Algebra & Number Theory

Volume 11
2017
No. 3

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2017 is US $325/year for the electronic version, and $520/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.505

The umbral moonshine module for the unique unimodular Niemeier root system

John F. R. Duncan and Jeffrey A. Harvey

We use canonically twisted modules for a certain super vertex operator algebra to construct the umbral moonshine module for the unique Niemeier lattice that coincides with its root sublattice. In particular, we give explicit expressions for the vector-valued mock modular forms attached to automorphisms of this lattice by umbral moonshine. We also characterize the vector-valued mock modular forms arising, in which four of Ramanujan's fifth-order mock theta functions appear as components.

1. Introduction

In his Ph.D thesis, Zwegers [2002] gave an intrinsic definition of mock theta functions and provided new insight into three families of such functions, constructed

(1) in terms of Appell–Lerch sums,

(2) as the Fourier coefficients of meromorphic Jacobi forms, and

(3) via theta functions attached to cones in lattices of indefinite signature.

The first two constructions have played a central role in recently observed moonshine connections between finite groups and mock theta functions. These started with the observation in [Eguchi et al. 2011] that the elliptic genus of a K3 surface has a decomposition into characters of the $N=4$ superconformal algebra with multiplicities that at low levels are equal to the dimensions of irreducible representations of the Mathieu group M_{24}. Appell–Lerch sums appear in this analysis in the so called "massless" characters. This Mathieu moonshine connection was conjectured in [Cheng et al. 2014a; 2014b] to be part of a much more general phenomenon, known as umbral moonshine, which attaches a vector-valued mock modular form H^X, a finite group G^X, and an infinite-dimensional graded G^X-module K^X to the root systems of each of the 23 Niemeier lattices. The analysis in [Cheng et al. 2014b] relied heavily on the construction of mock modular forms in terms of meromorphic Jacobi forms and built on the important work in [Dabholkar et al. 2012] extending

MSC2010: primary 17B69; secondary 11F22, 11F37.
Keywords: umbral moonshine, mock theta function, vertex operator algebra.

the analysis of [Zwegers 2002] and characterizing special Jacobi forms in terms of growth conditions.

Whilst the existence of the G^X-modules K^X has now been proven [Gannon 2016; Duncan et al. 2015b] for all Niemeier root systems X, no explicit construction of the modules K^X is yet known.

In this paper we construct the G^X-module K^X for the case that $X = E_8^3$. To do so we apply the third characterization of mock theta functions in terms of indefinite theta functions. This enables us to employ the formalism of vertex operator algebras [Borcherds 1986; Frenkel et al. 1988], which has been so fruitfully employed (in [Frenkel et al. 1988; Borcherds 1992] to name just two) in the understanding of monstrous moonshine [Conway and Norton 1979; Thompson 1979a; 1979b].

See [Duncan et al. 2015a] for a recent review of moonshine both monstrous and umbral, and many more references on these subjects.

To explain the methods of this paper in more detail, we first recall the *Pochammer symbol*

$$(x;q)_n := \prod_{k=0}^{n-1}(1-xq^k), \tag{1-1}$$

and the fifth-order mock theta functions

$$\chi_0(q) := \sum_{n\geq 0}\frac{q^n}{(q^{n+1};q)_n}, \qquad \chi_1(q) := \sum_{n\geq 0}\frac{q^n}{(q^{n+1};q)_{n+1}} \tag{1-2}$$

from Ramanujan's last letter to Hardy [Ramanujan 1988; 2000]. The conjectures of [Cheng et al. 2014b] (see also [Cheng et al. $\geq$ 2017]) imply the existence of a bigraded super vector space $K^X = \bigoplus_r K^X_r = \bigoplus_{r,d} K^X_{r,d}$ that is a module for $G^X \simeq S_3$ and satisfies

$$\begin{aligned}\operatorname{sdim}_q K^X_1 &= -2q^{-1/120} + \sum_{n>0} \dim K^X_{1,n-1/120}\, q^{n-1/120} \\ &= 2q^{-1/120}(\chi_0(q)-2), \\ \operatorname{sdim}_q K^X_7 &= \sum_{n>0} \dim K^X_{7,n-49/120}\, q^{n-49/120} = 2q^{71/120}\chi_1(q).\end{aligned} \tag{1-3}$$

Here $\operatorname{sdim}_q V := \sum_n (\dim(V_{\bar 0})_n - \dim(V_{\bar 1})_n)q^n$ for V a $\mathbb{Q}$-graded super space with even part $V_{\bar 0}$ and odd part $V_{\bar 1}$.

In this article we realize K^X explicitly in terms of canonically twisted modules for a super vertex operator algebra V^X, and we show that the graded trace functions for the resulting bigraded G^X-module are exactly compatible with the predictions of [Cheng et al. 2014b]. Thus we construct the analogue of the moonshine module $V^\natural$ of Frenkel, Lepowsky and Meurman [Frenkel et al. 1984; 1985; 1988], for the

$X = E_8^3$ case of umbral moonshine (although it should be noted that the group for us is $G^X \simeq S_3$, which is of course much less complicated than the monster).

To prove that our construction is indeed the $X = E_8^3$ counterpart to $V^\natural$, we verify the $X = E_8^3$ analogue of the Conway–Norton moonshine conjecture, proven by Borcherds [1992] in the case of the monster, which predicts that the trace functions arising are uniquely determined by their automorphy and their asymptotic behavior near cusps. Thus we verify the $X = E_8^3$ analogues of both of the two major conjectures of monstrous moonshine.

The main motivation for our construction of K^X stems from some q-series identities for $\chi_0(q)$ and $\chi_1(q)$ that were established by Zwegers [2009]. These are

$$\begin{aligned} 2-\chi_0(q) &= \frac{1}{(q;q)_\infty^2}\biggl(\sum_{k,l,m\geq 0}+\sum_{k,l,m<0}\biggr)(-1)^{k+l+m} \\ &\qquad\qquad \times q^{(k^2+l^2+m^2)/2+2(kl+lm+mk)+(k+l+m)/2}, \\ \chi_1(q) &= \frac{1}{(q;q)_\infty^2}\biggl(\sum_{k,l,m\geq 0}+\sum_{k,l,m<0}\biggr)(-1)^{k+l+m} \\ &\qquad\qquad \times q^{(k^2+l^2+m^2)/2+2(kl+lm+mk)+3(k+l+m)/2}, \end{aligned} \tag{1-4}$$

where $(x;q)_\infty := \prod_{n\geq 0}(1-xq^n)$. Upon regarding (1-4) we are led to consider a certain indefinite lattice of rank 3 (see Section 2D), which furnishes most of the vertex algebra structure that we use to define V^X. It is a curious circumstance that there seems to be no obvious relationship between the lattice we use and the $X = E_8^3$ structure which underpins the relevant case of umbral moonshine. For this reason it is not yet clear how the construction of K^X presented here should be generalized to the other cases of umbral moonshine.

We now formulate precise statements of our results. To prepare for this, recall that vector-valued functions $H_g^X(\tau) = (H_{g,r}^X(\tau))$ on the upper half plane $\mathbb{H}$ are considered in [Cheng et al. 2014b], for $g \in G^X \simeq S_3$, where the components are indexed by $r \in \mathbb{Z}/60\mathbb{Z}$. Define $o(g)$ to be the order of an element $g \in G^X$. The H_g^X are not uniquely determined in [Cheng et al. 2014b], except for the case that $g = e$ is the identity, $o(g) = 1$. But it is predicted that H_g^X is a mock modular form of weight $\frac{1}{2}$ for $\Gamma_0(o(g))$, with shadow given by a certain vector-valued unary theta function S_g^X (see (3-33)), and specified polar parts at the cusps of $\Gamma_0(o(g))$. In more detail, H_g^X should have the same polar parts as $H^X := H_e^X$ at the infinite cusp of $\Gamma_0(o(g))$, but should have vanishing polar parts at any noninfinite cusps. In practice, this amounts to the statement that we should have

$$\begin{aligned} &H_{g,r}^X(\tau) \\ &\quad = \begin{cases} \mp 2q^{-1/120} + O(q^{119/120}) & \text{if } r = \pm 1, \pm 11, \pm 19, \pm 29 \pmod{60}, \\ O(1) & \text{otherwise}, \end{cases} \end{aligned} \tag{1-5}$$

for $q = e^{2\pi i\tau}$, and all components of $H^X_g(\tau)$ should remain bounded as $\tau \to 0$, if $g \neq e$. (If $g \neq e$ then $o(g) = 2$ or $o(g) = 3$, and then $\Gamma_0(o(g))$ has only one cusp other than the infinite one, and this is the cusp represented by 0.)

Our main result is the following, where the functions T^X_g are defined in Section 3C (see (3-32)) in terms of traces of operators on canonically twisted modules for V^X.

Theorem 1.1. *Let $g \in G^X$. If $o(g) \neq 3$ then $2T^X_g$ is the Fourier expansion of the unique vector-valued mock modular form of weight $\frac{1}{2}$ for $\Gamma_0(o(g))$ whose shadow is S^X_g, and whose polar parts coincide with those of H^X_g. If $o(g) = 3$ then $2T^X_g$ is the Fourier expansion of the unique vector-valued modular form of weight $\frac{1}{2}$ for $\Gamma_0(3)$ which has the multiplier system $\rho_{3|3}\overline{\sigma^X}$ and polar parts coinciding with those of H^X_g.*

Here $\sigma^X : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_{60}(\mathbb{C})$ denotes the multiplier system of $S^X := S^X_e$ (see (3-34)), and $\rho_{3|3} : \Gamma_0(3) \to \mathbb{C}^\times$ is defined in (3-38). The shadow function S^X_g is defined in (3-33) and determines the multiplier system of $2T^X_g$ when $o(g) \neq 3$.

Armed with Theorem 1.1, we may now define the H^X_g concretely and explicitly, for $g \in G^X$, by setting

$$H^X_g(\tau) := 2T^X_g(\tau), \tag{1-6}$$

where $T^X_g(\tau)$ denotes the function obtained by substituting $e^{2\pi i\tau}$ for q in the series expression (3-32) for T^X_g.

Expressions for the components of H^X_g are given in §5.4 of [Cheng et al. 2014b], in terms of fifth-order mock theta functions of Ramanujan, for the cases that $o(g) = 1$ and $o(g) = 2$, but it is not verified there that these prescriptions define mock modular forms with the specified shadows. Our work confirms these statements, as the following theorem demonstrates.

Theorem 1.2. *We have the following identities*:

$$H^X_{1A,r}(\tau) = \begin{cases} \pm 2q^{-1/120}(\chi_0(q) - 2) & \text{if } r = \pm 1, \pm 11, \pm 19, \pm 29, \\ \pm 2q^{71/120}\chi_1(q) & \text{if } r = \pm 7, \pm 13, \pm 17, \pm 23, \end{cases} \tag{1-7}$$

$$H^X_{2A,r}(\tau) = \begin{cases} \mp 2q^{-1/120}\phi_0(-q) & \text{if } r = \pm 1, \pm 11, \pm 19, \pm 29, \\ \pm 2q^{-49/120}\phi_1(-q), & \text{if } r = \pm 7, \pm 13, \pm 17, \pm 23. \end{cases} \tag{1-8}$$

The fifth-order mock theta functions ϕ_0 and ϕ_1 were defined by Ramanujan (also in his last letter to Hardy) by setting

$$\phi_0(q) := \sum_{n \geq 0} q^{n^2}(-q; q^2)_n, \qquad \phi_1(q) := \sum_{n \geq 0} q^{(n+1)^2}(-q; q^2)_n. \tag{1-9}$$

The identities (1-7) follow immediately from Theorem 1.1, since the V^X-modules used to define the T^X_g have been constructed specifically so as to make Zwegers' identity (1-4) manifest. By contrast, the $o(g) = 2$ case of Theorem 1.2 requires some work, since the expressions we obtain naturally from our construction of T_g^X do not obviously coincide with (1-8). Thus the proof of Theorem 1.2 entails nontrivial q-series identities which may be of independent interest.

Corollary 1.3. *We have*

$$\Big(\sum_{k,m\geq 0} - \sum_{k,m<0}\Big)_{k=m \pmod 2} (-1)^m q^{k^2/2+m^2/2+4km+k/2+3m/2} = \prod_{n>0}(1+q^n)\Big(\sum_{k,m\geq 0} - \sum_{k,m<0}\Big)(-1)^{k+m} q^{3k^2+m^2/2+4km+k+m/2}, \tag{1-10}$$

$$\Big(\sum_{k,m\geq 0} - \sum_{k,m<0}\Big)_{k=m \pmod 2} (-1)^m q^{k^2/2+m^2/2+4km+3k/2+5m/2} = \prod_{n>0}(1+q^n)\Big(\sum_{k,m\geq 0} - \sum_{k,m<0}\Big)(-1)^{k+m} q^{3k^2+m^2/2+4km+3k+3m/2}. \tag{1-11}$$

The reader who is familiar with modularity results on trace functions attached to vertex operator algebras (see [Zhu 1996; Dong et al. 2000; Miyamoto 2004]) and super vertex operator algebras (see [Dong and Zhao 2005]) may find it surprising that the functions we construct are (generally) mock modular, rather than modular, and have weight $\frac{1}{2}$, rather than weight 0. In light of Zwegers' work [2002; 2009], it is clear that we can obtain trace functions with mock modular behavior by considering vertex algebras constructed according to the usual lattice vertex algebra construction, but with a cone (see Section 2D) taking on the role usually played by a lattice. A suitably chosen cone is the main ingredient for our construction of V^X.

Note however that the cone vertex algebra construction does not, on its own, naturally give rise to trace functions with weight $\frac{1}{2}$. For this we introduce a single "free fermion" to the cone vertex algebra that we use to construct V^X, and we insert the zero mode (i.e., $L(0)$-degree preserving component) of the canonically twisted vertex operator attached to a generator when we compute graded traces on canonically twisted modules for V^X. In practice, this has the effect of multiplying the cone vertex algebra trace functions by $\eta(\tau) := q^{1/24}\prod_{n>0}(1-q^n)$.

We remark that this technique may be profitably applied to other situations. For example, it is known (see, e.g., [Harada and Lang 1998]) that the moonshine module $V^\natural$, when regarded as a module for the Virasoro algebra, is a direct sum of modules $L(h, 24)$, for h ranging over nonnegative integers, satisfying

$$\operatorname{tr}_{L(h,24)} q^{L(0)-c/24} = \begin{cases} (1-q)q^{-23/24}\eta(\tau)^{-1} & \text{for } h=0, \\ q^{h-23/24}\eta(\tau)^{-1} & \text{for } h>0, \end{cases} \tag{1-12}$$

where $c = 24$. Also, the multiplicity of $L(0, 24)$ is 1, and the multiplicity of $L(1, 24)$ is 0. Consequently, the weight $\frac{1}{2}$ modular form $\eta(\tau)J(\tau)$, with $J(\tau) = q^{-1} + O(q)$

the (so normalized) elliptic modular invariant, is almost the generating function of the dimensions of the homogeneous spaces of Virasoro highest weight vectors in $V^\natural$. Indeed, the actual generating function is just $q^{1/24}\eta(\tau)J(\tau)+1$.

Certainly $\eta(\tau)J(\tau)$ has nicer modular properties than the Virasoro highest weight generating function of $V^\natural$, and moreover, an even more striking connection to the monster, as four of the dimensions of nontrivial irreducible representations for the monster appear as coefficients:

$$\eta(\tau)J(\tau) = \cdots + 196883q^{25/24} + 21296876q^{49/24} + 842609326q^{73/24} + 19360062527q^{97/24} + \cdots \quad \text{(1-13)}$$

(see p. 220 of [Conway et al. 1985]). This function $\eta(\tau)J(\tau)$ can be obtained naturally as a trace function on a canonically twisted module for a super vertex operator algebra. Indeed, if we take V to be the tensor product of $V^\natural$ with the super vertex operator algebra obtained by applying the Clifford module construction to a one-dimensional vector space (see Section 2C for details), then, choosing an irreducible canonically twisted module V_{tw} for V, and denoting by $p(0)$ the coefficient of z^{-1} in the canonically twisted vertex operator attached to a suitably scaled element $p \in V$ with $L(0)p = \frac{1}{2}p$, we have

$$\operatorname{tr}_{V_{\mathrm{tw}}} p(0)q^{L(0)-c/24} = \eta(\tau)J(\tau), \quad \text{(1-14)}$$

where now $c = \frac{49}{2}$. (See Section 2C for more detail.)

The importance of trace functions such as (1-14) within the broader context of modularity for super vertex operator algebras is analyzed in detail in [Van Ekeren 2013]; see also [Van Ekeren 2014].

