon > 0$, we have*

$$\left| \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}) \right| \leq c''(\epsilon) q^m q^{\epsilon \deg \mathfrak{C}},$$

where the constant $c''(\epsilon) > 0$ depends only on K and ϵ.

We also have the following elementary estimates:

Lemma 3. *Suppose K is a function field with field of constants $\mathbb{F}_q$. Let $\mathfrak{C} \in \operatorname{Div} K$ be an effective divisor. For all integers $n > 1$ and all $\epsilon > 0$,*

$$\sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1 + \cdots + \mathfrak{A}_n = \mathfrak{C}}} 1 \leq c_1(\epsilon)^{n-1} q^{(n-1)\epsilon \deg \mathfrak{C}},$$

where the constant $c_1(\epsilon) > 0$ depends only on K and ϵ. Also, for all positive integers m and all $\epsilon > 0$,

$$\sum_{\substack{\mathfrak{A} \geq 0 \\ \deg \mathfrak{A} \leq m}} q^{(\epsilon - 1) \deg \mathfrak{A}} \leq \frac{c_2}{\epsilon} q^{\epsilon m}, \qquad \sum_{\substack{\mathfrak{A} \geq 0 \\ \deg \mathfrak{A} \geq m}} q^{-(1+\epsilon) \deg \mathfrak{A}} \leq \frac{c_2}{\epsilon} q^{-\epsilon m},$$

and

$$\zeta_K(1+\epsilon) \leq \left(\frac{c_2}{\epsilon} \right)^{[K:\mathbb{F}_q(X)]},$$

where the constant $c_2 > 0$ depends only on K.

Proof. We prove the first part by induction on n. The case $n = 2$ follows directly from [Thunder 2013, Lemma 0]. Now assume $n > 2$. Then

$$\begin{aligned}
\sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1 + \cdots + \mathfrak{A}_n = \mathfrak{C}}} 1 &= \sum_{0 \leq \mathfrak{A}_n \leq \mathfrak{C}} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1 + \cdots + \mathfrak{A}_{n-1} = \mathfrak{C} - \mathfrak{A}_n}} 1 \\
&\leq \sum_{0 \leq \mathfrak{A}_n \leq \mathfrak{C}} c_1(\epsilon)^{n-2} q^{(n-2) \deg(\mathfrak{C} - \mathfrak{A}_n)} \\
&\leq c_1(\epsilon)^{n-2} q^{(n-2)\epsilon \deg \mathfrak{C}} \sum_{0 \leq \mathfrak{A}_n \leq \mathfrak{C}} 1 \\
&\leq c_1(\epsilon)^{n-1} q^{(n-1)\epsilon \deg \mathfrak{C}}.
\end{aligned}$$

For the next two inequalities, we see by (0) that there is a positive constant c, depending only on the field K, such that for all nonnegative integers j we have

$$\sum_{\substack{\mathfrak{A} \geq 0 \\ \deg \mathfrak{A} = j}} 1 \leq c q^j.$$

Finally, by the Euler product representation of the zeta function, we have

$$\zeta_K(s) \le (\zeta_{\mathbb{F}_q(X)}(s))^{[K:\mathbb{F}_q(X)]}$$

for all real $s > 1$. (See [Thunder and Widmer 2013, Lemma 2], for example.) Using the well-known formula

$$\zeta_{\mathbb{F}_q(X)}(s) = \frac{1}{(1-q^{-s})(1-q^{1-s})}$$

and substituting $s = 1+\epsilon$ gives

$$\zeta_K(1+\epsilon) \le \left(\frac{c'}{\epsilon}\right)^{[K:\mathbb{F}_q(X)]}$$

for some positive constant c' depending only on q. Setting c_2 to be the maximum of c and c' completes the proof. □

3. Proof of Theorem 1

We first deal with the summands in Lemma 1 where $\mathfrak{C} \in 2\,\mathrm{Div}\,K$. This is done in two steps.

Lemma 4. *Suppose K is a function field with field of constants $\mathbb{F}_q$ and $s \in \mathbb{C}$ with $\Re(s) > \frac{1}{2}$. Then, for all integers $n \ge 1$,*

$$\sum_{\mathfrak{C}\ge 0} \sum_{\substack{\mathfrak{A}_i \ge 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n = 2\mathfrak{C}}} q^{-2s\deg\mathfrak{C}} \prod_{v\in\mathrm{Supp}\,\mathfrak{C}} (1+q^{-\deg v})^{-1} = \sigma_K(s,n).$$

Proof. Set

$$\theta_n(\mathfrak{C}) = q^{-2s\deg\mathfrak{C}} \prod_{v\in\mathrm{Supp}\,\mathfrak{C}} (1+q^{-\deg v})^{-1} \sum_{\substack{\mathfrak{A}_i \ge 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n = 2\mathfrak{C}}} 1.$$

Note that $\theta_n(\mathfrak{C}+\mathfrak{D}) = \theta_n(\mathfrak{C})\theta_n(\mathfrak{D})$ for all n whenever $\mathfrak{C}, \mathfrak{D} \in \mathrm{Div}\,K$ have disjoint support. Thus

$$\sum_{\mathfrak{C}\ge 0} \theta_n(\mathfrak{C}) = \prod_{v\in M(K)} \left(1 + \sum_{k=1}^{\infty} \theta_n(kv)\right). \tag{1}$$

For all positive integers k and all places $v \in M(K)$,

$$\theta_n(kv) = (1+q^{-\deg v})^{-1} q^{-2ks\deg v} f(2k,n),$$

where

$$f(m,n) = \sum_{\substack{i_j \ge 0\\ i_1+\cdots+i_n = m}} 1 = \frac{(m+1)\cdots(m+n-1)}{(n-1)!}$$

for integers $m \geq 0$ and $n \geq 1$. Therefore

$$\sum_{k=1}^{\infty} \theta_n(kv) = (1+q^{-\deg v})^{-1} \sum_{k=1}^{\infty} q^{-2ks \deg v} \frac{(2k+1)\cdots(2k+n-1)}{(n-1)!}. \tag{2}$$