The organization of the paper is as follows. In Section 2 we recall some familiar constructions from the theory of vertex algebras and use these to construct the super vertex operator algebra V^X and its canonically twisted modules $V^{X,\pm}_{\mathrm{tw},a}$, which play the commanding role in this work. We recall the lattice construction of super vertex algebras in Section 2A, modules for lattice super vertex algebras in Section 2B, and the Clifford module super vertex algebra construction in Section 2C. We formulate the construction of V^X and the twisted modules $V^{X,\pm}_{\mathrm{tw},a}$ in Section 2D. We also equip these spaces with G^X-module structure in Section 2D, and compute explicit expressions (see Proposition 2.2) for the graded traces of elements of G^X.

In Section 3 our focus moves from representation theory to number theory, as we seek to determine the properties of the graded traces arising from the action of G^X on the $V^{\pm}_{\mathrm{tw},a}$. We recall the relationship between mock modular forms and harmonic Maass forms in Section 3A, and we recall some results on Zwegers' indefinite theta series in Section 3B. The proofs of our main results, Theorems 1.1 and 1.2, are the content of Section 3C.

We give tables with the first few coefficients of the H^X_g in the Appendix. We frequently employ the notational convention $e(x) := e^{2\pi i x}$.

2. Vertex algebra

This section begins with a review of the lattice (super) vertex algebra construction in Section 2A, and the natural generalization of this, which defines lattice vertex algebra modules, in Section 2B. We review the special case of the Clifford module super vertex algebra construction we require in Section 2C. We put all of this together for the construction of V^X, and its canonically twisted modules, in Section 2D.

2A. ***Lattice vertex algebra.*** We briefly recall, following [Borcherds 1986; Frenkel et al. 1988], the standard construction which associates a super vertex algebra V_L to a central extension of an integral lattice L. We also employ [Frenkel and Ben-Zvi 2004] as a reference. Set $\mathfrak{h} := L \otimes_{\mathbb{Z}} \mathbb{C}$, and extend the bilinear form on L to a symmetric $\mathbb{C}$-bilinear form on $\mathfrak{h}$ in the natural way. Set $\hat{\mathfrak{h}} := \mathfrak{h}[t, t^{-1}] \oplus \mathbb{C}\boldsymbol{c}$, for t a formal variable, and define a Lie algebra structure on $\hat{\mathfrak{h}}$ by declaring that $\boldsymbol{c}$ is central, and $[u \otimes t^m, v \otimes t^n] = m\langle u, v\rangle \delta_{m+n,0}\, \boldsymbol{c}$ for $u, v \in \mathfrak{h}$ and $m, n \in \mathbb{Z}$. We follow tradition and write $u(m)$ as a shorthand for $u \otimes t^m$. The Lie algebra $\hat{\mathfrak{h}}$ has a triangular decomposition $\hat{\mathfrak{h}} = \hat{\mathfrak{h}}^- \oplus \hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+$, where $\hat{\mathfrak{h}}^\pm := \mathfrak{h}[t^{\pm 1}]t^{\pm 1}$ and $\hat{\mathfrak{h}}^0 := \mathfrak{h} \oplus \mathbb{C}\boldsymbol{c}$.

We require a bilinear function $b : L \times L \to \mathbb{Z}/2\mathbb{Z}$ with the property that

$$b(\lambda, \mu) + b(\mu, \lambda) = \langle \lambda, \mu\rangle + \langle \lambda, \lambda\rangle\langle \mu, \mu\rangle + 2\mathbb{Z}.$$

If $\{\varepsilon_i\}$ is an ordered $\mathbb{Z}$-basis for L then we may take b to be the unique such function for which

$$b(\varepsilon_i, \varepsilon_j) = \begin{cases} 0 + 2\mathbb{Z} & \text{when } i \le j, \\ \langle \lambda, \mu\rangle + \langle \lambda, \lambda\rangle\langle \mu, \mu\rangle + 2\mathbb{Z} & \text{when } i > j. \end{cases} \tag{2-1}$$

Set $\beta(\lambda, \mu) := (-1)^{b(\lambda,\mu)}$, and define $\mathbb{C}_\beta[L]$ to be the ring generated by symbols $\boldsymbol{v}_\lambda$ for $\lambda \in L$ subject to the relations $\boldsymbol{v}_\lambda \boldsymbol{v}_\mu = \beta(\lambda, \mu)\boldsymbol{v}_{\lambda+\mu}$.

Remark 2.1. The algebra $\mathbb{C}_\beta[L]$ is isomorphic to the quotient $\mathbb{C}[\hat{L}]/\langle \kappa + 1\rangle$, where $\hat{L}$ is the unique (up to isomorphism) central extension of L by $\langle\kappa\rangle \simeq \mathbb{Z}/2\mathbb{Z}$ such that

$$aa' = \kappa^{\langle \bar{a}, \bar{a}'\rangle + \langle \bar{a}, \bar{a}\rangle\langle \bar{a}', \bar{a}'\rangle} a' a, \tag{2-2}$$

for $a, a' \in \hat{L}$ lying above $\bar{a}, \bar{a}' \in L$, respectively. See [Frenkel et al. 1988].

Now define an $\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+$-module structure on $\mathbb{C}_\beta[L]$ by setting $\boldsymbol{c}\boldsymbol{v}_\lambda = \boldsymbol{v}_\lambda$ and $u(m)\boldsymbol{v}_\lambda = \delta_{m,0}\langle u, \lambda\rangle \boldsymbol{v}_\lambda$ for $u \in \mathfrak{h}$ and $\lambda \in L$, and define V_L to be the induced $\hat{\mathfrak{h}}$-module,

$$V_L := U(\hat{\mathfrak{h}}) \otimes_{U(\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+)} \mathbb{C}_\beta[L]. \tag{2-3}$$

Then, according to §5.4.2 of [Frenkel and Ben-Zvi 2004], for example, V_L admits a unique super vertex algebra structure $Y : V_L \to (\operatorname{End} V_L)[[z, z^{-1}]]$ such that $1 \otimes \boldsymbol{v}_0$ is the vacuum vector,

$$Y(u(-1) \otimes \boldsymbol{v}_0, z) = \sum_{n \in \mathbb{Z}} u(n) z^{-n-1} \tag{2-4}$$

for $u \in \mathfrak{h}$, and

$$Y(1 \otimes \boldsymbol{v}_\lambda, z) = \exp\biggl(-\sum_{n<0} \frac{\lambda(n)}{n} z^{-n}\biggr) \exp\biggl(-\sum_{n>0} \frac{\lambda(n)}{n} z^{-n}\biggr) \boldsymbol{v}_\lambda z^{\lambda(0)} \tag{2-5}$$

for $\lambda \in L$. Here $\boldsymbol{v}_\lambda$ denotes the operator $v \otimes \boldsymbol{v}_\mu \mapsto \beta(\lambda, \mu) v \otimes \boldsymbol{v}_{\lambda+\mu}$, and $z^{\lambda(0)}(v \otimes \boldsymbol{v}_\mu)$ is $(v \otimes \boldsymbol{v}_\mu) z^{\langle \lambda, \mu \rangle}$. Note that we have

$$V_L \simeq S(\hat{\mathfrak{h}}^-) \otimes \mathbb{C}[L] \tag{2-6}$$

as modules for $\hat{\mathfrak{h}}^- \oplus \hat{\mathfrak{h}}^0$.

Given that $\{\varepsilon_i\}$ is a basis for L, choose $\varepsilon_i' \in L \otimes_{\mathbb{Z}} \mathbb{Q}$ such that $\langle \varepsilon_i', \varepsilon_j \rangle = \delta_{i,j}$, and define

$$\omega := \frac{1}{2} \sum_{i=1}^{3} \varepsilon_i'(-1) \varepsilon_i(-1) \otimes \boldsymbol{v}_0. \tag{2-7}$$

Then ω is a conformal element for V_L with central charge equal to the rank of L. If we define $L(n) \in \operatorname{End} V_L$ so that $Y(\omega, z) = \sum_{n \in \mathbb{Z}} L(n) z^{-n-2}$, then we have $[L(0), v(n)] = -n v(n)$ and $1 \otimes \boldsymbol{v}_\lambda$ is an eigenvector for $L(0)$ with eigenvalue $\frac{1}{2}\langle \lambda, \lambda \rangle$. Note that we do not assume that the bilinear form on L is positive-definite. Vectors of nonpositive length in L give rise to infinite-dimensional eigenspaces for $L(0)$, so in general $(V_L, Y, \boldsymbol{v}_0, \omega_u)$ is a conformal super vertex algebra, but not a super vertex operator algebra.

Automorphisms of L can be lifted to automorphisms of V_L. Indeed, suppose given $g \in \operatorname{Aut}(L)$ and a function $\alpha : L \to \{\pm 1\}$ satisfying

$$\alpha(\lambda + \mu) \beta(\lambda, \mu) = \alpha(\lambda) \alpha(\mu) \beta(g\lambda, g\mu) \tag{2-8}$$

for $\lambda, \mu \in L$. Then we obtain an automorphism $\hat{g}$ of $\operatorname{Aut}(V_L)$ by setting

$$\hat{g}(v \otimes \boldsymbol{v}_\lambda) := \alpha(\lambda)(g \cdot v) \otimes \boldsymbol{v}_{g\lambda} \tag{2-9}$$

for $v \in S(\hat{\mathfrak{h}}^-)$ and $\lambda \in L$, where $g \cdot v$ denotes the natural extension of the action of $\operatorname{Aut}(L)$ on $\mathfrak{h} = L \otimes_{\mathbb{Z}} \mathbb{C}$ to $S(\hat{\mathfrak{h}}^-)$, determined by $g \cdot u(m) = (gu)(m)$ for $u \in \mathfrak{h}$.

2B. ***Lattice vertex algebra modules.*** Let γ be an element of the dual lattice

$$L^* := \{\lambda \in L \otimes_{\mathbb{Z}} \mathbb{Q} \mid \langle \lambda, L \rangle \subset \mathbb{Z}\}.$$

Define $\mathbb{C}_\beta[L+\gamma]$ to be the complex vector space generated by symbols $\boldsymbol{v}_{\mu+\gamma}$ for $\mu \in L$, regarded as a $\mathbb{C}_\beta[L]$-module according to the rule $\boldsymbol{v}_\lambda \cdot \boldsymbol{v}_{\mu+\gamma} = \beta(\lambda, \mu)\boldsymbol{v}_{\lambda+\mu+\gamma}$. Equip $\mathbb{C}_\beta[L+\gamma]$ with a $U(\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+)$-module structure much as before, by letting $\boldsymbol{c}\boldsymbol{v}_{\mu+\gamma} = \boldsymbol{v}_{\mu+\gamma}$ and $u(m)\boldsymbol{v}_{\mu+\gamma} = \delta_{m,0}\langle u, \mu+\gamma\rangle \boldsymbol{v}_{\mu+\gamma}$ for $u \in \mathfrak{h}$ and $\mu \in L$. Let $V_{L+\gamma}$ be the $\hat{\mathfrak{h}}$-module defined by setting $V_{L+\gamma} := U(\hat{\mathfrak{h}}) \otimes_{U(\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+)} \mathbb{C}_\beta[L+\gamma]$. Then we have an isomorphism

$$V_{L+\gamma} \simeq S(\hat{\mathfrak{h}}^-) \otimes \mathbb{C}[L+\gamma] \tag{2-10}$$

of modules for $\hat{\mathfrak{h}}^-$. Define vertex operators $Y_\gamma : V_L \to (\operatorname{End} V_{L+\gamma})[\![z, z^{-1}]\!]$ using the same formulas as before, but interpret the operator $\boldsymbol{v}_\lambda$ in (2-5) as $\boldsymbol{v}_\lambda(v \otimes \boldsymbol{v}_{\mu+\gamma}) := \beta(\lambda, \mu) v \otimes \boldsymbol{v}_{\lambda+\mu+\gamma}$, according to the $\mathbb{C}_\beta[L]$-module structure on $\mathbb{C}_\beta[L+\gamma]$ prescribed above. Note that the construction of $V_{L+\gamma}$ depends upon the choice of coset representative $\gamma \in L^*$, so that $V_{L+\gamma}$ might be different from $V_{L+\gamma'}$, as a $\mathbb{C}_\beta[L]$-module, for example, even when $L+\gamma = L+\gamma'$, but different choices of coset representative are guaranteed to define isomorphic V_L-modules according to [Dong 1993].

The construction just described may be generalized so as to realize certain twisted modules for V_L. We give a brief description here, and refer to §3 of [Dong and Mason 1994] for more details.

Choose a vector $h \in \mathfrak{h}$. Then for $v \in S(\hat{\mathfrak{h}}^-)$ and $\lambda \in L$ we have $h(0)v \otimes \boldsymbol{v}_\lambda = \langle h, \lambda\rangle v \otimes \boldsymbol{v}_\lambda$. So if h is chosen to lie in $L \otimes_{\mathbb{Z}} \mathbb{Q}$, then

$$g_h := e^{2\pi i h(0)} \tag{2-11}$$

is a finite order automorphism of V_L, which acts as multiplication by $e^{2\pi i\langle h,\lambda\rangle}$ on the vector $v \otimes \boldsymbol{v}_\lambda$. The kernel of the map $L \otimes_{\mathbb{Z}} \mathbb{Q} \to \operatorname{Aut}(V_L)$ given by $h \mapsto g_h$ is exactly L^*, so $(L \otimes_{\mathbb{Z}} \mathbb{Q})/L^*$ is naturally a group of automorphisms of V_L. We may construct all the corresponding twisted modules for V_L explicitly.

To do this choose an h in $L \otimes_{\mathbb{Z}} \mathbb{Q}$ and let $\mathbb{C}[L+h]$ be the complex vector space generated by symbols $\boldsymbol{v}_{\lambda+h}$ for $\lambda \in L$. Just as before, we define a $U(\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+)$-module structure on $\mathbb{C}[L+h]$ by setting $\boldsymbol{c}\boldsymbol{v}_\mu = \boldsymbol{v}_\mu$ and $u(m)\boldsymbol{v}_\mu = \delta_{m,0}\langle u, \mu\rangle \boldsymbol{v}_\mu$ for $u \in \mathfrak{h}$ and $\mu \in L+h$. Define also the $\hat{\mathfrak{h}}$-module $V_{L+h} := U(\hat{\mathfrak{h}}) \otimes_{U(\hat{\mathfrak{h}}^0 \oplus \hat{\mathfrak{h}}^+)} \mathbb{C}[L+h]$, so that we have an isomorphism

$$V_{L+h} \simeq S(\hat{\mathfrak{h}}^-) \otimes \mathbb{C}[L+h] \tag{2-12}$$

of modules for $\hat{\mathfrak{h}}^-$. Taking M to be a positive integer such that $Mh \in L^*$, define vertex operators $Y_h : V_L \to (\operatorname{End} V_{L+h})[\![z^{1/M}, z^{-1/M}]\!]$ using the same formulas as before, but interpret the operator $\boldsymbol{v}_\lambda$ in (2-5) as $\boldsymbol{v}_\lambda(v \otimes \boldsymbol{v}_{\mu+h}) = \beta(\lambda, \mu) v \otimes \boldsymbol{v}_{\lambda+\mu+h}$. Then $V_{L+h} = (V_{L+h}, Y_h)$ is an irreducible g_h-twisted module for V_L, and any g_h-twisted module for V_L is of the form $V_{L+h'}$ for some $h' \in L \otimes_{\mathbb{Z}} \mathbb{Q}$ that is congruent to h modulo L^*.

Note that the action of $L\otimes_{\mathbb{Z}}\mathbb{Q}$ on V_L, given by $h\mapsto g_h$, extends to the $g_{h'}$-twisted module $V_{L+h'}$, for $h'\in L\otimes_{\mathbb{Z}}\mathbb{Q}$. For given $h,h'\in L\otimes_{\mathbb{Z}}\mathbb{Q}$, we may define

$$g_h(v\otimes \boldsymbol{v}_{\lambda+h'}) := e^{2\pi i\langle h,\lambda\rangle}(v\otimes \boldsymbol{v}_{\lambda+h'}) \tag{2-13}$$

for $v\in S(\hat{\mathfrak{h}}^-)$ and $\lambda\in L$. Then we have $g_hY_{h'}(v,z)v' = Y_{h'}(g_hv,z)g_hv'$ for $v\in V_L$ and $v'\in V_{L+h'}$.

2C. ***Clifford module vertex algebra.*** We also require the standard procedure which attaches a Clifford module super vertex operator algebra to a vector space equipped with a symmetric bilinear form; see [Feingold et al. 1991] for a general treatment, and [Feingold et al. 1996] for the special, one-dimensional case we consider here.