Differentiating term-by-term $n-1$ times yields

$$\frac{d^{n-1}\sum_{k=1}^{\infty} x^{2k+n-1}}{dx^{n-1}} = \sum_{k=1}^{\infty} x^{2k}(2k+1)\cdots(2k+n-1). \tag{3}$$

On the other hand,

$$\begin{aligned} \frac{d^{n-1}\sum_{k=1}^{\infty} x^{2k+n-1}}{dx^{n-1}} &= \frac{d^{n-1}x^{n-1}\sum_{k=1}^{\infty} x^{2k}}{dx^{n-1}} \\ &= \frac{d^{n-1}x^{n+1}\sum_{k=0}^{\infty} x^{2k}}{dx^{n-1}} \\ &= \frac{d^{n-1}x^{n+1}/(1-x^2)}{dx^{n-1}} \\ &= P_n(x)(1-x^2)^{-n}. \end{aligned} \tag{4}$$

The lemma follows from (1)–(4). □

Lemma 5. *Let K be a function field with field of constants $\mathbb{F}_q$. Suppose m is a nonnegative integer such that $N_K(m) > 0$. Then, for all $\epsilon > 0$ and all $s \in \mathbb{C}$ with $\Re(s) > (1+n\epsilon)/2$,*

$$\left| \frac{1}{N_K(m)} \sum_{\mathfrak{C} \geq 0} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n = 2\mathfrak{C}}} q^{-2s \deg \mathfrak{C}} \sum_{\substack{[F:K]=2 \\ g_F = m,\, q_F = q}} \chi(F/2\mathfrak{C}) - \sigma_K(s,n) \right| \leq \begin{cases} c_3(\epsilon)^{n+1} q^{-m(3/2-\epsilon)} & \text{if } q \text{ is odd}, \\ c_3(\epsilon)^n q^{-m} & \text{if } q \text{ is even}, \end{cases}$$

where $c_3(\epsilon) > 0$ depends only on K and ϵ.

Proof. We have

$$\begin{aligned} &\left| \sum_{\mathfrak{C} \geq 0} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n = 2\mathfrak{C}}} q^{-2s \deg \mathfrak{C}} \sum_{\substack{[F:K]=2 \\ g_F = m,\, q_F = q}} \chi(F/2\mathfrak{C}) - N_K(m)\sigma_K(s,n) \right| \\ &\leq \left| \sum_{\mathfrak{C} \geq 0} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n = 2\mathfrak{C}}} q^{-2s \deg \mathfrak{C}} \sum_{\substack{[F:K]=2 \\ g_F = m,\, q_F = q}} \chi(F/2\mathfrak{C}) - N'_K(m)\sigma_K(s,n) \right| \\ &\quad + |N_K(m) - N'_K(m)|\, |\sigma_K(s,n)|. \end{aligned} \tag{5}$$

By Lemma 3 and using $2\Re(s)-(n-1)\epsilon > 1+\epsilon$,

$$\text{(6)}\quad \left|\sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2s\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}\right|$$
$$\begin{aligned}&\leq \sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2\Re(s)\deg\mathfrak{C}}\\ &\leq c_1\Big(\frac{\epsilon}{2}\Big)^{n-1}\sum_{\mathfrak{C}\geq 0} q^{(-2\Re(s)+(n-1)\epsilon)\deg\mathfrak{C}}\\ &< c_1\Big(\frac{\epsilon}{2}\Big)^{n-1}\zeta_K(1+\epsilon)\\ &\leq c_4\epsilon^n,\end{aligned}$$

where $c_4(\epsilon)=\max\{c_1(\epsilon/2),\,(c_2/\epsilon)^{[K:\mathbb{F}_q(X)]}\}$.

Now, by Proposition 1, Lemma 4 and (6),

$$\text{(7a)}\quad \left|\sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2s\deg\mathfrak{C}}\sum_{\substack{[F:K]=2\\ g_F=m,\,q_F=q}}\chi(F/2\mathfrak{C})-N'_K(m)\sigma_K(s,n)\right|$$
$$\leq c'(\epsilon)q^{(1/2+\epsilon)m}\sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2\Re(s)\deg\mathfrak{C}}\leq c'(\epsilon)c_4(\epsilon)^n q^{(1/2+\epsilon)m}$$

if q is odd, and

$$\text{(7b)}\quad \left|\sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2s\deg\mathfrak{C}}\sum_{\substack{[F:K]=2\\ g_F=m,\,q_F=q}}\chi(F/2\mathfrak{C})-N'_K(m)\sigma_K(s,n)\right|$$
$$\ll c'(1)q^{m}\sum_{\mathfrak{C}\geq 0}\sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=2\mathfrak{C}}} q^{-2\Re(s)\deg\mathfrak{C}}\leq c'(1)c_4(\epsilon)^n q^{m}$$

if q is even. Also, by Proposition 1, Lemma 4 and (6)

$$\text{(8)}\quad |N_K(m)-N'_K(m)|\,|\sigma_K(s,n)|\leq\begin{cases}c'(\epsilon)c_4(\epsilon)^n q^{(1/2+\epsilon)m} & \text{if } q \text{ is odd,}\\ c'(1)c_4(\epsilon)^n q^m & \text{if } q \text{ is even.}\end{cases}$$

Finally, if $N_K(m)$ isn't zero, then

$$\text{(9)}\quad c_5q^{2m}\leq N_K(m)\leq c_6q^{2m}$$

by Proposition 1, where $c_5, c_6>0$ depend only on K. The lemma follows from (5) and (7a)–(9) when q is odd, and (5), (7b), (8) and (9) when q is even. □

With the sums over the main terms done, we now turn to the sums over the error terms, i.e., the sums where $\mathfrak{A}_1+\cdots+\mathfrak{A}_n\notin 2\operatorname{Div}K$. We have the following:

Lemma 6. *Suppose K is a function field with field of constants $\mathbb{F}_q$ and m is a nonnegative integer. Fix an integer $n \geq 2$ and an $\epsilon > 0$. Suppose that $s \in \mathbb{C}$ with $\Re(s) > \frac{1}{2} + (n-1)\epsilon$. Then, for any quadratic extension $F \supset K$ with $g_F = m$ and $q_F = q$,*

$$\sum_{\mathfrak{A}_i \geq 0} q^{-\Re(s)\deg(\mathfrak{A}_1+\cdots+\mathfrak{A}_{n-1})} \Bigg| \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n > 2m-2g_K \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n \notin 2\operatorname{Div} K}} q^{-s\deg \mathfrak{A}_n} \chi(F/\mathfrak{A}) \Bigg| \leq c_7 \epsilon^{n+1} q^{-2m(\Re(s)-1/2)},$$

where the constant $c_7(\epsilon) > 0$ depends only on K and ϵ.

Proof. For the moment, fix $\mathfrak{A}_1, \dots, \mathfrak{A}_{n-1}$ and set $\mathfrak{B} = \mathfrak{A}_1 + \cdots + \mathfrak{A}_{n-1}$. Write $\mathfrak{B} = \mathfrak{B}' + 2\mathfrak{B}''$, where $\mathfrak{B}'$ and $\mathfrak{B}''$ are both effective divisors and $\operatorname{ord}_v(\mathfrak{B}') = 1$ for all $v \in \operatorname{Supp} \mathfrak{B}'$. Fix an integer $j > 2m - 2g_K$. As shown in the proof of [Thunder 2013, Lemma 25],

$$\Bigg| \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n = j \\ \mathfrak{B}+\mathfrak{A}_n \notin 2\operatorname{Div} K}} \chi(F/\mathfrak{A}) q^{-s\deg \mathfrak{A}_n} \Bigg| \leq c_8 q^{-\deg(\mathfrak{B}')/2} q^{-j(\Re(s)-1/2)} \tag{10}$$

for some $c_8 > 0$ depending only on K. Now, since $\Re(s) - \frac{1}{2} > (n-1)\epsilon$,

$$\begin{aligned} \sum_{j > 2m-2g_K} q^{-j(\Re(s)-1/2)} &< \sum_{j > 2m-2g_K} q^{-j(n-1)\epsilon} \\ &\leq c_9 q^{-2m(n-1)\epsilon} \sum_{j \geq 0} q^{-j(n-1)\epsilon} \\ &= c_9 q^{-2m(\Re(s)-1/2)} (1 - q^{-(n-1)\epsilon})^{-1} \\ &\leq c_9 c_{10} q^{-2m(\Re(s)-1/2)} ((n-1)\epsilon)^{-1}, \end{aligned} \tag{11}$$

where $c_9, c_{10} > 0$ depend only on K. By Lemma 3,

$$\begin{aligned} \sum_{\mathfrak{B} \geq 0} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_{n-1} = \mathfrak{B}}} & q^{-\Re(s)\deg \mathfrak{B}} q^{-\deg(\mathfrak{B}')/2} \\ &\leq c_1(\epsilon)^{n-2} \sum_{\mathfrak{B} \geq 0} q^{((n-2)\epsilon - \Re(s))\deg \mathfrak{B}} q^{-\deg(\mathfrak{B}')/2} \\ &= c_1(\epsilon)^{n-2} \sum_{\mathfrak{B} \geq 0} q^{((n-2)\epsilon - 1/2 - \Re(s))\deg \mathfrak{B}'} q^{2((n-2)\epsilon - \Re(s))\deg \mathfrak{B}''} \\ &\leq c_1(\epsilon)^{n-2} \sum_{\mathfrak{B}' \geq 0} q^{((n-2)\epsilon - 1/2 - \Re(s))\deg \mathfrak{B}'} \sum_{\mathfrak{B}'' \geq 0} q^{2((n-2)\epsilon - \Re(s))\deg \mathfrak{B}''} \\ &< c_1(\epsilon)^{n-2} \zeta_K(1+\epsilon) \zeta_K(1+2\epsilon) < c_{11}(\epsilon)^n, \end{aligned} \tag{12}$$

where $c_{11}(\epsilon) = \max\{c_1(\epsilon), (c_2/\epsilon)^{[K:\mathbb{F}_q(X)]}\}$. The lemma follows from (10)–(12). $\square$

Proof of Theorem 1. Suppose first that q is odd. Since the cases $n=2$ and $n=1$ of Theorem 1 follow directly from [Thunder 2013, Theorem 1, Corollary 1], we will assume that $n \geq 3$. We may also assume that $N_K(m) > 0$. Rearranging the sums and then using Lemma 6 yields

$$\left| \frac{1}{N_K(m)} \sum_{\substack{\mathfrak{C}\geq 0 \\ \mathfrak{C}\notin 2\,\mathrm{Div}\,K}} \sum_{\substack{\mathfrak{A}_i\geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C} \\ \deg\mathfrak{A}_n > 2m-2g_K}} \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} q^{-s\deg\mathfrak{C}} \chi(F/\mathfrak{C}) \right| \leq c_7(\epsilon)^{n+1} q^{-2m(\Re(s)-1/2)} \tag{13}$$

whenever $\Re(s) > \frac{1}{2} + (n-1)\epsilon$.