So let $\mathfrak{p}$ be a one-dimensional complex vector space equipped with a nondegenerate symmetric bilinear form $\langle\cdot\,,\cdot\rangle$. Set $\hat{\mathfrak{p}}=\mathfrak{p}[t,t^{-1}]t^{1/2}$ and write $a(r)$ for $a\otimes t^r$. Extend the bilinear form from $\mathfrak{p}$ to $\hat{\mathfrak{p}}$ by requiring that $\langle a(r),b(s)\rangle=\langle a,b\rangle\delta_{r+s,0}$. Set $\hat{\mathfrak{p}}^{\pm}=\mathfrak{p}[t^{\pm1}]t^{\pm1/2}$, write $\langle\hat{\mathfrak{p}}^{\pm}\rangle$ for the subalgebra of $\mathrm{Cliff}(\hat{\mathfrak{p}})$ generated by $\hat{\mathfrak{p}}^{\pm}$, and define a one-dimensional $\langle\hat{\mathfrak{p}}^{+}\rangle$-module $\mathbb{C}\boldsymbol{v}$ by requiring that $\mathbf{1}\boldsymbol{v}=\boldsymbol{v}$ and $a(r)\boldsymbol{v}=0$ for $a\in\mathfrak{p}$ and $r>0$. Here $\mathrm{Cliff}(\hat{\mathfrak{p}})$ denotes the *Clifford algebra* attached to $\hat{\mathfrak{p}}$, which we take to be the quotient of the tensor algebra $T(\hat{\mathfrak{p}})=\mathbb{C}\mathbf{1}\oplus\hat{\mathfrak{p}}\oplus\hat{\mathfrak{p}}^{\otimes 2}\oplus\cdots$ by the ideal generated by expressions of the form $u\otimes u+\frac{1}{2}\langle u,u\rangle\mathbf{1}$ for $u\in\hat{\mathfrak{p}}$.

Observe that the induced $\mathrm{Cliff}(\hat{\mathfrak{p}})$-module, $A(\mathfrak{p})=\mathrm{Cliff}(\hat{\mathfrak{p}})\otimes_{\langle\hat{\mathfrak{p}}^+\rangle}\mathbb{C}\boldsymbol{v}$, is isomorphic to $\bigwedge(\hat{\mathfrak{p}}^-)\boldsymbol{v}$ as a $\langle\hat{\mathfrak{p}}^-\rangle$-module. We obtain a super vertex algebra structure on $A(\mathfrak{p})$ by setting

$$Y\big(a\big(-\tfrac{1}{2}\big)\boldsymbol{v},z\big)=\sum_{n\in\mathbb{Z}}a\big(n+\tfrac{1}{2}\big)z^{-n-1} \tag{2-14}$$

for $a\in\mathfrak{p}$, for the reconstruction theorem of [Frenkel and Ben-Zvi 2004] ensures that this rule extends uniquely to a super vertex algebra structure $Y: A(\mathfrak{p})\otimes A(\mathfrak{p})\to A(\mathfrak{p})((z))$ with $Y(\boldsymbol{v},z)=\mathrm{Id}$.

Let $p\in\mathfrak{p}$ such that $\langle p,p\rangle=-2$. We obtain a super vertex operator algebra structure, with central charge $c=\frac{1}{2}$, by taking

$$\omega=\tfrac{1}{4}p\big(-\tfrac{3}{2}\big)p\big(-\tfrac{1}{2}\big)\boldsymbol{v} \tag{2-15}$$

to be the conformal element.

To construct canonically twisted modules for $A(\mathfrak{p})$ set $\hat{\mathfrak{p}}_{\mathrm{tw}}=\mathfrak{p}[t,t^{-1}]$ and extend the bilinear form from $\mathfrak{p}$ to $\hat{\mathfrak{p}}_{\mathrm{tw}}$ as before, by requiring that $\langle a(r),b(s)\rangle=\langle a,b\rangle\delta_{r+s,0}$. Set $\hat{\mathfrak{p}}_{\mathrm{tw}}^{>}=\mathfrak{p}[t]t$ and $\hat{\mathfrak{p}}_{\mathrm{tw}}^{\leq}=\mathfrak{p}[t^{-1}]$, and define a one-dimensional $\langle\hat{\mathfrak{p}}_{\mathrm{tw}}^{>}\rangle$-module $\mathbb{C}\boldsymbol{v}_{\mathrm{tw}}$ by requiring, much as before, that $\mathbf{1}\boldsymbol{v}_{\mathrm{tw}}=\boldsymbol{v}_{\mathrm{tw}}$ and $a(r)\boldsymbol{v}_{\mathrm{tw}}=0$ for $a\in\mathfrak{p}$ and $r>0$. Then for the induced $\mathrm{Cliff}(\hat{\mathfrak{p}}_{\mathrm{tw}})$-module

$$A(\mathfrak{p})_{\mathrm{tw}} := \mathrm{Cliff}(\hat{\mathfrak{p}}_{\mathrm{tw}})\otimes_{\langle\hat{\mathfrak{p}}_{\mathrm{tw}}^{>}\rangle}\mathbb{C}\boldsymbol{v}_{\mathrm{tw}}, \tag{2-16}$$

there is a unique linear map $Y_{\mathrm{tw}} : A(\mathfrak{p}) \otimes A(\mathfrak{p})_{\mathrm{tw}} \to A(\mathfrak{p})_{\mathrm{tw}}((z^{1/2}))$ such that

$$Y_{\mathrm{tw}}\big(u\big(-\tfrac{1}{2}\big)\boldsymbol{v}, z\big) = \sum_{n\in\mathbb{Z}} u(n) z^{-n-1/2} \tag{2-17}$$

for $u \in \mathfrak{p}$, and $(A(\mathfrak{p})_{\mathrm{tw}}, Y_{\mathrm{tw}})$ is a canonically twisted module for $A(\mathfrak{p})$. One may use a suitably modified formulation of the reconstruction theorem of [Frenkel and Ben-Zvi 2004] to see this; see [Frenkel and Szczesny 2004]. We refer to [Feingold et al. 1996] for a concrete and detailed description of Y_{tw}. Note that $A(\mathfrak{p})$ is isomorphic to $\bigwedge(\hat{\mathfrak{p}}_{\mathrm{tw}}^{\leq})\boldsymbol{v}$ as a $\langle\hat{\mathfrak{p}}_{\mathrm{tw}}^{\leq}\rangle$-module.

With $p \in \mathfrak{p}$ as above, such that $\langle p, p\rangle = -2$, we have $p(0)^2 = \mathbf{1}$ in $\mathrm{Cliff}(\mathfrak{p})$. Set

$$\boldsymbol{v}_{\mathrm{tw}}^{\pm} := (\mathbf{1} \pm p(0))\boldsymbol{v}_{\mathrm{tw}}, \tag{2-18}$$

so that $p(0)\boldsymbol{v}_{\mathrm{tw}}^{\pm} = \pm\boldsymbol{v}_{\mathrm{tw}}^{\pm}$. Then $A(\mathfrak{p})_{\mathrm{tw}} = A(\mathfrak{p})_{\mathrm{tw}}^{+} \oplus A(\mathfrak{p})_{\mathrm{tw}}^{-}$ is a decomposition of $A(\mathfrak{p})_{\mathrm{tw}}$ into irreducible canonically twisted $A(\mathfrak{p})$-modules, where $A(\mathfrak{p})_{\mathrm{tw}}^{\pm}$ denotes the submodule of $A(\mathfrak{p})_{\mathrm{tw}}$ generated by $\boldsymbol{v}_{\mathrm{tw}}^{\pm}$:

$$A(\mathfrak{p})_{\mathrm{tw}}^{\pm} := \mathrm{Cliff}(\hat{\mathfrak{p}}_{\mathrm{tw}}) \otimes_{\langle\hat{\mathfrak{p}}_{\mathrm{tw}}^{>}\rangle} \mathbb{C}\boldsymbol{v}_{\mathrm{tw}}^{\pm}. \tag{2-19}$$

From (2-17) we see that the $L(0)$-degree preserving component of $Y_{\mathrm{tw}}\big(p\big(-\tfrac{1}{2}\big)\boldsymbol{v}, z\big)$ is $p(0)$. Computing the graded-trace of $p(0)$ on $A(\mathfrak{p})_{\mathrm{tw}}^{\pm}$, we find

$$\mathrm{tr}_{A(\mathfrak{p})_{\mathrm{tw}}^{\pm}}\, p(0) q^{L(0)-c/24} = \pm q^{1/24} \prod_{n>0} (1-q^n), \tag{2-20}$$

where the factor $q^{1/24}$ appears because $L(0)\boldsymbol{v}_{\mathrm{tw}}^{\pm} = \frac{1}{16}\boldsymbol{v}_{\mathrm{tw}}^{\pm}$ and $c = \frac{1}{2}$.

2D. ***Main construction.*** We now take $L = \mathbb{Z}\varepsilon_1 + \mathbb{Z}\varepsilon_2 + \mathbb{Z}\varepsilon_3$ to be the rank 3 lattice with bilinear form $\langle\cdot\,,\cdot\rangle$ determined by

$$\langle\varepsilon_i, \varepsilon_j\rangle = 2 - \delta_{i,j}. \tag{2-21}$$

This is an integral, noneven lattice with signature $(1, 2)$. Set $\rho := \frac{1}{5}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$ and observe that

$$\langle\lambda, \rho\rangle = k + l + m \tag{2-22}$$

for $\lambda = k\varepsilon_1 + l\varepsilon_2 + m\varepsilon_3$, so ρ belongs to the dual L^* of L. In fact, L^*/L is cyclic of order 5, and $\rho + L$ is a generator. If we set

$$L^j := \{\lambda \in L \mid \langle\lambda, \rho\rangle = j \pmod 2\}, \tag{2-23}$$

then $L = L^0 \cup L^1$ is the decomposition of L into its even and odd parts, by which we mean that $\langle\lambda, \lambda\rangle$ is even or odd according as λ lies in L^0 or L^1.

Let V_L be the super vertex operator algebra attached to L via the construction of Section 2A, where the bilinear function $b : L \times L \to \mathbb{Z}/2\mathbb{Z}$ is determined by setting

$$b(\varepsilon_i, \varepsilon_j) := \begin{cases} 0 & \text{when } i \le j, \\ 1 & \text{when } i > j. \end{cases} \tag{2-24}$$

There is an obvious action of the symmetric group S_3 on L, by permutations of the basis vectors ε_i. We lift this action to V_L in the following way. Recall from Section 2A that a lift $\hat{g} \in \mathrm{Aut}(V_L)$ of an automorphism $g \in \mathrm{Aut}(L)$ is determined by a choice of function $\alpha : L \to \{\pm 1\}$ satisfying (2-8). Taking $\mu = k\lambda$ in (2-8) we have $\alpha((k+1)\lambda) = \alpha(\lambda)\alpha(k\lambda)\beta(\lambda,\lambda)^k\beta(g\lambda, g\lambda)^k$, since β is bimultiplicative. So given (2-24) we see that $\beta(\lambda,\lambda) = k_1k_2 + k_2k_3 + k_3k_1$ for $\lambda = k_1\varepsilon_1 + k_2\varepsilon_2 + k_3\varepsilon_3$, which is invariant under the action of S_3. So actually $\beta(\lambda,\lambda) = \beta(g\lambda, g\lambda)$, and thus we may assume $\alpha(k\lambda) = \alpha(\lambda)^k$ in (2-8) for $\lambda \in L$ and k a positive integer, when g acts by permuting the ε_i. Observe also that for $\lambda, \mu, \nu \in L$ we have

$$\begin{aligned} &\alpha(\lambda+\mu+\nu)\beta(\lambda,\mu)\beta(\mu,\nu)\beta(\nu,\lambda) \\ &\qquad = \alpha(\lambda)\alpha(\mu)\alpha(\nu)\beta(g\lambda, g\mu)\beta(g\mu, g\nu)\beta(g\nu, g\lambda) \end{aligned} \tag{2-25}$$

according to (2-8), which specializes to

$$\begin{aligned} &\alpha(\lambda)\beta(\varepsilon_1,\varepsilon_2)^{k_1k_2}\beta(\varepsilon_2,\varepsilon_3)^{k_2k_3}\beta(\varepsilon_3,\varepsilon_1)^{k_3k_1} \\ &\quad = \alpha(\varepsilon_1)^{k_1}\alpha(\varepsilon_2)^{k_2}\alpha(\varepsilon_3)^{k_3}\beta(g\varepsilon_1, g\varepsilon_2)^{k_1k_2}\beta(g\varepsilon_2, g\varepsilon_3)^{k_2k_3}\beta(g\varepsilon_3, g\varepsilon_1)^{k_3k_1} \end{aligned} \tag{2-26}$$

for $\lambda = k_1\varepsilon_1 + k_2\varepsilon_2 + k_3\varepsilon_3$.

Consider the case that $g = \sigma$ is the cyclic permutation (123). From (2-26) we see that we may lift σ to $\mathrm{Aut}(V_L)$ by taking $\alpha(\varepsilon_i) = 1$ for $i \in \{1, 2, 3\}$, and more generally $\alpha(k_1\varepsilon_1 + k_2\varepsilon_2 + k_3\varepsilon_3) = (-1)^{k_2k_3+k_3k_1}$, in the construction of Section 2A. We denote the corresponding automorphism of V_L by $\hat{\sigma}$. In the notation of (2-9) we have

$$\hat{\sigma}(v \otimes \boldsymbol{v}_{k_1\varepsilon_1+k_2\varepsilon_2+k_3\varepsilon_3}) := (-1)^{k_2k_3+k_3k_1}(\sigma \cdot v) \otimes \boldsymbol{v}_{k_3\varepsilon_1+k_1\varepsilon_2+k_2\varepsilon_3}. \tag{2-27}$$

Next consider $g = \tau := (12)$. Applying (2-26) we see that we may lift τ to $\mathrm{Aut}(V_L)$ by taking $\alpha(\varepsilon_i) = 1$ as before, and more generally $\alpha(k_1\varepsilon_1 + k_2\varepsilon_2 + k_3\varepsilon_3) = (-1)^{k_1k_2}$. We denote the corresponding automorphism of V_L by $\hat{\tau}$, so that

$$\hat{\tau}(v \otimes \boldsymbol{v}_{k_1\varepsilon_1+k_2\varepsilon_2+k_3\varepsilon_3}) := (-1)^{k_1k_2}(\tau \cdot v) \otimes \boldsymbol{v}_{k_2\varepsilon_1+k_1\varepsilon_2+k_3\varepsilon_3}. \tag{2-28}$$

Using (2-27) and (2-28) one can check that $\hat{\sigma}^3 = \hat{\tau}^2 = (\hat{\tau}\hat{\sigma})^2 = \mathrm{Id}$ in $\mathrm{Aut}(V_L)$, so $\hat{\sigma}$ and $\hat{\tau}$ generate a group

$$\hat{G} := \langle \hat{\sigma}, \hat{\tau} \rangle \simeq S_3 \tag{2-29}$$

in $\mathrm{Aut}(V_L)$.

Note that $V_L = V_{L^0} \oplus V_{L^1}$ is the decomposition of V_L into its even and odd parity subspaces, where L^j is defined by (2-23). Thus the canonical involution of V_L, acting as $+1$ on the even subspace V_{L^0} and -1 on the odd subspace V_{L^1}, is realized by $g_{\rho/2}$ in the notation of (2-11). So the canonically twisted modules for V_L are exactly the $V_{L+a\rho/2}$, for $a \in \{1, 3, 5, 7, 9\}$ (see Section 2B).

The prescription (2-13) furnishes an extension of the action of the canonical involution $g_{\rho/2}$ from V_L to $V_{L+a\rho/2}$. Since ρ is S_3-invariant we may also extend the actions of $\hat{\sigma}$ and $\hat{\tau}$ to $V_{L+a\rho/2}$ by setting

$$\begin{aligned} \hat{\sigma}(v \otimes \boldsymbol{v}_{\lambda+a\rho/2}) &:= (-1)^{k_2k_3+k_3k_1}(\sigma \cdot v) \otimes \boldsymbol{v}_{\sigma\lambda+a\rho/2}, \\ \hat{\tau}(v \otimes \boldsymbol{v}_{\lambda+a\rho_2}) &:= (-1)^{k_1k_2+(a-1)/2}(\tau \cdot v) \otimes \boldsymbol{v}_{\tau\lambda+a\rho/2}, \end{aligned} \tag{2-30}$$

for $v \in S(\hat{\mathfrak{h}}^-)$ and $\lambda = k_1\varepsilon_1 + k_2\varepsilon_2 + k_3\varepsilon_3$. We use these rules to equip the $V_{L+a\rho/2}$ with $\hat{G}$- and $\langle g_{\rho/2}\rangle$-module structures.