Let $\delta > 0$, to be chosen later. Using Proposition 2 and setting $\mathfrak{B} = \mathfrak{A}_1 + \cdots + \mathfrak{A}_{n-1}$ in what follows, we have

$$\begin{aligned} &\left| \sum_{\substack{\mathfrak{C}\geq 0 \\ \mathfrak{C}\notin 2\,\mathrm{Div}\,K \\ \deg\mathfrak{C}<4m}} \sum_{\substack{\mathfrak{A}_i\geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C} \\ \deg\mathfrak{A}_n\leq 2m-2g_K}} q^{-s\deg\mathfrak{C}} \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}) \right| \\ &\leq c''(\delta) q^m \sum_{\substack{\mathfrak{A}_n\geq 0 \\ \deg\mathfrak{A}_n\leq 2m-2g_K}} q^{(\delta+1/4-\Re(s))\deg\mathfrak{A}_n} \times \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg\mathfrak{B}<4m-\deg\mathfrak{A}_n}} \sum_{\substack{\mathfrak{A}_i\geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_{n-1}=\mathfrak{B}}} q^{(\delta+1/4-\Re(s))\deg\mathfrak{B}} \\ &\leq c''(\delta) c_1(\delta)^{n-2} q^m \sum_{\substack{\mathfrak{A}_n\geq 0 \\ \deg\mathfrak{A}_n\leq 2m-2g_K}} q^{(\delta+1/4-\Re(s))\deg\mathfrak{A}_n} \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg\mathfrak{B}<4m-\deg\mathfrak{A}_n}} q^{((n-1)\delta+1/4-\Re(s))\deg\mathfrak{B}}. \end{aligned} \tag{14}$$

If $\Re(s) \leq \frac{5}{4} + (n-2)\epsilon$ we set $\delta = \epsilon$ above. Since $\frac{5}{4} + (n-1)\epsilon - \Re(s) \geq \epsilon$, Lemma 3 implies that

$$\begin{aligned} &\sum_{\substack{\mathfrak{A}_n\geq 0 \\ \deg\mathfrak{A}_n\leq 2m-2g_K}} q^{(\epsilon+1/4-\Re(s))\deg\mathfrak{A}_n} \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg\mathfrak{B}<4m-\deg\mathfrak{A}_n}} q^{((n-1)\epsilon+1/4-\Re(s))\deg\mathfrak{B}} \\ &\leq \frac{c_2}{\epsilon} q^{4m(5/4+(n-1)\epsilon-\Re(s))} \sum_{\substack{\mathfrak{A}_n\geq 0 \\ \deg\mathfrak{A}_n\leq 2m-2g_K}} q^{-(1+(n-2)\epsilon)\deg\mathfrak{A}_n} \\ &< \frac{c_2}{\epsilon} \zeta_K(1+(n-2)\epsilon) q^{4m(5/4+(n-1)\epsilon-\Re(s))} \\ &\leq \frac{c_2}{\epsilon} \left(\frac{c_2}{\epsilon}\right)^{[K:\mathbb{F}_q(X)]} q^{4m(5/4+(n-1)\epsilon-\Re(s))}. \end{aligned} \tag{15}$$

If $\Re(s) > \frac{5}{4} + (n-1)\epsilon$ we set $n\delta = \Re(s) - \frac{5}{4}$ and note that $\delta > (n-2)\epsilon/n \geq \epsilon/3$ since $\Re(s) > 1 + (n-1)\epsilon$. We thus may assume that $c''(\delta)c_1(\delta)^{n-2} \leq c_{12}(\epsilon)^{n-2}$

for some $c_{12}(\epsilon) > 0$ depending only on K and ϵ. Also, by Lemma 3,

$$\begin{aligned}(16)\quad \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n \leq 2m-2g_K}} q^{(\delta+1/4-\Re(s))\deg \mathfrak{A}_n} \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg \mathfrak{B} < 4m-\deg \mathfrak{A}_n}} q^{((n-1)\delta+1/4-\Re(s))\deg \mathfrak{B}} \\ < \sum_{\mathfrak{A}_n \geq 0} q^{-((n-1)\delta+1)\deg \mathfrak{A}_n} \sum_{\mathfrak{B}\geq 0} q^{-(\delta+1)\deg \mathfrak{B}} \\ < \zeta_K(1+(n-1)\delta)\zeta_K(1+\delta) \\ \leq \left(\frac{c_2}{\delta}\right)^{2[K:\mathbb{F}_q(X)]} \\ < \left(\frac{3c_2}{\epsilon}\right)^{2[K:\mathbb{F}_q(X)]}.\end{aligned}$$

When $\deg \mathfrak{C} \geq 4m$, we trivially estimate

$$\begin{aligned}(17)\quad \Bigg| \sum_{\substack{\mathfrak{C}\geq 0 \\ \mathfrak{C}\notin 2\operatorname{Div} K \\ \deg \mathfrak{C}\geq 4m}} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C} \\ \deg \mathfrak{A}_n \leq 2m-2g_K}} q^{-s\deg \mathfrak{C}} \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C})\Bigg| \\ \leq N_K(m) \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n \leq 2m-2g_K}} q^{-\Re(s)\deg \mathfrak{A}_n} \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg \mathfrak{B}\geq 4m-\deg \mathfrak{A}_n}} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_{n-1}=\mathfrak{B}}} q^{-\Re(s)\deg \mathfrak{B}} \\ \leq N_K(m)c_1(\epsilon)^{n-2} \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n \leq 2m-2g_K}} q^{-\Re(s)\deg \mathfrak{A}_n} \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg \mathfrak{B}\geq 4m-\deg \mathfrak{A}_n}} q^{((n-2)\epsilon-\Re(s))\deg \mathfrak{B}}.\end{aligned}$$

Since $\Re(s) > 1+(n-1)\epsilon$ by hypothesis, Lemma 3 implies that

$$(18)\quad \sum_{\substack{\mathfrak{B}\geq 0 \\ \deg \mathfrak{B}\geq 4m-\deg \mathfrak{A}_n}} q^{((n-2)\epsilon-\Re(s))\deg \mathfrak{B}} \leq \frac{c_2}{\epsilon} q^{(4m-\deg \mathfrak{A}_n)((n-2)\epsilon+1-\Re(s))}$$

and also

$$\begin{aligned}(19)\quad \sum_{\substack{\mathfrak{A}_n \geq 0 \\ \deg \mathfrak{A}_n \leq 2m-2g_K}} q^{-\Re(s)\deg \mathfrak{A}_n} q^{(\Re(s)-1-(n-2)\epsilon)\deg \mathfrak{A}_n} \\ < \zeta_K(1+(n-2)\epsilon) \leq \left(\frac{c_2}{\epsilon}\right)^{[K:\mathbb{F}_q(X)]}.\end{aligned}$$

Combining (9) with (13)–(19) yields

$$\begin{aligned}(20)\quad \Bigg| \frac{1}{N_K(m)} \sum_{\substack{\mathfrak{C}\geq 0 \\ \mathfrak{C}\notin 2\operatorname{Div} K}} \sum_{\substack{\mathfrak{A}_i \geq 0 \\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C}}} \sum_{\substack{[F:K]=2 \\ g_F=m,\, q_F=q}} q^{-s\deg \mathfrak{C}} \chi(F/\mathfrak{C})\Bigg| \\ \leq c_{13}(\epsilon)^{n+1}(q^{-m}+q^{-4m(\Re(s)-1-(n-1)\epsilon)})\end{aligned}$$

for some $c_{13}(\epsilon) > 0$ depending only on K and ϵ. The case where q is odd (and $n \geq 3$) in Theorem 1 follows from Lemma 1, Lemma 5 and (20).

Suppose now that q is even. This time we use Lemma 6 to get

$$(21)\quad \left|\frac{1}{N_K(m)} \sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\,\mathrm{Div}\,K}} \sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C}\\ \deg\mathfrak{A}_i>2m-2g_K \text{ for some } i}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} q^{-s\deg\mathfrak{C}}\chi(F/\mathfrak{C})\right| \leq nc_7(\epsilon)^{n+1}q^{-2m(\Re(s)-1/2)}$$

Let $\delta > 0$, to be chosen later. By Proposition 3,

$$(22)\quad \left|\sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\,\mathrm{Div}\,K}} \sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C}\\ \deg\mathfrak{A}_i\leq 2m-2g_K}} q^{-s\deg\mathfrak{C}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C})\right| \leq c''(\delta)q^m\left(\sum_{\substack{\mathfrak{A}\geq 0\\ \deg\mathfrak{A}\leq 2m-2g_K}} q^{(\delta-\Re(s))\deg\mathfrak{A}}\right)^n.$$

If $\Re(s) \leq 1+\epsilon/2$ we set $\delta = \epsilon$. Since $1+\epsilon-\Re(s) \geq \epsilon/2$, Lemma 3 implies that

$$(23)\quad \sum_{\substack{\mathfrak{A}\geq 0\\ \deg\mathfrak{A}\leq 2m-2g_K}} q^{(\epsilon-\Re(s))\deg\mathfrak{A}} \leq \frac{2c_2}{\epsilon}q^{(2m-2g_K)(1+\epsilon-\Re(s))} \leq c_{14}(\epsilon)q^{-2m(\Re(s)-1-\epsilon)},$$

where $c_{14}(\epsilon) > 0$ depends only on K and ϵ. If $\Re(s) > 1+\epsilon/2$ then we set $\delta = \epsilon/4$. We now have $c''(\delta) = c_{15}(\epsilon)$ and, by Lemma 3,

$$(24)\quad \sum_{\substack{\mathfrak{A}\geq 0\\ \deg\mathfrak{A}\leq 2m-2g_K}} q^{(\delta-\Re(s))\deg\mathfrak{A}} < \zeta_K(1+\epsilon/4) \leq \left(\frac{4c_2}{\epsilon}\right)^{[K:\mathbb{F}_q(X)]}.$$

Combining (9) and (21)–(24) gives

$$(25)\quad \left|\frac{1}{N_K(m)} \sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\,\mathrm{Div}\,K}} \sum_{\substack{\mathfrak{A}_i\geq 0\\ \mathfrak{A}_1+\cdots+\mathfrak{A}_n=\mathfrak{C}}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} q^{-s\deg\mathfrak{C}}\chi(F/\mathfrak{C})\right| \leq c_{16}(\epsilon)^n(q^{-m}+q^{-2mn(\Re(s)-1-\epsilon+1/2n)})$$

for some $c_{16}(\epsilon) > 0$ depending only on K and ϵ. The case where q is even in Theorem 1 follows from Lemma 1, Lemma 5 and (25). □

4. Proof of Theorem 2 and Corollaries

Proof of Theorem 2. We know that $M_K(s, m, 1)$ is a polynomial in q^{-s} thanks to a theorem of Weil (see [Rosen 2002, Theorem 9.16B]). It's an even function of q^{-s} by Lemma 2 and Lemma 1. Also, $a_0 = 1$ since $\chi(F/0) = 1$ by definition. The functional equation for $M_K(s, m, 1)$ follows directly from the functional equations for $L_K(q^{-s})$ and $L_F(q^{-s})$ for all quadratic extensions $F \supset K$. The identity

$$a_{2(m-g_K-j)} = q^{m-g_K-2j} a_{2j}, \quad j = 0, \ldots, m - g_K,$$

follows immediately from the functional equation.