Now observe that if $K \subset L$ is closed under addition and contains 0 then the $\hat{\mathfrak{h}}$-submodule $V_K < V_L$ generated by the $\boldsymbol{v}_\lambda$ for $\lambda \in K$,

$$V_K \simeq S(\hat{\mathfrak{h}}^-) \otimes \mathbb{C}[K], \tag{2-31}$$

is a sub-super vertex algebra of V_L, and the ω of (2-7) is a conformal element for V_K. Furthermore, if $K' \subset L + \frac{1}{2}a\rho$ satisfies $K + K' \subset K'$ then the restriction of the twisted vertex operators $u \otimes v \mapsto Y_{a\rho/2}(u, z^{1/2})v$ to $V_K \otimes V_{K'} < V_L \otimes V_{L+a\rho/2}$ equips $V_{K'}$ with a canonically twisted module structure over V_K (for a odd). To describe the choices of K and K' that are relevant to us, define P to be the monoid of nonnegative integer combinations of the ε_i,

$$P := \left\{ \sum_i k_i\varepsilon_i \in L \;\middle|\; k_i \geq 0, \ \forall i \right\}. \tag{2-32}$$

Then V_P is naturally a conformal super vertex algebra, and for a odd, $V_{P+a\rho/2}$ is a canonically twisted module for it. By our choice of basis $\{\varepsilon_i\}$, all nontrivial vectors of the monoid P have strictly positive norm. So the eigenspaces for the action of $L(0)$ on V_P are finite-dimensional, and the eigenvalues of $L(0)$ are contained in $\frac{1}{2}\mathbb{Z}$ and bounded from below. Thus V_P is actually a super vertex operator algebra.

Observe that the actions (2-30) restrict to $V_{P+a\rho/2}$ for any a, since P and ρ are invariant under coordinate permutations. The canonical involution $g_{\rho/2}$, acting on $V_{L+a\rho/2}$ according to (2-13), also preserves the subspace $V_{P+a\rho/2}$.

We now let V^X denote the tensor product super vertex operator algebra

$$V^X := A(\mathfrak{p}) \otimes V_P. \tag{2-33}$$

For a an odd integer we define $V^{\pm}_{\mathrm{tw},a}$ to be the canonically twisted V^X-module

$$V^{\pm}_{\mathrm{tw},a} := A(\mathfrak{p})^{\pm}_{\mathrm{tw}} \otimes V_{P+a\rho/2} \oplus A(\mathfrak{p})^{\mp}_{\mathrm{tw}} \otimes V_{P+(10-a)\rho/2}. \tag{2-34}$$

We extend the action of $\hat{G} \simeq S_3$ to V^X and $V^{\pm}_{\mathrm{tw},a}$ by letting $\hat{G}$ act trivially on the Clifford module factors, setting

$$\hat{\sigma}(u \otimes v) := u \otimes \hat{\sigma}(v), \qquad \hat{\tau}(u \otimes v) := u \otimes \hat{\tau}(v) \tag{2-35}$$

for $u \in A(\mathfrak{p})$ and $v \in V_P$, and for $u \in A(\mathfrak{p})^{\pm}_{\mathrm{tw}}$ and $v \in V_{P+a\rho/2}$.

Given $g \in \hat{G}$ and a an odd integer, we now define $T^{\pm}_{g,a}$ to be the trace of the operator $g g_{\rho/2} p(0) q^{L(0)-c/24}$ on the canonically twisted V^X-module $V^{\pm}_{\mathrm{tw},a}$,

$$T^{\pm}_{g,a} := \mathrm{tr}_{V^{\pm}_{\mathrm{tw},a}}\, g g_{\rho/2} p(0) q^{L(0)-c/24}. \tag{2-36}$$

Note that $c = \frac{7}{2}$ here. Also, the identities

$$T^{\pm}_{g,a} = -T^{\mp}_{g,a} = T^{\mp}_{g,10-a} \tag{2-37}$$

follow directly from (2-34) and (2-36). In particular, we have $T^{\pm}_{g,5} = 0$ for all g.

Recall that $(q;q)_\infty = \prod_{n>0}(1-q^n)$ (see (1-1)). Our concrete construction allows us to compute explicit formulas for the trace functions $T^{\pm}_{g,a}$.

Proposition 2.2. *The trace functions $T^{\pm}_{g,a}$ admit the following expressions*:

$$T^{\pm}_{e,a} = \pm \frac{q^{-1/12}}{(q;q)^2_\infty} \Biggl(\sum_{k,l,m \geq 0} + \sum_{k,l,m<0} \Biggr) (-1)^{k+l+m} \times q^{(k^2+l^2+m^2)/2 + 2(kl+lm+mk) + a(k+l+m)/2 + 3a^2/40}, \tag{2-38}$$

$$T^{\pm}_{\hat{\tau},a} = \pm (-1)^{(a-1)/2} \frac{q^{-1/12}}{(q^2;q^2)_\infty} \Biggl(\sum_{k,m \geq 0} - \sum_{k,m<0} \Biggr) (-1)^{k+m} \times q^{3k^2+m^2/2+4km+a(2k+m)/2+3a^2/40}, \tag{2-39}$$

$$T^{\pm}_{\hat{\sigma},a} = \pm q^{-1/12} \frac{(q;q)_\infty}{(q^3;q^3)_\infty} \sum_{k \in \mathbb{Z}} (-1)^k q^{15k^2/2+3ak/2+3a^2/40}. \tag{2-40}$$

Proof. First consider the case that $g = e$ is the identity. From the definition (2-36) of $T^{\pm}_{e,a}$ we derive

$$T^{\pm}_{e,a} = \pm \frac{1}{(q;q)^2_\infty} \sum_{\mu \in P + a\rho/2} (-1)^{\langle \mu - a\rho/2, \rho \rangle} q^{\langle \mu, \mu \rangle/2 - 1/12} \mp \frac{1}{(q;q)^2_\infty} \sum_{\mu \in P + (10-a)\rho/2} (-1)^{\langle \mu - (10-a)\rho/2, \rho \rangle} q^{\langle \mu, \mu \rangle/2 - 1/12}. \tag{2-41}$$

We replace μ with $k\varepsilon_1+l\varepsilon_2+m\varepsilon_3+\frac{1}{2}a\rho$ in the first summation, where $k,l,m\geq 0$, and obtain

$$\pm\frac{q^{-1/12}}{(q;q)_\infty^2}\sum_{k,l,m\geq 0}(-1)^{k+l+m}q^{(k^2+l^2+m^2)/2+2(kl+lm+mk)+a(k+l+m)/2+3a^2/40} \tag{2-42}$$

using (2-21) and (2-22) and the fact that $\langle\rho,\rho\rangle=\frac{3}{5}$. For the second summation we replace $P+\frac{1}{2}(10-a)\rho$ with $-P-\frac{1}{2}(10-a)\rho=-P^++\frac{1}{2}a\rho$, where

$$P^+:=\left\{\sum_i k_i\varepsilon_i\in L\;\middle|\;k_i>0,\ \forall i\right\}, \tag{2-43}$$

and replace μ with $-\mu$ in the summands. We obtain

$$\pm\frac{q^{-1/12}}{(q;q)_\infty^2}\sum_{k,l,m<0}(-1)^{k+l+m}q^{(k^2+l^2+m^2)/2+2(kl+lm+mk)+a(k+l+m)/2+3a^2/40}, \tag{2-44}$$

which together with (2-42) gives (2-38) as required.

Next take $g=\hat{\tau}$. Using (2-36) and the formula (2-30) we compute

$$\begin{aligned}T^{\pm}_{\hat{\tau},a}&=\pm(-1)^{(a-1)/2}\frac{1}{(q^2;q^2)_\infty}\\&\quad\times\sum_{\substack{\mu\in P+a\rho/2\\ \tau\mu=\mu}}(-1)^{\langle\mu-a\rho/2,\rho+\varepsilon_1'\rangle}q^{\langle\mu,\mu\rangle/2-1/12}\mp(-1)^{(a-1)/2}\frac{1}{(q^2;q^2)_\infty}\\&\qquad\qquad\times\sum_{\substack{\mu\in P+(10-a)\rho/2\\ \tau\mu=\mu}}(-1)^{\langle\mu-(10-a)\rho/2,\rho+\varepsilon_1'\rangle}q^{\langle\mu,\mu\rangle/2-1/12}.\end{aligned} \tag{2-45}$$

We now replace μ with $k\varepsilon_1+k\varepsilon_2+m\varepsilon_3+\frac{1}{2}a\rho$ in the first summation, where $k,m\geq 0$, and obtain

$$\pm(-1)^{(a-1)/2}\frac{q^{-1/12}}{(q^2;q^2)_\infty}\sum_{k,m\geq 0}(-1)^{k+m}q^{3k^2+m^2/2+4km+a(2k+m)/2+3a^2/40}. \tag{2-46}$$

Note that the factor $(-1)^k$ in $(-1)^{k+m}$, corresponding to $(-1)^{\langle\mu-a\rho/2,\varepsilon_1'\rangle}$ in (2-45), arises from the factor $(-1)^{k_1k_2}=(-1)^{k^2}=(-1)^k$ in (2-30). For the second summation we proceed as before, replacing $P+\frac{1}{2}(10-a)\rho$ with $-P^++\frac{1}{2}a\rho$ and μ with $-\mu$, and obtain

$$\mp(-1)^{(a-1)/2}\frac{q^{-1/12}}{(q^2;q^2)_\infty}\sum_{k,m<0}(-1)^{k+m}q^{3k^2+m^2/2+4km+a(2k+m)/2+3a^2/40}. \tag{2-47}$$

This taken together with (2-46) gives the required expression (2-39).

Finally we consider $g=\hat{\sigma}$ (see (2-30)). Then the appropriate analogue of (2-41) and (2-45) is

$$T^{\pm}_{\hat{\sigma},a}=\pm\frac{(q;q)_\infty}{(q^3;q^3)_\infty}\sum_{\substack{\mu\in P+a\rho/2\\ \sigma\mu=\mu}}(-1)^{\langle\mu-a\rho/2,\rho\rangle}q^{\langle\mu,\mu\rangle/2-1/12}$$
$$\mp\frac{(q;q)_\infty}{(q^3;q^3)_\infty}\sum_{\substack{\mu\in P+(10-a)\rho/2\\ \sigma\mu=\mu}}(-1)^{\langle\mu-(10-a)\rho/2,\rho\rangle}q^{\langle\mu,\mu\rangle/2-1/12}. \tag{2-48}$$

We obtain (2-40) from (2-48) in much the same way as for $g=e$, so we omit the remaining details. □

3. Mock theta functions

In this section we consider the modular properties of the trace functions defined in Section 2D, computed explicitly in Proposition 2.2. We recall some basic facts about Maass forms in Section 3A, including their relationship to mock modular forms. We require some facts about theta series of cones in indefinite lattices due to Zwegers [2002], which we recall in Section 3B. The proof of our main result, Theorem 1.1, appears in Section 3C. In particular, we identify the umbral McKay–Thompson series attached to $X=E_8^3$ as trace functions arising from the action of G^X on canonically twisted modules for V^X in Section 3C.

3A. *Harmonic Maass forms.* Define the weight $\frac{1}{2}$ *Casimir operator* $\Omega_{\frac{1}{2}}$, a differential operator on smooth functions $H:\mathbb{H}\to\mathbb{C}$, by setting

$$(\Omega_{\frac{1}{2}}H)(\tau):=-4\Im(\tau)^2\frac{\partial^2 H}{\partial\tau\partial\bar{\tau}}(\tau)+i\Im(\tau)\frac{\partial H}{\partial\bar{\tau}}(\tau)+\frac{3}{16}H(\tau). \tag{3-1}$$

Note that $\Omega_{\frac{1}{2}}=\Delta_{\frac{1}{2}}+\frac{3}{16}$, where Δ_k is the hyperbolic Laplace operator in weight k.

Following [Bruinier and Funke 2004] (see also [Ono 2009; Zagier 2009]), a *harmonic weak Maass form* of weight $\frac{1}{2}$ for $\Gamma<\mathrm{SL}_2(\mathbb{Z})$ is defined to be a smooth function $H:\mathbb{H}\to\mathbb{C}$ that transforms as a (not necessarily holomorphic) modular form of weight $\frac{1}{2}$ for Γ, is an eigenfunction for $\Omega_{\frac{1}{2}}$ with eigenvalue $\frac{3}{16}$, and has at most exponential growth as τ approaches cusps of Γ.

Define $\beta(x)$ for $x\in\mathbb{R}_{\geq 0}$ by setting

$$\beta(x):=\int_x^\infty u^{-1/2}e^{-\pi u}\,\mathrm{d}u. \tag{3-2}$$

Note that β is related to the incomplete gamma function by $\sqrt{\pi}\beta(x)=\Gamma\left(\frac{1}{2},\pi x\right)$. If H is a harmonic weak Maass form of weight $\frac{1}{2}$ then we can canonically decompose

H into its *holomorphic* and *nonholomorphic* parts, $H = H^+ + H^-$, where

$$H^+(\tau) = \sum_{n \gg -\infty} c_H^+(n) q^n, \tag{3-3}$$

$$H^-(\tau) = 2i c_H^-(0)\sqrt{2\Im(\tau)} - i \sum_{n>0} c_H^-(n) \frac{1}{\sqrt{2n}} \beta(4n\Im(\tau)) q^{-n}, \tag{3-4}$$

for some uniquely determined values $c_H^{\pm}(n) \in \mathbb{C}$. (See §3 of [Bruinier and Funke 2004]. See also §5 of [Zagier 2009] and §7.1 of [Dabholkar et al. 2012].) Note that n should be allowed to range over rational values in (3-3) and (3-4).

We may define the *mock modular forms* of weight $\frac{1}{2}$ to be those holomorphic functions $H^+ : \mathbb{H} \to \mathbb{C}$ which arise as the holomorphic parts of harmonic weak Maass forms of weight $\frac{1}{2}$. For $H^{\pm}$ as above, the *shadow* of H^+ is defined, up to a choice of scaling factor C, by

$$g(\tau) := C\sqrt{2\Im(\tau)}\,\overline{\frac{\partial H^-}{\partial \bar{\tau}}} = C \sum_{n \geq 0} c_H^-(n) q^n. \tag{3-5}$$

Then so long as $c_H^-(0) = 0$ (i.e., g is a cusp form), the function H^- is the *Eichler integral* of g,

$$H^-(\tau) = \frac{e\left(-\frac{1}{8}\right)}{C} \int_{-\bar{\tau}}^{\infty} \frac{\overline{g(-\bar{z})}}{\sqrt{z+\tau}}\, \mathrm{d}z. \tag{3-6}$$

In this setting, the weak harmonic Maass form $H = H^+ + H^-$ is called the *completion* of H^+.

Various choices for C can be found in the literature. In [Cheng et al. 2014b] we find $C = \sqrt{2m}$ in the case that $H = (H_r)$ is a $2m$-vector-valued Maass form for some $\Gamma_0(N)$, such that

$$(H \cdot \theta)(\tau, z) := \sum_r H_r(\tau)\theta_{m,r}(\tau, z) \tag{3-7}$$

transforms like a (not necessarily holomorphic in τ) Jacobi form of weight 1 and index m for $\Gamma_0(N)$, where

$$\theta_{m,r}(\tau, z) := \sum_{k \in \mathbb{Z}} q^{(2km+r)^2/4m} e^{2\pi i z(2km+r)}. \tag{3-8}$$

The cases of relevance to us here all have $m = 30$, so we take $C = \sqrt{60}$ henceforth in (3-5) and (3-6). All the shadows arising in this work are linear combinations of the unary theta functions

$$S_{m,r}(\tau) := \frac{1}{2\pi i} \frac{\partial}{\partial z} \theta_{m,r}(\tau, z)\Big|_{z=0} = \sum_{k \in \mathbb{Z}} (2km+r) q^{(2km+r)^2/4m}, \tag{3-9}$$

where $m = 30$ and $r \neq 0 \pmod{30}$. In particular, we do not encounter any examples for which the shadow g (see (3-5)) is not a cusp form.

3B. ***Indefinite theta series.*** Even though our main construction uses a lattice of signature $(1, 2)$, it will develop in Section 3C that the trace functions (2-38) and (2-39) can be analyzed in terms of theta series of indefinite lattices with signature $(1, 1)$, whereas (2-40) is essentially a theta series with rank 1 and consequently can be handled by classical methods. With this in mind we now recall some results from [Zwegers 2002] on theta series for lattices of signature $(r-1, 1)$, restricting to the case that $r = 2$.