Similar to the proof of Lemma 4, for an effective divisor $\mathfrak{C} \in \operatorname{Div} K$ set

$$\theta(\mathfrak{C}) = q^{-s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1}$$

and set $f(s) = \sum_{\mathfrak{C} \geq 0} \theta(\mathfrak{C})$. Since $\theta(\mathfrak{C} + \mathfrak{D}) = \theta(\mathfrak{C})\theta(\mathfrak{D})$ whenever $\mathfrak{C}$ and $\mathfrak{D}$ have disjoint support,

$$\begin{aligned}
f(s) &= \prod_{v \in M(K)} \left(1 + \sum_{k=1}^{\infty} \theta(kv)\right) \\
&= \prod_{v \in M(K)} \left(1 + \frac{q^{-s \deg v}}{(1 + q^{-\deg v})(1 - q^{-s \deg v})}\right) \\
&= \prod_{v \in M(K)} \left(1 + \frac{q^{-s \deg v}(1 - q^{-\deg v})}{(1 + q^{-2 \deg v})(1 - q^{-s \deg v})}\right) \\
&= \zeta_K(2)\zeta_K(s) \prod_{v \in M(K)} \big((1 - q^{-2 \deg v})(1 - q^{-s \deg v}) + q^{-s \deg v}(1 - q^{-\deg v})\big) \\
&= \zeta_K(2)\zeta_K(s) \prod_{v \in M(K)} (1 - q^{-2 \deg v} - q^{-(s+1) \deg v} + q^{-(s+2) \deg v}).
\end{aligned}$$

For any $\epsilon > 0$, $f(s)$ is holomorphic on $\{s \in \mathbb{C} : \Re(s) \geq \epsilon, -\pi / \log q \leq \Im(s) < \pi / \log q\}$ except for a simple pole at $s = 1$, where the residue is

$$\begin{aligned}
\operatorname{Res}_{s=1} f(s) &= \zeta_K(2) \prod_{v \in M(K)} (1 - 2q^{-2 \deg v} + q^{-3 \deg v}) \operatorname{Res}_{s=1} \zeta_K(s) \\
&= \frac{J_K q^{1-g_K} \zeta_K(2)}{(q-1) \log q} \prod_{v \in M(K)} (1 - 2q^{-2 \deg v} + q^{-3 \deg v})
\end{aligned}$$

(see [Weil 1974, Chapter VII], for example, for the residue of the zeta function). Now by a Tauberian argument (see [Rosen 2002, Theorem 17.1], for example)

$$\sum_{\substack{\mathfrak{C} \geq 0 \\ \deg \mathfrak{C} = j}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} = q^j \frac{J_K q^{1-g_K} \zeta_K(2)}{q-1} \prod_{v \in M(K)} (1 - q^{-2 \deg v} + q^{-3 \deg v}) + O(M q^{\epsilon j}),$$

where the implicit constant is absolute and

$$M = \max_{\Re(s)=\epsilon} |f(s)|,$$

which is clearly bounded above by a constant that depends only on K and ϵ. Therefore

$$\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \prod_{v\in\operatorname{Supp}\mathfrak{C}} (1+q^{-\deg v})^{-1} = q^j \frac{J_K q^{1-g_K}\zeta_K(2)}{q-1} \prod_{v\in M(K)} (1-q^{-2\deg v}+q^{-3\deg v}) + O(q^{\epsilon j}), \tag{26}$$

where the implicit constant depends only on K and ϵ.

We may assume that $N_K(m)>0$. For the remainder of the proof, all implicit constants depend only on K and ϵ. Fix an index j between 0 and $m-g_K$ and an $\epsilon>0$. Then by Lemma 1 (separating out those divisors of degree $2j$ that are twice an effective divisor and those that aren't)

$$\begin{aligned} N_K(m)a_{2j} &= \sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=2j}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}) \\ &= \sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/2\mathfrak{C}) + \sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\operatorname{Div}K\\ \deg\mathfrak{C}=2j}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/\mathfrak{C}). \end{aligned} \tag{27}$$

Now, by (0) and Proposition 1,

$$\begin{aligned} \sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}} \chi(F/2\mathfrak{C}) &= N'_K(m) \sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \prod_{v\in\operatorname{Supp}\mathfrak{C}} (1+q^{-\deg v})^{-1} \\ &\quad + \begin{cases} O(q^{(1/2+\epsilon)m}q^{(1+\epsilon)j}) & \text{if } q \text{ is odd,}\\ O(q^m q^j + q^{\epsilon m}q^{(1+\epsilon)j}) & \text{if } q \text{ is even.}\end{cases} \end{aligned} \tag{28}$$

Using the estimate for $|N'_K(m)-N_K(m)|$ in Proposition 1 and (26), we get

$$\begin{aligned} N'_K(m) &\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \prod_{v\in\operatorname{Supp}\mathfrak{C}} (1+q^{-\deg v})^{-1} \\ &= N_K(m) \sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} \prod_{v\in\operatorname{Supp}\mathfrak{C}} (1+q^{-\deg v})^{-1} + \begin{cases} O(q^{(1/2+\epsilon)m}q^j) & \text{if } q \text{ is odd,}\\ O(q^m q^j) & \text{if } q \text{ is even.}\end{cases} \end{aligned} \tag{29}$$

Combining (28), (29) and (9) yields

$$(30)\quad (N_K(m))^{-1}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}}\sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}}\chi(F/2\mathfrak{C})$$
$$=\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}+\begin{cases}O(q^{-(3/2-\epsilon)m}q^{(1+\epsilon)j}) & \text{if } q \text{ is odd},\\ O(q^{-m}q^{j}) & \text{if } q \text{ is even}.\end{cases}$$

Using (0) in conjunction with Propositions 2 and 3, we get

$$\sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\operatorname{Div}K\\ \deg\mathfrak{C}=2j}}\sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}}\chi(F/\mathfrak{C})=\begin{cases}O(q^{m}q^{(5/4+\epsilon)2j}) & \text{if } q \text{ is odd},\\ O(q^{m}q^{(1+\epsilon)2j}) & \text{if } q \text{ is even}.\end{cases}$$