Given a symmetric 2×2 matrix A, we define a quadratic form $Q : \mathbb{R}^2 \to \mathbb{R}$, by setting

$$Q(\boldsymbol{x}) := \tfrac{1}{2}(\boldsymbol{x}, A\boldsymbol{x}), \tag{3-10}$$

where $(\,\cdot\,,\cdot\,)$ denotes the usual Euclidean inner product on $\mathbb{R}^2$. The associated bilinear form is

$$B(\boldsymbol{x}, \boldsymbol{y}) := (\boldsymbol{x}, A\boldsymbol{y}) = Q(\boldsymbol{x}+\boldsymbol{y}) - Q(\boldsymbol{x}) - Q(\boldsymbol{y}). \tag{3-11}$$

Henceforth assume that A has signature $(1, 1)$. Then the set of vectors $\boldsymbol{c} \in \mathbb{R}^2$ with $Q(\boldsymbol{c}) < 0$ is nonempty and has two components. Let C_Q be one of these components. Two vectors $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}$ belong to the same component if $B(\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}) < 0$. Thus, picking a vector $\boldsymbol{c}_0$ in C_Q, we may identify

$$C_Q = \{\boldsymbol{c} \in \mathbb{R}^2 \mid Q(\boldsymbol{c}) < 0,\ B(\boldsymbol{c}, \boldsymbol{c}_0) < 0\}. \tag{3-12}$$

Zwegers also defines a set of representatives of *cusps*,

$$S_Q := \{\boldsymbol{c} \in \mathbb{Z}^2 \mid \boldsymbol{c} \text{ primitive},\ Q(\boldsymbol{c}) = 0,\ B(\boldsymbol{c}, \boldsymbol{c}_0) < 0\}. \tag{3-13}$$

Define the *indefinite theta function* with characteristics $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^2$, with respect to $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)} \in C_Q$, by setting

$$\vartheta_{\boldsymbol{a},\boldsymbol{b}}^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}(\tau) := \sum_{\nu \in \boldsymbol{a}+\mathbb{Z}^2} \left(E\left(\frac{B(\boldsymbol{c}^{(1)}, \nu)}{\sqrt{-Q(\boldsymbol{c}^{(1)})}} \sqrt{\Im(\tau)} \right) - E\left(\frac{B(\boldsymbol{c}^{(2)}, \nu)}{\sqrt{-Q(\boldsymbol{c}^{(2)})}} \sqrt{\Im(\tau)} \right) \right) q^{Q(\nu)} e^{2\pi i B(\nu, \boldsymbol{b})}, \tag{3-14}$$

where $E(z) := \operatorname{sgn}(z)(1 - \beta(z^2))$. Corollary 2.9 of [Zwegers 2002] (see also Theorem 3.1 of [Zagier 2009]) shows that $\vartheta_{\boldsymbol{a},\boldsymbol{b}}^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}(\tau)$ is a nonholomorphic modular form of weight 1.

Presently we will see that these indefinite theta functions can be used to define harmonic Maass forms whose nonholomorphic parts can be written in terms of the functions

$$R_{a,b}(\tau) := \sum_{\nu \in a+\mathbb{Z}} \operatorname{sgn}(\nu)\beta(2\nu^2 \Im(\tau)) q^{-\nu^2/2} e^{-2\pi i \nu b}. \tag{3-15}$$

Note that the $R_{a,b}$ are Eichler integrals (see (3-6)) of unary theta functions of weight $\frac{3}{2}$. Indeed, we have

$$R_{a,b}(\tau) = e\big(-\tfrac{1}{8}\big) \int_{-\bar{\tau}}^{i\infty} \frac{g_{a,-b}(z)}{\sqrt{z+\tau}}\,\mathrm{d}z, \tag{3-16}$$

for

$$g_{a,b}(\tau) := \sum_{\nu \in a+\mathbb{Z}} \nu q^{\nu^2/2} e^{2\pi i \nu b}. \tag{3-17}$$

Note that $\overline{g_{a,b}(-\bar{z})} = g_{a,-b}(z)$. Observe also that

$$g_{r/2m,0}(m\tau) = \frac{1}{2m} S_{m,r}(\tau) \tag{3-18}$$

(see (3-9)), which is useful for comparing the results of [Zwegers 2002] to those of [Cheng et al. 2014b].

Define $\langle \boldsymbol{c}\rangle_{\mathbb{Z}}^{\perp} := \{\xi \in \mathbb{Z}^r \mid B(\boldsymbol{c}, \xi) = 0\}$. For future use we quote the $r = 2$ case of Proposition 4.3 from [Zwegers 2002].

Proposition 3.1 (Zwegers). *Let $\boldsymbol{c} \in C_Q \cap \mathbb{Z}^2$ be primitive. Let $P_0 \subset \mathbb{R}^2$ be any finite set such that*

$$\Big\{\mu \in \boldsymbol{a} + \mathbb{Z}^2 \;\Big|\; 0 \le \frac{B(\boldsymbol{c}, \mu)}{2Q(\boldsymbol{c})} < 1\Big\} = \bigsqcup_{\mu_0 \in P_0} \big(\mu_0 + \langle \boldsymbol{c}\rangle_{\mathbb{Z}}^{\perp}\big). \tag{3-19}$$

Then we have

$$\begin{aligned} &\sum_{\nu \in \boldsymbol{a}+\mathbb{Z}^2} \operatorname{sgn}(B(\boldsymbol{c}, \nu))\beta\Big(-\frac{B(\boldsymbol{c}, \nu)^2}{Q(\boldsymbol{c})}\Im(\tau)\Big) e^{2\pi i Q(\nu)\tau + 2\pi i B(\nu, \boldsymbol{b})} \\ &\quad = -\sum_{\mu_0 \in P_0} R_{B(\boldsymbol{c},\mu_0)/2Q(\boldsymbol{c}),\, B(\boldsymbol{c},\boldsymbol{b})}(-2Q(\boldsymbol{c})\tau) \cdot \sum_{\xi \in \mu_0^{\perp} + \langle \boldsymbol{c}\rangle_{\mathbb{Z}}^{\perp}} e^{2\pi i Q(\xi)\tau + 2\pi i B(\xi, \boldsymbol{b}^{\perp})}, \end{aligned} \tag{3-20}$$

where $\mu_0^{\perp} = \mu_0 - \dfrac{B(\boldsymbol{c}, \mu_0)}{2Q(\boldsymbol{c})}\boldsymbol{c}$ *and* $\boldsymbol{b}^{\perp} = \boldsymbol{b} - \dfrac{B(\boldsymbol{c}, \boldsymbol{b})}{2Q(\boldsymbol{c})}\boldsymbol{c}$.

Note that the term

$$\sum_{\xi \in \mu_0^{\perp} + \langle \boldsymbol{c}\rangle_{\mathbb{Z}}^{\perp}} e^{2\pi i Q(\xi)\tau + 2\pi i B(\xi, \boldsymbol{b}^{\perp})} \tag{3-21}$$

is a classical (positive-definite) theta function of weight $\frac{1}{2}$.

The indefinite theta function construction (3-14) is applied in [Zwegers 2002] to mock theta functions of Ramanujan (other than χ_0 and χ_1, which are treated in [Zwegers 2009]). Amongst those appearing are the four functions F_0, F_1, ϕ_0 and

ϕ_1, where ϕ_0 and ϕ_1 are defined in (1-9), and

$$F_0(q) := \sum_{n\geq 0} \frac{q^{2n^2}}{(q;q^2)_n},\qquad F_1(q) := \sum_{n\geq 0} \frac{q^{2n(n+1)}}{(q;q^2)_{n+1}}. \tag{3-22}$$

These are amongst the fifth-order mock theta functions introduced by Ramanujan in his last letter to Hardy.

To study these functions, 6-vector-valued mock modular forms

$$F_{5,1}(\tau) = (F_{5,1,r}(\tau)), \quad F_{5,2}(\tau) = (F_{5,2,r}(\tau)) \tag{3-23}$$

are introduced on pages 74 and 79, respectively, of [Zwegers 2002]. Inspecting their definitions, and substituting 2τ for τ, we find that

$$F_{5,1,3}(2\tau) = q^{-1/120}(F_0(q)-1), \qquad F_{5,2,3}(2\tau) = q^{-1/120}\phi_0(-q), \tag{3-24}$$

$$F_{5,1,4}(2\tau) = q^{71/120}F_1(q), \qquad F_{5,2,4}(2\tau) = -q^{-49/120}\phi_1(-q). \tag{3-25}$$

The content of Proposition 4.10 of [Zwegers 2002] is that

$$H_{5,1}(\tau) = F_{5,1}(\tau) - G_{5,1}(\tau), \tag{3-26}$$

where the vector-valued functions $H_{5,1}$ and $G_{5,1}$ are such that the components of $2\eta(\tau)H_{5,1}(\tau)$ are nonholomorphic indefinite theta functions of the form $\vartheta_{\boldsymbol{a},\boldsymbol{b}}^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}(\tau)$ (see (3-14)), and the third and fourth components of $G_{5,1}$ satisfy

$$G_{5,1,3}(2\tau) = -\tfrac{1}{2}\big(R_{\frac{19}{60},0} + R_{\frac{29}{60},0} - R_{\frac{49}{60},0} - R_{\frac{59}{60},0}\big)(60\tau), \tag{3-27}$$

$$G_{5,1,4}(2\tau) = -\tfrac{1}{2}\big(R_{\frac{13}{60},0} + R_{\frac{23}{60},0} - R_{\frac{43}{60},0} - R_{\frac{53}{60},0}\big)(60\tau). \tag{3-28}$$

(See (3-15) for $R_{a,b}$.) Moreover, $H_{5,1}(\tau)$ is an eigenfunction for $\Omega_{\frac{1}{2}}$ with eigenvalue $\frac{3}{16}$ (see (3-1)). In other words, the components of $H_{5,1} = (H_{5,1,r})$ are harmonic weak Maass forms of weight $\frac{1}{2}$ (see Section 3A).

Proposition 4.13 of [Zwegers 2002] establishes a similar result for $F_{5,2}$, namely

$$H_{5,2}(\tau) = F_{5,2}(\tau) - G_{5,2}(\tau), \tag{3-29}$$

where $H_{5,2}$ is again a harmonic weak Maass form of weight $\frac{1}{2}$, and $G_{5,2} = -G_{5,1}$.

The left-hand sides of (3-26) and (3-29) are harmonic weak Maass forms of weight $\frac{1}{2}$, so they admit canonical decompositions into holomorphic (see (3-3)) and nonholomorphic (see (3-4)) parts. The summands $F_{5,1}$ and $F_{5,2}$ on the right-hand sides are holomorphic by construction, and the $R_{a,b}$ are of the same form as (3-4) by construction (see (3-15)), so the right-hand sides of (3-26) and (3-29) are precisely the decompositions of $H_{5,1}$ and $H_{5,2}$ into their holomorphic and nonholomorphic parts.

Equivalently, the four functions $F_{5,j,r}$ are mock modular forms of weight $\frac{1}{2}$ with completions given by the $H_{5,j,r}$, and the $G_{5,j,r}$ are the Eichler integrals of their shadows. Thus we can describe their shadows explicitly. Applying (3-16), (3-17) and (3-18), and the identities $g_{1-a,0} = g_{-a,0} = -g_{a,0}$, we see that $F_{5,1,3}(2\tau)$ and $-F_{5,2,3}(2\tau)$ have the same shadow

$$\tfrac{1}{2}(S_{30,1} + S_{30,11} + S_{30,19} + S_{30,29})(\tau), \tag{3-30}$$

while $F_{5,1,4}(2\tau)$ and $-F_{5,2,4}(2\tau)$ both have shadow

$$\tfrac{1}{2}(S_{30,7} + S_{30,13} + S_{30,17} + S_{30,23})(\tau). \tag{3-31}$$

3C. ***McKay–Thompson series.*** We now prove our main result, Theorem 1.1, that the trace functions arising from the action of G^X on the $V^{\pm}_{\mathrm{tw},a}$ recover the Fourier expansions of the mock modular forms H^X_g attached to $g \in G^X \simeq S_3$ by umbral moonshine at $X = E_8^3$.

To formulate this precisely, let $T^X_g = (T^X_{g,r})$ be the vector of Laurent series in $q^{1/120}$, with components indexed by $\mathbb{Z}/60\mathbb{Z}$, such that

$$T^X_{g,r} := \begin{cases} T^{\mp}_{g,1} & \text{for } r = \pm 1, \pm 11, \pm 19, \pm 29 \pmod{60}, \\ T^{\mp}_{g,7} & \text{for } r = \pm 7, \pm 13, \pm 17, \pm 23 \pmod{60}, \\ 0 & \text{else.} \end{cases} \tag{3-32}$$

(Recall from Section 2D that $T^{\mp}_{g,a} = -T^{\mp}_{g,10-a}$ and in particular $T^{\mp}_{g,5} = 0$ for all g.) Define the *polar part at infinity* of T^X_g to be the vector of polynomials in $q^{-1/120}$ obtained by removing all nonnegative rational powers of q in each component $T^X_{g,r}$. Let $g \mapsto \bar{\chi}^X_g$ be the natural permutation character of G^X, so that $\bar{\chi}_g$ is 3, 1 or 0, according as g has order 1, 2 or 3, and define a vector $S^X_g = (S^X_{g,r})$ of theta series, with components indexed by $\mathbb{Z}/60\mathbb{Z}$, by setting

$$S^X_{g,r} := \begin{cases} \pm\bar{\chi}_g(S_{30,1} + S_{30,11} + S_{30,19} + S_{30,29}) \\ \qquad\qquad\qquad \text{for } r = \pm 1, \pm 11, \pm 19, \pm 29 \pmod{60}, \\ \pm\bar{\chi}_g(S_{30,7} + S_{30,13} + S_{30,17} + S_{30,23}) \\ \qquad\qquad\qquad \text{for } r = \pm 7, \pm 13, \pm 17, \pm 23 \pmod{60}, \\ 0 \qquad\qquad\quad\;\; \text{else.} \end{cases} \tag{3-33}$$

(See (3-9) for $S_{m,r}$.)

Set $S^X := S^X_e$, and let $\sigma^X : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_{60}(\mathbb{C})$ denote the multiplier system of S^X, so that

$$\sigma^X(\gamma) S^X(\gamma\tau)(c\tau + d)^{-3/2} = S^X(\tau) \tag{3-34}$$

for $\tau \in \mathbb{H}$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, when (c,d) is the lower row of γ. Our next goal (to be realized in Proposition 3.2) is to show that $2T_g^X$ is a mock modular form with shadow S_g^X for $g \in G^X$. This condition tells us what the multiplier system of T_g^X must be, at least when $o(g)$ is 1 or 2 (as S_g^X is identically zero when $o(g)=3$). For the convenience of the reader we describe this multiplier system in more detail now.

It is cumbersome to work with matrices in $\mathrm{GL}_{60}(\mathbb{C})$, but we can avoid this since any nonzero component of T_g^X is ± 1 times $T_{g,1}^X$ or $T_{g,7}^X$. In other words, we can work with the 2-vector-valued functions $\check{T}_g^X := (T_{g,1}^X, T_{g,7}^X)$ and $\check{S}_g^X := (S_{g,1}^X, S_{g,7}^X)$. If $h=(h_r)$ is a modular form of weight $\frac{1}{2}$ with multiplier system conjugate to that of S^X, and satisfying

$$h_r := \begin{cases} h_1 & \text{for } r = \pm 1, \pm 11, \pm 19, \pm 29 \pmod{60}, \\ h_7 & \text{for } r = \pm 7, \pm 13, \pm 17, \pm 23 \pmod{60}, \\ 0 & \text{else}, \end{cases} \tag{3-35}$$

then, setting $\check{h} = (h_1, h_7)$, we have

$$\check{h}\Big(\frac{a\tau+b}{c\tau+d}\Big)\check{\nu}\Big(\frac{a\tau+b}{c\tau+d}\Big)(c\tau+d)^{-1/2} = \check{h}(\tau), \tag{3-36}$$

where $\check{\nu} : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{GL}_2(\mathbb{C})$ is determined by the rules

$$\begin{aligned} \check{\nu}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} e\big(-\frac{1}{120}\big) & 0 \\ 0 & e\big(-\frac{49}{120}\big) \end{pmatrix}, \\ \check{\nu}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} &= \frac{2e\big(\frac{3}{8}\big)}{\sqrt{15}} \begin{pmatrix} \sin\big(\frac{1}{30}\pi\big)+\sin\big(\frac{11}{30}\pi\big) & \sin\big(\frac{7}{30}\pi\big)+\sin\big(\frac{13}{30}\pi\big) \\ \sin\big(\frac{7}{30}\pi\big)+\sin\big(\frac{13}{30}\pi\big) & -\sin\big(\frac{1}{30}\pi\big)-\sin\big(\frac{11}{30}\pi\big) \end{pmatrix}. \end{aligned} \tag{3-37}$$

We now return to our main objective: the determination of the modularity of T_g^X for $g \in G^X$. To describe the multiplier system for T_g^X when $o(g)=3$, we require the function $\rho_{3|3} : \Gamma_0(3) \to \mathbb{C}^\times$, defined by setting

$$\rho_{3|3}\begin{pmatrix} a & b \\ c & d \end{pmatrix} := e\Big(\frac{cd}{9}\Big). \tag{3-38}$$

Evidently $\rho_{3|3}$ has order 3, and restricts to the identity on $\Gamma_0(9)$.

Proposition 3.2. *Let $g \in G^X$. Then $2T_g^X$ is the Fourier series of a mock modular form for $\Gamma_0(o(g))$, whose shadow is S_g^X. The polar part at infinity of $2T_g^X$ is given by*

$$T_{g,r}^X = \begin{cases} \mp 2q^{-1/120} + O(1) & \textit{for } r = \pm 1, \pm 11, \pm 19, \pm 29 \pmod{60}, \\ O(1) & \textit{otherwise}, \end{cases} \tag{3-39}$$

and $2T_g^X$ has vanishing polar part at all noninfinite cusps of $\Gamma_0(o(g))$. If $o(g)=3$ then the multiplier system of $2T_g^X$ is given by $\gamma \mapsto \rho_{3|3}(\gamma)\overline{\sigma^X(\gamma)}$.