Combining this with (9) yields

$$(31)\quad (N_K(m))^{-1}\sum_{\substack{\mathfrak{C}\geq 0\\ \mathfrak{C}\notin 2\operatorname{Div}K\\ \deg\mathfrak{C}=2j}}\sum_{\substack{[F:K]=2\\ g_F=m,\, q_F=q}}\chi(F/\mathfrak{C})=\begin{cases}O(q^{-m}q^{(5/4+\epsilon)2j}) & \text{if } q \text{ is odd},\\ O(q^{-m}q^{(1+\epsilon)2j}) & \text{if } q \text{ is even}.\end{cases}$$

Finally, by (27), (30) and (31),

$$a_{2j}=\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}+\begin{cases}O(q^{-m}q^{(5/4+\epsilon)2j}) & \text{if } q \text{ is odd},\\ O(q^{-m}q^{(1+\epsilon)2j}) & \text{if } q \text{ is even}.\end{cases}$$

This completes the proof of Theorem 2. □

Proof of Corollary 1. Set $m'=m-g_K$ in what follows for notational convenience. We first note that $a_{2j}=O(q^j)$ for all $j=0,\dots,m'$ by Theorem 2, where the implicit constant depends only on K. Let $x\leq 1$ to be chosen later. Then, whenever $\Re(s)>\frac{1}{2}$,

$$(32)\quad M_K(s,m,1)=\sum_{j\leq xm'}a_{2j}q^{-2sj}+O\biggl(\sum_{j>xm'}q^{-j(2\Re(s)-1)}\biggr)$$
$$=\sum_{j\leq xm'}a_{2j}q^{-2sj}+O\biggl(\frac{q^{-xm'(2\Re(s)-1)}}{1-q^{1-2\Re(s)}}\biggr),$$

where the implicit constants depend only on K. Also, by Theorem 2, for any $\delta>0$,

$$(33)\quad \sum_{j\leq xm'}a_{2j}q^{-2sj}=\sum_{j\leq xm'}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}}q^{-2s\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}$$
$$+\begin{cases}O\bigl(q^{-m}\sum_{j\leq xm'}q^{2j(5/4+\delta-\Re(s))}\bigr) & \text{if } q \text{ is odd},\\ O\bigl(q^{-m}\sum_{j\leq xm'}q^{2j(1+\delta-\Re(s))}\bigr) & \text{if } q \text{ is even},\end{cases}$$

$$= \sum_{j \le xm'} \sum_{\substack{\mathfrak{C} \ge 0 \\ \deg \mathfrak{C} = j}} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} + \begin{cases} O(q^{-m}(1 + q^{2xm'(5/4+\delta-\Re(s))})) & \text{if } q \text{ is odd,} \\ O(q^{-m}(1 + q^{2xm'(1+\delta-\Re(s))})) & \text{if } q \text{ is even,} \end{cases}$$

where the implicit constants depend only on K and δ. We may assume that $\epsilon \le \frac{1}{6}$. We now choose x and δ such that

$$x = \begin{cases} \dfrac{1}{3/2+2\delta} = \dfrac{2}{3} - \epsilon & \text{if } q \text{ is odd,} \\ \dfrac{1}{1+2\delta} = 1 - \epsilon & \text{if } q \text{ is even,} \end{cases}$$

so that

$$-xm'(2\Re(s) - 1) = \begin{cases} -m' + 2xm'\left(\frac{5}{4} + \delta - \Re(s)\right) & \text{if } q \text{ is odd,} \\ -m' + 2xm'(1 + \delta - \Re(s)) & \text{if } q \text{ is even.} \end{cases}$$

Then, by (32), (33) and the definition of m',

$$M_K(s, m, 1) = \sum_{j \le xm'} \sum_{\substack{\mathfrak{C} \ge 0 \\ \deg \mathfrak{C} = j}} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} + \begin{cases} O\left(\dfrac{q^{-m(2/3-\epsilon)(2\Re(s)-1)}}{(1 - q^{1-2\Re(s)})}\right) & \text{if } q \text{ is odd,} \\ O\left(\dfrac{q^{-m(1-\epsilon)(2\Re(s)-1)}}{(1 - q^{1-2\Re(s)})}\right) & \text{if } q \text{ is even,} \end{cases} \tag{34}$$

where the implicit constants depend only on K and ϵ. Also, by Theorem 2,

$$\begin{aligned} \sum_{j \le xm'} \sum_{\substack{\mathfrak{C} \ge 0 \\ \deg \mathfrak{C} = j}} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} &= \sum_{\mathfrak{C} \ge 0} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} + O\left(\sum_{j > xm'} q^{-j(2\Re(s)-1)}\right) \\ &= \sum_{\mathfrak{C} \ge 0} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} + \begin{cases} O\left(\dfrac{q^{-m(2/3-\epsilon)(2\Re(s)-1)}}{(1 - q^{1-2\Re(s)})}\right) & \text{if } q \text{ is odd,} \\ O\left(\dfrac{q^{-m(1-\epsilon)(2\Re(s)-1)}}{(1 - q^{1-2\Re(s)})}\right) & \text{if } q \text{ is even.} \end{cases} \end{aligned} \tag{35}$$

Finally, by Lemma 3,

$$\sum_{\mathfrak{C} \ge 0} q^{-2s \deg \mathfrak{C}} \prod_{v \in \operatorname{Supp} \mathfrak{C}} (1 + q^{-\deg v})^{-1} = \sigma_K(s, 1). \tag{36}$$

Corollary 1 follows from (34)–(36). □

Proof of Corollary 2. Set $m' = m - g_K$ again. By Theorem 2,

$$\sum_{j\geq m'/2}\Bigg(\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}\Bigg)-C(K) = O\Bigg(\sum_{j\geq m'/2} q^{-j(1-2\epsilon)}\Bigg) = O(q^{-m'(1/2-\epsilon)}), \tag{37}$$

where the implicit constant depends only on K and ϵ. This shows that the series $C'(K)$ converges. Also, by Theorem 2,

$$M_K\big(\tfrac{1}{2}, m, 1\big) = \begin{cases} 2\sum_{j<m'/2} a_{2j}q^{-j} & \text{if } m' \text{ is odd,}\\ 2\sum_{j<m'/2} a_{2j}q^{-j} + a_{m'}q^{-m'/2} & \text{if } m' \text{ is even,}\end{cases} \tag{38}$$

and

$$\begin{aligned} 2\sum_{j<m'/2} a_{2j}q^{-j} &= 2\sum_{j<m'/2}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1} + O\Bigg(\sum_{j<m'/2} q^{-m'}q^{j(3/2+2\epsilon)}\Bigg)\\ &= 2\sum_{j<m'/2}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1} + O(q^{-m'(1/4-\epsilon)}) \end{aligned} \tag{39a}$$

if q is odd. If q is even, similar estimates give

$$2\sum_{j<m'/2} a_{2j}q^{-j} = 2\sum_{j<m'/2}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1} + O(q^{-m'(1/2-\epsilon)}). \tag{39b}$$

Now, by (37),

$$\begin{aligned} &2\sum_{j<m'/2}\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}\\ &= 2C'(K) + 2\sum_{j<m'/2} C(K) - 2\sum_{j\geq m'/2}\Bigg(\sum_{\substack{\mathfrak{C}\geq 0\\ \deg\mathfrak{C}=j}} q^{-\deg\mathfrak{C}}\prod_{v\in\operatorname{Supp}\mathfrak{C}}(1+q^{-\deg v})^{-1}\Bigg)-C(K)\\ &= 2C'(K) + O(q^{-m'(1/2-\epsilon)}) + \begin{cases} m'C(K) & \text{if } m' \text{ is even,}\\ (m'+1)C(K) & \text{if } m' \text{ is odd.}\end{cases} \end{aligned} \tag{40}$$

Finally, if m' is even, Theorem 2 gives

$$a_{m'}q^{-m'/2} = C(K) + O(q^{-m'(1/2-\epsilon)}). \tag{41}$$

The remainder of Corollary 2 follows from (38)–(41). □

Proof of Corollary 3. It is well known that $\#\{v \in M(F) : \deg v = 1\} - q - 1$ is equal to the coefficient of q^{-s} in the polynomial $L_F(q^{-s})$ for all function fields F with $q_F = q$. (See [Stichtenoth 1993, Theorem V.1.15], for example.) By Theorem 2, the coefficient of q^{-s} in the polynomial $L_K(q^{-s})M_K(s, m, 1)$ is just the coefficient of q^{-s} in the polynomial $L_K(q^{-s})$. □

References

[Andrade and Keating 2012] J. C. Andrade and J. P. Keating, "The mean value of $L(\frac{1}{2}, \chi)$ in the hyperelliptic ensemble", *J. Number Theory* **132**:12 (2012), 2793–2816. MR 2965192 Zbl 1278.11082

[Chinta et al. 2006] G. Chinta, S. Friedberg, and J. Hoffstein, "Multiple Dirichlet series and automorphic forms", pp. 3–41 in *Multiple Dirichlet series, automorphic forms, and analytic number theory* (Bretton Woods, NH, 2005), edited by S. Friedberg et al., Proc. Sympos. Pure Math. **75**, American Mathematical Society, Providence, RI, 2006. MR 2008c:11072 Zbl 1124.11023

[Fisher and Friedberg 2004] B. Fisher and S. Friedberg, "Double Dirichlet series over function fields", *Compos. Math.* **140**:3 (2004), 613–630. MR 2005a:11183 Zbl 1082.11053

[Goldfeld and Hoffstein 1985] D. Goldfeld and J. Hoffstein, "Eisenstein series of $\frac{1}{2}$-integral weight and the mean value of real Dirichlet L-series", *Invent. Math.* **80**:2 (1985), 185–208. MR 86m:11029 Zbl 0564.10043

[Hoffstein and Rosen 1992] J. Hoffstein and M. Rosen, "Average values of L-series in function fields", *J. Reine Angew. Math.* **426** (1992), 117–150. MR 93c:11022 Zbl 0754.11036

[Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Mathematics **210**, Springer, New York, 2002. MR 2003d:11171 Zbl 1043.11079

[Siegel 1944] C. L. Siegel, "The average measure of quadratic forms with given determinant and signature", *Ann. of Math.* (2) **45** (1944), 667–685. MR 7,51a Zbl 0063.07007

[Stichtenoth 1993] H. Stichtenoth, *Algebraic function fields and codes*, Graduate Texts in Mathematics **254**, Springer, Berlin, 1993. MR 94k:14016 Zbl 0816.14011

[Thunder 2013] J. L. Thunder, "On sums involving products and quotients of L-functions over function fields", preprint, 2013. arXiv 1310.8572

[Thunder and Widmer 2013] J. L. Thunder and M. Widmer, "Counting points of fixed degree and given height over function fields", *Bull. Lond. Math. Soc.* **45**:2 (2013), 283–300. MR 3064414 Zbl 06160957

[Weil 1974] A. Weil, *Basic number theory*, 3rd ed., Die Grundlehren der Mathematischen Wissenschaften **144**, Springer, New York, 1974. MR 55 #302 Zbl 0326.12001

Received December 10, 2013. Revised July 10, 2014.

JEFFREY LIN THUNDER
DEPARTMENT OF MATHEMATICAL SCIENCES
NORTHERN ILLINOIS UNIVERSITY
DEKALB, IL 60115-2874
UNITED STATES

jthunder@math.niu.edu

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 274 No. 1 March 2015