Proof. According to our definition (3-32), the components of T_g^X are $T_{g,1}^{\pm}$ or $T_{g,7}^{\pm}$. In practice it is more convenient to work with $T_{g,3}^{\pm}$ than $T_{g,7}^{\pm}$, and we may do so because $T_{g,7}^{\pm}=-T_{g,3}^{\pm}$ according to (2-37).

We now verify that the series T_g^X are Fourier expansions of vector-valued mock modular forms, and determine their shadows. For the case $g=e$ we compute $\frac{3}{40}-\frac{1}{12}=-\frac{1}{120}$ and $\frac{27}{40}-\frac{1}{12}=\frac{71}{120}$, and see, upon comparison of (2-38) with (1-4), that $T_{e,1}^{\pm}(q)=\pm q^{-1/120}(2-\chi_0(q))$ and $T_{e,3}^{\pm}=\pm q^{71/120}\chi_1(q)$. In particular,

$$
\begin{aligned}
2T_{e,1}^{-}&=2q^{-1/120}(\chi_0(q)-2),\\
2T_{e,7}^{-}&=2q^{71/120}\chi_1(q).
\end{aligned}
\tag{3-40}
$$

Note that identities $H_{e,1}^X=2q^{-1/120}(\chi_0(q)-2)$ and $H_{e,7}^X=2q^{71/120}\chi_1(q)$ are predicted in §5.4 of [Cheng et al. 2014b], but it is not verified there that this specification yields a mock modular form with shadow $S^X=S_e^X$.

We determine the modular properties of $2T_{e,1}^{-}$ and $2T_{e,7}^{-}$ by applying the results of Zwegers on F_0, F_1, ϕ_0 and ϕ_1 that we summarized in Section 3B. To apply these results we first recall the expressions

$$
\begin{aligned}
\chi_0(q)&=2F_0(q)-\phi_0(-q),\\
\chi_1(q)&=2F_1(q)+q^{-1}\phi_1(-q),
\end{aligned}
\tag{3-41}
$$

which are proven in §3 of [Watson 1937]. (The first of these was given by Ramanujan in his last letter to Hardy, where he also mentioned the existence of a similar formula relating χ_1, F_1 and ϕ_1.) Thus we obtain

$$2T_{e,1}^{-}=4F_{5,1,3}(2\tau)-2F_{5,2,3}(2\tau),\tag{3-42}$$

$$2T_{e,7}^{-}=4F_{5,1,4}(2\tau)-2F_{5,2,4}(2\tau),\tag{3-43}$$

upon comparison of (3-24), (3-25), (3-40) and (3-41).

Applying the results of Zwegers on $F_{5,1}$ and $F_{5,2}$ recalled in Section 3B, and the equations (3-30) and (3-31) in particular, we conclude that $2T_{e,1}^{-}$ and $2T_{e,7}^{-}$ are mock modular forms of weight $\frac{1}{2}$, with respective shadows given by

$$3(S_{30,1}+S_{30,11}+S_{30,19}+S_{30,29})(\tau),\tag{3-44}$$

$$3(S_{30,7}+S_{30,13}+S_{30,17}+S_{30,23})(\tau).\tag{3-45}$$

In other words, the shadow of T_e^X is precisely S_e^X, as required. The modular transformation formulas for $H_{5,1}(\tau)$ and $H_{5,2}(\tau)$ given in Propositions 4.10 and 4.13 of [Zwegers 2002], respectively, show that T_e^X transforms in the desired way under $\mathrm{SL}_2(\mathbb{Z})$.

We now consider the case $o(g)=2$. We may take $g=\hat{\tau}$. We again begin by using the results recalled in Section 3B to analyze the components $T_{\hat{\tau},1}^{-}$ and $T_{\hat{\tau},7}^{-}$

separately. For $T^{-}_{\hat{\tau},1}$ let

$$A=\begin{pmatrix}6&4\\4&1\end{pmatrix},\quad \boldsymbol{a}=\begin{pmatrix}\frac{1}{10}\\ \frac{1}{10}\end{pmatrix},\quad \boldsymbol{b}=\begin{pmatrix}\frac{3}{20}\\ -\frac{2}{20}\end{pmatrix},\quad \boldsymbol{c}^{(1)}=\begin{pmatrix}-1\\ 4\end{pmatrix},\quad \boldsymbol{c}^{(2)}=\begin{pmatrix}-2\\ 3\end{pmatrix}. \tag{3-46}$$

Then a direct computation using

$$\begin{gathered}\nu=\begin{pmatrix}k+\frac{1}{10}\\ m+\frac{1}{10}\end{pmatrix},\qquad Q(\nu)=3k^2+\tfrac{1}{2}m^2+4km+k+\tfrac{1}{2}m+\tfrac{3}{40},\\ B(\nu,\boldsymbol{b})=\tfrac{1}{2}(k+m)+\tfrac{1}{10},\\ \operatorname{sgn}(B(\boldsymbol{c}^{(1)},\nu))=\operatorname{sgn}\bigl(k+\tfrac{1}{10}\bigr),\quad \operatorname{sgn}(B(\boldsymbol{c}^{(2)},\nu))=\operatorname{sgn}\bigl(-m-\tfrac{1}{10}\bigr),\end{gathered} \tag{3-47}$$

gives

$$2T^{-}_{\hat{\tau},1}=-\frac{e\bigl(-\frac{1}{10}\bigr)}{\eta(2\tau)}\sum_{\nu\in\boldsymbol{a}+\mathbb{Z}^2}\bigl(\operatorname{sgn}(B(\boldsymbol{c}^{(1)},\nu)) - \operatorname{sgn}(B(\boldsymbol{c}^{(2)},\nu))\bigr)e^{2\pi iQ(\nu)\tau+2\pi iB(\nu,\boldsymbol{b})}. \tag{3-48}$$

Comparing this to the indefinite theta function construction (3-14) we find that

$$\begin{aligned}\vartheta^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}_{\boldsymbol{a},\boldsymbol{b}}(\tau)&=-e\bigl(\tfrac{1}{10}\bigr)\eta(2\tau)2T^{-}_{\hat{\tau},1}(\tau)\\ &\quad+\sum_{\nu\in\boldsymbol{a}+\mathbb{Z}^2}\Biggl(\sum_{k=1}^{2}(-1)^k\operatorname{sgn}(B(\boldsymbol{c}^{(k)},\nu))\beta\biggl(-\frac{B(\boldsymbol{c}^{(k)},\nu)^2\Im(\tau)}{Q(\boldsymbol{c}^{(k)})}\biggr)\Biggr)\\ &\qquad\times q^{Q(\nu)}e^{2\pi iB(\nu,\boldsymbol{b})}.\end{aligned} \tag{3-49}$$

We now use Proposition 3.1 to rewrite the terms involving $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$ in the second line of (3-49). For the term with $\boldsymbol{c}^{(1)}$ the set P_0 of Proposition 3.1 has one element, $\mu_0=\frac{1}{10}\bigl(\begin{smallmatrix}-9\\1\end{smallmatrix}\bigr)$, and we find

$$\langle\boldsymbol{c}^{(1)}\rangle^{\perp}_{\mathbb{Z}}=\bigl\{\bigl(\begin{smallmatrix}0\\m\end{smallmatrix}\bigr)\mid m\in\mathbb{Z}\bigr\},\quad \boldsymbol{b}^{\perp}=\tfrac{1}{2}\bigl(\begin{smallmatrix}0\\1\end{smallmatrix}\bigr),\quad \mu_0^{\perp}=\tfrac{1}{2}\bigl(\begin{smallmatrix}0\\-7\end{smallmatrix}\bigr). \tag{3-50}$$

Thus

$$\sum_{\xi\in\mu_0^{\perp}+\langle c\rangle^{\perp}_{\mathbb{Z}}}e^{2\pi iQ(\xi)\tau+2\pi iB(\xi,b^{\perp})}=e\bigl(-\tfrac{1}{4}\bigr)\sum_{m\in\mathbb{Z}}(-1)^m q^{(m-1/2)^2/2}=0, \tag{3-51}$$

so this term vanishes.

For the term with $\boldsymbol{c}^{(2)}$ the set P_0 has three elements, $\mu_0=\frac{1}{10}\bigl(\begin{smallmatrix}1\\1\end{smallmatrix}\bigr),\frac{1}{10}\bigl(\begin{smallmatrix}1\\11\end{smallmatrix}\bigr),\frac{1}{10}\bigl(\begin{smallmatrix}1\\21\end{smallmatrix}\bigr)$, and we have $B(\boldsymbol{c}^{(2)},\mu_0)/2Q(\boldsymbol{c}^{(2)})=\frac{1}{30},\frac{11}{30},\frac{21}{30}$, in the respective cases. The last value of μ_0 also leads to a vanishing contribution, while the other two lead to

$$-e\bigl(\tfrac{1}{12}\bigr)R_{\frac{1}{30},-\frac{1}{2}}(15\tau)\eta(2\tau)-e\bigl(-\tfrac{1}{12}\bigr)R_{\frac{11}{30},-\frac{1}{2}}(15\tau)\eta(2\tau), \tag{3-52}$$

which we see by applying Euler's identity

$$q^{1/12}\sum_{k\in\mathbb{Z}}(-1)^k q^{3k^2+k}=\eta(2\tau). \tag{3-53}$$

We thus have

$$-e\left(-\tfrac{1}{10}\right)\frac{\vartheta^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}_{\boldsymbol{a},\boldsymbol{b}}(\tau)}{\eta(2\tau)} \\ = 2T^-_{\hat{\tau},1}-e\left(-\tfrac{1}{60}\right)R_{\frac{1}{30},-\frac{1}{2}}(15\tau)-e\left(-\tfrac{11}{60}\right)R_{\frac{11}{30},-\frac{1}{2}}(15\tau). \tag{3-54}$$

In particular, $T^-_{\hat{\tau},1}$ is the Fourier expansion of a holomorphic function on $\mathbb{H}$, which we henceforth denote $T^-_{\hat{\tau},1}(\tau)$.

Since $T^-_{\hat{\tau},1}(\tau)$ is holomorphic, the function (3-54) is a harmonic weak Maass form of weight $\frac{1}{2}$, according to Proposition 4.2 of [Zwegers 2002] (see also Section 3A). Thus we are in a directly similar situation to that encountered at the end of Section 3B. Namely, we have that $T^-_{\hat{\tau},1}(\tau)$ is a mock modular form of weight $\frac{1}{2}$ for some congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and the second and third summands of the right-hand side of (3-54) comprise the Eichler integral of its shadow. Applying (3-16), (3-17) and (3-18), and also

$$e\left(-\tfrac{1}{60}\right)g_{\frac{1}{30},\frac{1}{2}}(15\tau)+e\left(-\tfrac{11}{60}\right)g_{\frac{11}{30},\frac{1}{2}}(15\tau) \\ =\tfrac{1}{30}(S_{30,1}+S_{30,11}+S_{30,19}+S_{30,29})(\tau), \tag{3-55}$$

we conclude that the shadow of $2T^-_{\hat{\tau},1}(\tau)$ is indeed $S^X_{\hat{\tau},1}(\tau)$ (see (3-33)).

For $T^-_{\hat{\tau},7}$ we take A, $\boldsymbol{b}$, $\boldsymbol{c}^{(1)}$, $\boldsymbol{c}^{(2)}$ as before but set $\boldsymbol{a}=\frac{1}{10}\binom{3}{3}$. We now have

$$\nu=\begin{pmatrix}k+\frac{3}{10}\\ m+\frac{3}{10}\end{pmatrix},\quad Q(\nu)=3k^2+\tfrac{1}{2}m^2+4km+3k+\tfrac{3}{2}m-\tfrac{27}{40}, \\ B(\nu,\boldsymbol{b})=\tfrac{1}{2}(k+m)+\tfrac{3}{10}, \\ \operatorname{sgn}(B(\boldsymbol{c}^{(1)},\nu))=\operatorname{sgn}\left(k+\tfrac{3}{10}\right),\quad \operatorname{sgn}(B(\boldsymbol{c}^{(2)},\nu))=\operatorname{sgn}\left(-m-\tfrac{3}{10}\right). \tag{3-56}$$

Proceeding as we did for $T^-_{\hat{\tau},1}$, the contribution from the $\boldsymbol{c}^{(1)}$ term vanishes again. For the $\boldsymbol{c}^{(2)}$ term, P_0 consists of the three values $\mu_0=\frac{1}{10}\binom{3}{3},\frac{1}{10}\binom{3}{13},\frac{1}{10}\binom{3}{23}$, and we have

$$\frac{B(\boldsymbol{c}^{(2)},\mu_0)}{2Q(\boldsymbol{c}^{(2)})}=\frac{3}{30},\ \frac{13}{30},\ \frac{23}{30}, \tag{3-57}$$

respectively. The first value of μ_0 leads to a vanishing contribution while the other two terms lead to

$$-e\big(-\tfrac{3}{10}\big)\frac{\vartheta^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}_{\boldsymbol{a},\boldsymbol{b}}(\tau)}{\eta(2\tau)} = 2T^-_{\hat{\tau},7} - e\big(-\tfrac{13}{60}\big)R_{\frac{13}{30},-\frac{1}{2}}(15\tau) - e\big(-\tfrac{23}{60}\big)R_{\frac{23}{30},-\frac{1}{2}}(15\tau). \quad \text{(3-58)}$$

We conclude thus that $T^-_{\hat{\tau},7}$ is the Fourier expansion of a mock modular form of weight $\frac{1}{2}$, and using

$$e\big(-\tfrac{13}{60}\big)g_{\frac{13}{30},\frac{1}{2}}(15\tau) + e\big(-\tfrac{23}{60}\big)g_{\frac{23}{30},\frac{1}{2}}(15\tau) = \tfrac{1}{30}(S_{30,7}+S_{30,13}+S_{30,17}+S_{30,23})(\tau), \quad \text{(3-59)}$$

we see that the shadow of $2T^-_{\hat{\tau},1}(\tau)$ is $S^X_{\hat{\tau},1}(\tau)$ (see (3-33)). So we have verified that the shadow of $2T^-_g = (2T^-_{g,r})$ is $S^X_g = (S^X_{g,r})$ for $o(g) = 2$.

Corollary 2.9 of [Zwegers 2002] details the modular transformation properties of the indefinite theta functions $\vartheta^{\boldsymbol{c}^{(1)},\boldsymbol{c}^{(2)}}_{\boldsymbol{a},\boldsymbol{b}}(\tau)$. Applying these formulas, much as in the proofs of Propositions 4.10 and 4.13. in [Zwegers 2002], we see that $2T^-_{\hat{\tau}}$ transforms in the desired way under the action of $\Gamma_0(2)$.

Corollary 2.9 also enables us to compute the expansion of $2T^-_{\hat{\tau}}$ at the cusp of $\Gamma_0(2)$ represented by 0. We ultimately find that both $T^-_{\hat{\tau},1}(\tau)$ and $T^-_{\hat{\tau},7}(\tau)$ vanish as $\tau \to 0$. Thus $2T^-_{\hat{\tau}}$ has no poles away from the infinite cusp.

It remains to consider the case $o(g) = 3$, but this can be handled by applying classical results on positive-definite theta functions, since the formula (2-40) gives $T^-_{\hat{\sigma},1}$ and $T^-_{\hat{\sigma},7}$ explicitly in terms of the Dedekind eta function and the theta series of a rank 1 lattice. We easily check that these functions transform in the desired way under $\Gamma_0(3)$, and have no poles away from the infinite cusp of $\Gamma_0(3)$. In particular, $2T^-_{\hat{\sigma}}$ is modular, and has vanishing shadow. □

We are now ready to prove our main results.

Proof of Theorem 1.1. Proposition 3.2 demonstrates that the functions $2T^X_g$ are mock modular forms of weight $\frac{1}{2}$ with the claimed shadows, multiplier systems and polar parts. It remains to verify that they are the unique such functions.

The uniqueness in case $g = e$ is shown in Corollary 4.2 of [Cheng et al. 2014b], using the fact that there are no weak Jacobi forms of weight 1 (see Theorem 9.7 in [Dabholkar et al. 2012]). We give a different (but certainly related) argument here.

Consider first the case that $o(g)$ is 1 or 2. It suffices to show that if $h = (h_r)$ is a modular form of weight $\frac{1}{2}$, transforming with the same multiplier system as H^X under $\Gamma_0(2)$, with h_r vanishing whenever r does not belong to

$$\{\pm 1, \pm 7, \pm 11, \pm 13, \pm 17, \pm 19, \pm 23, \pm 29\}, \quad \text{(3-60)}$$

then h vanishes identically. The multiplier system for H^X is trivial when restricted to $\Gamma(120)$, so the components h_r are modular forms for $\Gamma_0(2) \cap \Gamma(120) = \Gamma(120)$.

Satz 5.2 of [Skoruppa 1985] is an effective version of the celebrated theorem of Serre and Stark [1977] on modular forms of weight $\frac{1}{2}$ for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$. It tells us that the space of modular forms of weight $\frac{1}{2}$ for $\Gamma(120)$ is spanned by certain linear combinations of the thetanullwerte $\theta^0_{n,r}(\tau) := \theta_{n,r}(\tau, 0)$, and the only n that can appear are those that divide 30. On the other hand, the restriction (3-60) implies that any nonzero component h_r must belong to one of $q^{-1/120}\mathbb{C}[\![q]\!]$ or $q^{71/120}\mathbb{C}[\![q]\!]$. We conclude that all the h_r are necessarily zero by checking, using

$$\theta^0_{n,r}(\tau) = \sum_{k\in\mathbb{Z}} q^{(2kn+r)^2/4n}, \tag{3-61}$$

that none of the $\theta^0_{n,r}$ belong to either space, for n a divisor of 30.

The case that $o(g) = 3$ is very similar, except that the h_r are now modular forms on $\Gamma_0(9)\cap\Gamma(120)$, which contains $\Gamma(360)$, and the relevant thetanullwerte are those $\theta^0_{n,r}$ with n a divisor of 90. We easily check using (3-61) that there are nonzero possibilities for h_r, and this completes the proof. □

Proof of Theorem 1.2. Taking now (1-6) as the definition of H^X_g, the identities (1-7) follow directly from the definition (3-32) of T^X_g, and the explicit expressions (2-38) for the components of T^X_e.

The identities (1-8) follow from the characterization of H^X_g for $o(g) = 2$ that is entailed in Theorem 1.1. Indeed, using Zwegers' results (viz., Propositions 4.10 and 4.13 in [Zwegers 2002]) on the modularity of $\phi_0(-q)$ and $\phi_1(-q)$, we see that the function defined by the right-hand side of (1-8) is a vector-valued mock modular form with exactly the same shadow as $2T^X_{\hat{\tau}}$, transforming with the same multiplier system under $\Gamma_0(2)$, and having the same polar parts at both the infinite and noninfinite cusps of $\Gamma_0(2)$. So it must coincide with $H^X_{2A,1} = 2T^X_{\hat{\tau}}$ according to Theorem 1.1. This completes the proof. □

Proof of Corollary 1.3. Andrews [1986] established Hecke-type "double sum" identities for ϕ_0 and ϕ_1. Rewriting these slightly, we find

$$\phi_0(-q) = \frac{(q;q)_\infty}{(q^2;q^2)^2_\infty}\Biggl(\sum_{k,m\geq 0} - \sum_{k,m<0}\Biggr)_{k=m \pmod 2} (-1)^m \times q^{k^2/2+m^2/2+4km+k/2+3m/2}, \tag{3-62}$$

$$-q^{-1}\phi_1(-q) = \frac{(q;q)_\infty}{(q^2;q^2)^2_\infty}\Biggl(\sum_{k,m\geq 0} - \sum_{k,m<0}\Biggr)_{k=m \pmod 2} (-1)^m \times q^{k^2/2+m^2/2+4km+3k/2+5m/2}. \tag{3-63}$$

Armed with the identities (1-8), we obtain (1-10) and (1-11) by comparing (3-62) and (3-63) with the explicit expression (2-39) for the components of $T^X_{\hat{\tau}}$. □

Appendix: Coefficients

$[g]$	1A	2A	3A
Γ_g	1|1	2|1	3|3
−1	−2	−2	−2
119	2	2	2
239	2	−2	2
359	4	0	−2
479	2	−2	2
599	6	2	0
719	4	0	−2
839	6	2	0
959	6	−2	0
1079	10	2	−2
1199	6	−2	0
1319	12	0	0
1439	10	−2	−2
1559	14	2	2
1679	14	−2	2
1799	18	2	0
1919	14	−2	2
2039	24	4	0
2159	22	−2	−2
2279	26	2	2
2399	26	−2	2
2519	34	2	−2
2639	30	−2	0
2759	42	2	0
2879	40	−4	−2
2999	48	4	0
3119	48	−4	0
3239	58	2	−2
3359	56	−4	2
3479	72	4	0
3599	70	−2	−2
3719	80	4	2
3839	84	−4	0
3959	100	4	−2
4079	96	−4	0
4199	116	4	2
4319	116	−4	−4
4439	134	6	2
4559	140	−4	2

Table 1. $H^X_{g,1}$, $X = E_8^3$

$[g]$	1A	2A	3A
Γ_g	1|1	2|1	3|3
71	2	−2	2
191	4	0	−2
311	4	0	−2
431	6	2	0
551	6	−2	0
671	8	0	2
791	8	0	2
911	12	0	0
1031	10	−2	−2
1151	14	2	2
1271	16	0	−2
1391	18	2	0
1511	18	−2	0
1631	24	0	0
1751	24	0	0
1871	30	2	0
1991	30	−2	0
2111	36	0	0
2231	38	−2	2
2351	46	2	−2
2471	46	−2	−2
2591	54	2	0
2711	60	0	0
2831	66	2	0
2951	68	−4	2
3071	82	2	−2
3191	84	0	0
3311	98	2	2
3431	102	−2	0
3551	114	2	0
3671	122	−2	2
3791	138	2	0
3911	144	−4	0
4031	162	2	0
4151	174	−2	0
4271	192	4	0
4391	200	−4	2
4511	226	2	−2
4631	238	−2	−2

Table 2. $H^X_{g,7}$, $X = E_8^3$

Acknowledgements

We thank Miranda Cheng for particularly helpful discussions and advice that took place in the early stages of this work, and we thank Ching Hung Lam for discussions on the vertex operator algebra structure here employed. We thank the referees for many helpful comments and suggestions. The research of Duncan was supported in part by the Simons Foundation (#316779). Both authors gratefully acknowledge support from the U.S. National Science Foundation (grants 1203162, 1214409 and 1601306).

References

[Andrews 1986] G. E. Andrews, "The fifth and seventh order mock theta functions", *Trans. Amer. Math. Soc.* **293**:1 (1986), 113–134. MR Zbl

[Borcherds 1986] R. E. Borcherds, "Vertex algebras, Kac–Moody algebras, and the monster", *Proc. Nat. Acad. Sci. U.S.A.* **83**:10 (1986), 3068–3071. MR Zbl

[Borcherds 1992] R. E. Borcherds, "Monstrous moonshine and monstrous Lie superalgebras", *Invent. Math.* **109**:2 (1992), 405–444. MR Zbl

[Bruinier and Funke 2004] J. H. Bruinier and J. Funke, "On two geometric theta lifts", *Duke Math. J.* **125**:1 (2004), 45–90. MR Zbl

[Cheng et al. 2014a] M. C. N. Cheng, J. F. R. Duncan, and J. A. Harvey, "Umbral moonshine", *Commun. Number Theory Phys.* **8**:2 (2014), 101–242. MR Zbl

[Cheng et al. 2014b] M. C. N. Cheng, J. F. R. Duncan, and J. A. Harvey, "Umbral moonshine and the Niemeier lattices", *Res. Math. Sci.* **1** (2014), art. id. 3, 1–81. MR Zbl

[Cheng et al. ≥ 2017] M. C. N. Cheng, J. F. R. Duncan, and J. A. Harvey, "Weight one Jacobi forms and umbral moonshine", in preparation.

[Conway and Norton 1979] J. H. Conway and S. P. Norton, "Monstrous moonshine", *Bull. London Math. Soc.* **11**:3 (1979), 308–339. MR Zbl

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups*, Oxford University Press, 1985. MR Zbl

[Dabholkar et al. 2012] A. Dabholkar, S. Murthy, and D. Zagier, "Quantum black holes, wall crossing, and mock modular forms", preprint, 2012. arXiv

[Dong 1993] C. Dong, "Vertex algebras associated with even lattices", *J. Algebra* **161**:1 (1993), 245–265. MR Zbl

[Dong and Mason 1994] C. Dong and G. Mason, "Nonabelian orbifolds and the boson-fermion correspondence", *Comm. Math. Phys.* **163**:3 (1994), 523–559. MR Zbl

[Dong and Zhao 2005] C. Dong and Z. Zhao, "Modularity in orbifold theory for vertex operator superalgebras", *Comm. Math. Phys.* **260**:1 (2005), 227–256. MR Zbl

[Dong et al. 2000] C. Dong, H. Li, and G. Mason, "Modular-invariance of trace functions in orbifold theory and generalized moonshine", *Comm. Math. Phys.* **214**:1 (2000), 1–56. MR Zbl

[Duncan et al. 2015a] J. F. R. Duncan, M. J. Griffin, and K. Ono, "Moonshine", *Res. Math. Sci.* **2** (2015), art. id. 11, 1–57. MR Zbl

[Duncan et al. 2015b] J. F. R. Duncan, M. J. Griffin, and K. Ono, "Proof of the umbral moonshine conjecture", *Res. Math. Sci.* **2** (2015), art. id. 26, 1–47. MR Zbl

[Eguchi et al. 2011] T. Eguchi, H. Ooguri, and Y. Tachikawa, "Notes on the $K3$ surface and the Mathieu group M_{24}", *Exp. Math.* **20**:1 (2011), 91–96. MR Zbl

[Feingold et al. 1991] A. J. Feingold, I. B. Frenkel, and J. F. X. Ries, *Spinor construction of vertex operator algebras, triality, and $E_8^{(1)}$*, Contemporary Mathematics **121**, American Mathematical Society, Providence, RI, 1991. MR Zbl

[Feingold et al. 1996] A. J. Feingold, J. F. X. Ries, and M. D. Weiner, "Spinor construction of the $c = \frac{1}{2}$ minimal model", pp. 45–92 in *Moonshine, the monster, and related topics* (South Hadley, MA, 1994), edited by C. Dong and G. Mason, Contemp. Math. **193**, American Mathematical Society, Providence, RI, 1996. MR Zbl

[Frenkel and Ben-Zvi 2004] E. Frenkel and D. Ben-Zvi, *Vertex algebras and algebraic curves*, 2nd ed., Mathematical Surveys and Monographs **88**, American Mathematical Society, Providence, RI, 2004. MR Zbl

[Frenkel and Szczesny 2004] E. Frenkel and M. Szczesny, "Twisted modules over vertex algebras on algebraic curves", *Adv. Math.* **187**:1 (2004), 195–227. MR Zbl

[Frenkel et al. 1984] I. B. Frenkel, J. Lepowsky, and A. Meurman, "A natural representation of the Fischer–Griess monster with the modular function J as character", *Proc. Nat. Acad. Sci. U.S.A.* **81**:10, phys. sci. (1984), 3256–3260. MR Zbl

[Frenkel et al. 1985] I. B. Frenkel, J. Lepowsky, and A. Meurman, "A moonshine module for the monster", pp. 231–273 in *Vertex operators in mathematics and physics* (Berkeley, 1983), edited by J. Lepowsky et al., Math. Sci. Res. Inst. Publ. **3**, Springer, New York, 1985. MR Zbl

[Frenkel et al. 1988] I. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the monster*, Pure and Applied Mathematics **134**, Academic Press, Boston, 1988. MR Zbl

[Gannon 2016] T. Gannon, "Much ado about Mathieu", *Adv. Math.* **301** (2016), 322–358. MR Zbl

[Harada and Lang 1998] K. Harada and M. L. Lang, "Modular forms associated with the monster module", pp. 59–83 in *The monster and Lie algebras* (Columbus, OH, 1996), edited by J. Ferrar and K. Harada, Ohio State Univ. Math. Res. Inst. Publ. **7**, de Gruyter, Berlin, 1998. MR Zbl

[Miyamoto 2004] M. Miyamoto, "Modular invariance of vertex operator algebras satisfying C_2-cofiniteness", *Duke Math. J.* **122**:1 (2004), 51–91. MR Zbl

[Ono 2009] K. Ono, "Unearthing the visions of a master: harmonic Maass forms and number theory", pp. 347–454 in *Current developments in mathematics* (Cambridge, MA, 2008), edited by D. Jerison et al., International Press, Boston, 2009. MR Zbl

[Ramanujan 1988] S. Ramanujan, *The lost notebook and other unpublished papers*, Springer, Berlin, 1988. MR Zbl

[Ramanujan 2000] S. Ramanujan, *Collected papers of Srinivasa Ramanujan*, edited by G. H. Hardy et al., AMS Chelsea Publishing, Providence, RI, 2000. MR Zbl

[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 Mathematics **627**, Springer, Berlin, 1977. MR Zbl

[Skoruppa 1985] N.-P. Skoruppa, *Über den Zusammenhang zwischen Jacobiformen und Modulformen halbganzen Gewichts*, Bonner Mathematische Schriften **159**, Universität Bonn, 1985. MR

[Thompson 1979a] J. G. Thompson, "Finite groups and modular functions", *Bull. London Math. Soc.* **11**:3 (1979), 347–351. MR Zbl

[Thompson 1979b] J. G. Thompson, "Some numerology between the Fischer–Griess monster and the elliptic modular function", *Bull. London Math. Soc.* **11**:3 (1979), 352–353. MR Zbl

[Van Ekeren 2013] J. Van Ekeren, "Modular invariance for twisted modules over a vertex operator superalgebra", *Comm. Math. Phys.* **322**:2 (2013), 333–371. MR Zbl

[Van Ekeren 2014] J. Van Ekeren, "Vertex operator superalgebras and odd trace functions", pp. 223–234 in *Advances in Lie superalgebras* (Rome, 2012), edited by M. Gorelik and P. Papi, Springer INdAM Ser. **7**, Springer, Cham, Switzerland, 2014. MR Zbl

[Watson 1937] G. N. Watson, "The mock theta functions (2)", *Proc. London Math. Soc.* **S2-42**:1 (1937), 274. MR Zbl

[Zagier 2009] D. Zagier, "Ramanujan's mock theta functions and their applications (after Zwegers and Ono–Bringmann)", exposé 986, 143–164 in *Séminaire Bourbaki*, Astérisque **326**, 2009. MR Zbl

[Zhu 1996] Y. Zhu, "Modular invariance of characters of vertex operator algebras", *J. Amer. Math. Soc.* **9**:1 (1996), 237–302. MR Zbl

[Zwegers 2002] S. Zwegers, *Mock theta functions*, Ph.D. thesis, Universiteit Utrecht, 2002, available at https://dspace.library.uu.nl/handle/1874/878. Zbl

[Zwegers 2009] S. Zwegers, "On two fifth order mock theta functions", *Ramanujan J.* **20**:2 (2009), 207–214. MR Zbl

Communicated by Richard Borcherds
Received 2015-01-27 Revised 2016-10-05 Accepted 2016-12-18

john.duncan@emory.edu *Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, United States*

j-harvey@uchicago.edu *Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, IL 60637, United States*

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.537

Geometry on totally separably closed schemes

Stefan Schröer

We prove, for quasicompact separated schemes over ground fields, that Čech cohomology coincides with sheaf cohomology with respect to the Nisnevich topology. This is a partial generalization of Artin's result that for noetherian schemes such an equality holds with respect to the étale topology, which holds under the assumption that every finite subset admits an affine open neighborhood (AF-property). Our key result is that on the absolute integral closure of separated algebraic schemes, the intersection of any two irreducible closed subsets remains irreducible. We prove this by establishing general modification and contraction results adapted to inverse limits of schemes. Along the way, we characterize schemes that are acyclic with respect to various Grothendieck topologies, study schemes all local rings of which are strictly henselian, and analyze fiber products of strict localizations.

MSC2010: primary 14F20; secondary 14E05, 13B22, 13J15.

Keywords: Absolute algebraic closure, acyclic schemes, étale and Nisnevich topology, henselian rings, Čech and sheaf cohomology, contractions.

Introduction

An integral scheme X that is normal and whose function field $k(X)$ is algebraically closed is called *totally algebraically closed*, or *absolutely algebraically closed*. These notion were introduced by Enochs [1968] and Artin [1971] and further studied, for example, in [Hochster 1970; Borho and Weber 1971].

Artin used such schemes to prove that Čech cohomology coincides with sheaf cohomology for any abelian sheaf in the étale topos of a noetherian scheme X that satisfies the *AF-property*, that is, every finite subset admits an affine open neighborhood. This is a rather large and very useful class of schemes, which includes all schemes that are quasiprojective over some affine scheme. Note, however, that there are smooth threefolds lacking this property (see Hironaka's example [Hartshorne 1977, Appendix B, Example 3.4.1]).

In recent years, absolutely closed domains were studied in connection with tight closure theory. Hochster and Huneke [1992] showed that the absolute algebraic closure is a big Cohen–Macaulay module in characteristic $p > 0$. This was further extended by Huneke and Lyubeznik [2007]. Schoutens [2004] and Aberbach [2005] used the absolute closure to characterize regularity for local rings. Huneke [2011] also gave a nice survey on absolute algebraic closure.

Other applications include Gabber's rigidity property [1994] for abelian torsion sheaves for affine henselian pairs, or Rydh's study [Rydh 2010] of descent questions, or the proof by Bhatt and de Jong [2014] of the Lefschetz theorem for local Picard groups. Closely related ideas occur in the proétale site, introduced by Bhatt and Scholze [2015], or the construction of universal coverings for schemes by Vakil and Wickelgren [2011]. Absolute integral closure in characteristic $p > 0$ provides examples of flat ring extensions that are not a filtered colimit of finitely presented flat ring extensions, as Bhatt [2014] noted. I have used absolute closure to study points in the fppf topos [Schröer 2014].

For aesthetic reasons, and also to stress the relation to strict localization and étale topology, I prefer to work with separable closure instead of algebraic closure: if X_0 is an integral scheme, the *total separable closure* $X = \mathrm{TSC}(X_0)$ is the integral closure of X_0 in some chosen separable closure of the function field. The goal of this paper is to make a systematic study of geometric properties of $X = \mathrm{TSC}(X_0)$, and to apply it to cohomological questions. One of our main results is the following rather counterintuitive property:

Theorem (see Theorem 12.1). *Let X_0 be separated and of finite type over a ground field k, and $X = \mathrm{TSC}(X_0)$. Then for every pair of closed irreducible subsets $A, B \subset X$, the intersection $A \cap B$ remains irreducible.*

Note that for algebraic surfaces, this means that for all closed irreducible subsets $A \neq B$ in X, the intersection $A \cap B$ contains at most one point. I conjecture that our result actually holds true for arbitrary integral schemes that are quasicompact and separated.

Such geometric properties were crucial for Artin to establish the equality

$$\check{H}^p(X_{\text{et}}, F) = H^p(X_{\text{et}}, F).$$

He achieved this by assuming the AF-property. This condition, however, appears to be somewhat alien to the problem. Indeed, for locally factorial schemes, the AF-property is equivalent to quasiprojectivity, according to Kleiman's proof of the Chevalley conjecture [Kleiman 1966, Theorem 3, p. 327]. Recently, this was extended to arbitrary normal schemes by Benoist [2013]. Note, however, that nonnormal schemes may have the AF-property without being quasiprojective, as examples of Horrocks [1971] and Ferrand [2003] show.

For schemes X lacking the AF-property, Artin's argument break down at two essential steps: First, he uses the AF-property to reduce the analysis of fiber products $\operatorname{Spec}(\mathbb{O}^s_{X,x}) \times_X \operatorname{Spec}(\mathbb{O}^s_{X,y})$ of strict henselizations to the more accessible case of integral affine schemes $X = \operatorname{Spec}(R)$, in fact to the situation $A = \mathbb{Z}[T_1, \dots, T_n]$, which then leads to the join construction $B = [A^h_p, A^h_q]$ inside some algebraic closure of the field of fractions $\operatorname{Frac}(A)$ (see [Artin 1971, Theorem 2.2]). Second, he needs the affine situation to form meaningful semilocal intersection rings $R = A^h_p \cap A^h_q$, in order to prove that B has separably closed residue field (see [Artin 1971, Theorem 2.5]).

My motivation to study totally separably closed schemes was to bypass the AF-property in Artin's arguments. In some sense, I was able to generalize half of his reasoning to arbitrary schemes X, namely those steps that pertain to henselization $\mathbb{O}^h_{X,x}$ for points $x \in X$ rather then strict localizations $\mathbb{O}^s_{X,a}$ for geometric points $a : \operatorname{Spec}(\Omega) \to X$, the latter having in addition separably closed residue fields.

In terms of Grothendieck topologies, we thus get results on the *Nisnevich topology* rather then the étale topology. This is one of the more recent Grothendieck topologies, which was considered in connection with motivic questions. Recall that the covering families of the Nisnevich topology on $(\text{Et}\,/\,X)$ are those $(U_\lambda \to U)_\lambda$ so that each $U_\lambda \to U$ is *completely decomposed*, that is, over each point lies at least one point with the same residue field, and that $\bigcup U_\lambda \to U$ is surjective. We refer to Nisnevich [1989] and the Stacks project [2005–] for more details. Note that the Nisnevich topology plays an important role in Voevodsky's theory of sheaves with transfer and motivic cohomology, see [Voevodsky et al. 2000, Chapter 3] and [Mazza et al. 2006]. Our second main result is:

Theorem (see Theorem 13.1). *Let X be a quasicompact and separated scheme over a ground field k. Then*

$$\check{H}^p(X_{\text{Nis}}, F) = H^p(X_{\text{Nis}}, F)$$

for all abelian Nisnevich sheaves on X.

It is quite sad that my methods apparently need a ground field, in order to use the geometry of contractions, which lose some of their force over more general ground rings. Again I conjecture that the result holds true for quasicompact and separated schemes, even for the étale topology. Indeed, this paper contains several general *reduction steps*, which reveal that it suffices to prove this conjecture merely for separated integral $\mathbb{Z}$-schemes of finite type. It seems likely that it suffices that the diagonal $\Delta : X \to X \times X$ is affine, rather than closed.

Along the way, it is crucial to characterize schemes that are acyclic with respect to various Grothendieck topologies. With analogous results for the Zariski and the étale topology, we have:

Theorem (see Theorem 4.2). *Let X be a quasicompact scheme. Then the following are equivalent*:

(i) $H^p(X_{\mathrm{Nis}}, F) = 0$ *for every abelian Nisnevich sheaf F and every $p \geq 1$.*

(ii) *Every completely decomposed étale surjection $U \to X$ admits a section.*

(iii) *The scheme X is affine*, *and each connected component is local henselian.*

(iv) *The scheme X is affine*, *each irreducible component is local henselian*, *and the space* $\mathrm{Max}(X)$ *is at most zero-dimensional.*

The paper is organized as follows: In Section 1, we review some properties of schemes that are stable by integral surjections. These will be important for many reduction steps that follow. Section 2 contains a discussion of schemes all of whose local rings are strictly local. Such schemes X have the crucial property that any integral morphism $f : Y \to X$ with Y irreducible must be injective. In Section 3 we discuss the total separable closure $X = \mathrm{TSC}(X_0)$ of an integral scheme X_0, which are the most important examples of schemes that are everywhere strictly local.

Section 4 contains our characterizations of acyclic schemes with respect to three Grothendieck topologies, namely Zariski, Nisnevich and étale. In Section 5, we introduce technical conditions, namely the weak/strong Cartan–Artin properties, which roughly speaking means that the fiber products of strict localizations are acyclic with respect to the Nisnevich/étale topology. Note that such fiber products are almost always nonnoetherian. Sections 6 and 7 contain reduction arguments, which basically show that it suffices to check the Cartan–Artin properties on total separable closures $X = \mathrm{TSC}(X_0)$. On the latter, it translates into a simple, but rather counterintuitive geometric condition on the intersection of irreducible closed subsets.

Section 8 reveals in the special case of algebraic surfaces that this geometric condition indeed holds. It is used here that one understands very well which integral curves on a normal surface are contractible to a point. The next three sections prepare the ground to generalize this to higher dimensions: In Section 9, we collect some facts on noetherian schemes concerning quasiprojectivity and connectedness of divisors. The latter is an application of Grothendieck's connectedness theorem. In Section 10 we show how to make a closed subset contractible on some modification. In Section 11, we introduce the technical notion of cyclic systems, which is better suited to understand contractions in inverse limits like $X = \mathrm{TSC}(X_0)$. Having this, Section 12 contains Theorem 12.1: on total separable closures there are no cyclic systems, in particular the intersection of irreducible closed subsets remains irreducible. The application to Nisnevich cohomology appears in Section 13. There are also three appendices, discussing Lazard's observation on connected components of schemes, H. Cartan's argument on equality of Čech and sheaf cohomology, and some results on inductive dimension in general topology used in this paper.

1. Integral surjections

Throughout the paper, integral surjections and inverse limits play an important role. We start by reexamining these concepts.

Let X_λ, $\lambda \in L$, be a filtered inverse system of schemes with affine transition morphisms $X_\mu \to X_\lambda$, $\lambda \leq \mu$. Then the corresponding inverse limit exists as a scheme. For its construction, one may tacitly assume that there is a smallest index $\lambda = 0$, and regard the $X_\lambda = \mathrm{Spec}(\mathscr{A}_\lambda)$ as relatively affine schemes over X_0. Then

$$X = \varprojlim(X_\lambda) = \mathrm{Spec}(\mathscr{A}), \quad \mathscr{A} = \varinjlim \mathscr{A}_\lambda.$$

Moreover, the underlying topological space of X is the inverse limit of the underlying topological spaces for X_λ, by [EGA IV$_3$ 1966, Proposition 8.2.9]. In turn, for each point $x = (x_\lambda) \in X$ the canonical map $\varinjlim(\mathscr{O}_{X_\lambda, x_\lambda}) \to \mathscr{O}_{X,x}$ is bijective. Consequently, the residue field $\kappa(x)$ is the union of the $\kappa(x_\lambda)$, viewed as subfields.

Recall that a homomorphism of rings $R \to A$ is *integral* if each element in A is the root of a monic polynomial with coefficients in R. A morphism of schemes $f : X \to Y$ is called integral if there is a affine open covering $Y = \bigcup V_i$ so that $U_i = f^{-1}(V_i)$ is affine and $A_i = \Gamma(V_i, \mathscr{O}_X)$ is integral as an algebra over $R_i = \Gamma(V_i, \mathscr{O}_Y)$. Note that the underlying topological space of the fibers $f^{-1}(y)$ are profinite.

Of particular interest are those $f : X \to Y$ that are integral and surjective. The following locution will be useful throughout: Let $\mathscr{P}$ be a class of schemes. We say that $\mathscr{P}$ is *stable under images of integral surjections* if for each integral surjection $f : X \to Y$ where the domain X belongs to $\mathscr{P}$, the range Y belongs to $\mathscr{P}$ as well. For example:

Theorem 1.1. *The class of affine schemes is stable under images of integral surjections.*

In this generality, the result is due to Rydh [2015], who established it even for algebraic spaces. It generalizes Chevalley's theorem [EGA II 1961, Theorem 6.7.1], where Y is a noetherian scheme and f is finite surjective. See also [Conrad 2007, Corollary A.2] for the case that Y is an arbitrary scheme and f is finite surjective.

A scheme Y is called *local* if it is quasicompact and contains precisely one closed point. Equivalently Y is the spectrum of a local ring. We have the following permanence property:

Proposition 1.2. *The class of local schemes is stable under images of integral surjections.*

Proof. Let $f : X \to Y$ be an integral surjection, with X local. We have to show that Y is local. According to Theorem 1.1, the scheme Y is affine. It follows that Y contains at least one closed point. Let $y, y' \in Y$ be two closed points. Since f is surjective, there are closed points $x, x' \in X$ mapping to y, y', respectively. Since X is local, we have $x = x'$, whence $y = y'$. □

A scheme Y is called *local henselian* if it is local, and for every finite morphism $g : Y' \to Y$, the domain Y' is a sum of local schemes. Of course, there are only finitely many such summands, because $g^{-1}(b)$, where $b \in Y$ is the closed point, contains only finitely many points and g is a closed map.

Proposition 1.3. *The class of local henselian schemes is stable under images of integral surjections.*

Proof. Let $f : X \to Y$ be an integral surjection, with X local henselian. By Proposition 1.2, the scheme Y is local. Now let $Y' \to Y$ be a finite morphism, and consider the induced finite morphism $X' = Y' \times_Y X \to X$. Then $X' = X'_1 \amalg \cdots \amalg X'_n$ for some local schemes X'_i, $1 \leq i \leq n$. Their images $Y'_i \subset Y'$ are closed, because f and the induced morphism $X' \to Y'$ are integral. Moreover, these images form a closed covering of Y, since f and the induced morphism $X' \to Y'$ is surjective. Regarding the Y_i as reduced schemes, the canonical morphism $X'_i \to Y'_i$ is integral and surjective. Using Proposition 1.2 again, we conclude that the Y'_i are local. It follows that there are at most n closed points on Y'. Denote them by $b_1, \ldots, b_m$, for some $m \leq n$. For each closed point b_j, let $C_j \subset Y'$ be the union of those Y'_i that contain b_j. Then the C_j are closed connected subsets, pairwise disjoint and finite in number. We conclude that the C_j are the connected components of Y'. By construction, each C_j is quasicompact and contains only one closed point, namely b_j, such that C_j is local. It follows that Y' is a sum of local schemes, thus Y is local henselian. □

A scheme Y is called *strictly local*, if it is local henselian, and the residue field of the closed point is separably closed. The class of such schemes is not stable under images of surjective integral morphisms, for example $\mathrm{Spec}(\mathbb{C}) \to \mathrm{Spec}(\mathbb{R})$.

We say that a class $\mathcal{P}$ of schemes is stable under *images of integral surjections with radical residue field extensions* if for every morphism $f : X \to Y$ that is integral, surjective and whose field extensions $\kappa(y) \subset \kappa(x)$, $y = f(x)$, $x \in X$ are radical (that is, algebraic and purely inseparable), and with domain X belonging to $\mathcal{P}$, the domain Y also belongs to $\mathcal{P}$.

Proposition 1.4. *The class of strictly local schemes is stable under images of integral surjections with radical residue field extensions.*

Proof. This follows from Proposition 1.3, together with the fact that a field K is separably closed if it admits a radical extension $K \subset E$ that is separably closed. $\square$

Recall that a ring R is called a pm-*ring*, if every prime ideal $\mathfrak{p} \subset R$ is contained in exactly one maximal ideal $\mathfrak{m} \subset R$, see [De Marco and Orsatti 1971]. Clearly, it suffices to check this for minimal prime ideals, which correspond to the irreducible components of $\operatorname{Spec}(R)$. Therefore, we call a scheme X a pm-*scheme* if it is quasicompact, and each irreducible component is local. These notions will show up in Section 4. For later use, we observe the following fact:

Proposition 1.5. *The class of* pm-*schemes is stable under images of integral surjections.*

Proof. Let $f : X \to Y$ be integral surjective, with X a pm-scheme. Clearly, Y is quasicompact. Let $Y' \subset Y$ be an irreducible component. We have to check that Y' is local. Since f is surjective and closed, there is an irreducible closed subset $X' \subset Y'$ with $f(X') = Y'$. Replacing X and Y be these subschemes, we may assume that X and Y are irreducible. Let $y, y' \in Y$ be two closed points. Choose closed points $x, x' \in X$ mapping to them. Then $x = x'$, because X is pm, whence $y = y'$. $\square$

2. Schemes that are everywhere strictly local

We now introduce a class of schemes that, in my opinion, seems rather natural with respect to the étale topology. Recall that a *strictly local* ring R is a local ring that is henselian and has separably closed residue field $\kappa = R/\mathfrak{m}$.

Definition 2.1. A scheme X is called *everywhere strictly local* if the local rings $\mathcal{O}_{X,x}$ are strictly local, for all $x \in X$.

Similarly, we call a ring R everywhere strictly local if the $R_\mathfrak{p}$ are strictly local for all prime ideals $\mathfrak{p} \subset R$. Note that strictly local rings are usually not everywhere strictly local. The relation between these classes of rings appears to be similar in nature to the relation between valuation rings and Prüfer rings (see, for example, [Endler 1972, §11]). We have the following permanence property:

Proposition 2.2. *If X is everywhere strictly local, then the same holds for each quasifinite X-scheme.*

Proof. Let $U \to X$ be quasifinite, and $u \in U$, with image point $x \in X$. Obviously, the residue field $\kappa(u)$ is separably closed. To check that $\mathbb{O}_{U,u}$ is henselian, we may replace X by the spectrum of $\mathbb{O}_{X,x}$. According to [EGA IV$_4$ 1967, Theorem 18.12.1] there is an étale morphism $X' \to X$, a point $u' \in U' = U \times_X X'$ lying over $u \in U$, and an open neighborhood $u' \in V' \subset U'$ so that the projection $V' \to X'$ is finite. Let $x' \in X'$ be the image of u'. After replacing X' by some affine open neighborhood of x', we may assume that $X' \to X$ is separated. There is a section $s : X \to X'$ with $s(x) = x'$ for the projection $X' \to X$, because $\mathbb{O}_{X,x}$ is strictly local, by [EGA IV$_4$ 1967, Proposition 18.8.1]. Its image is an open connected component, according to [EGA IV$_4$ 1967, Corollary 17.9.4] so shrinking further we may assume that $X' = X$. By [EGA IV$_4$ 1967, Proposition 18.6.8] the local ring $\mathbb{O}_{U,u} = \mathbb{O}_{V',u'}$ is henselian. □

Recall that a morphism $f : Y \to X$ is referred to as *radical* if it is universally injective. Equivalently, the map is injective, and the induced residue field extension $\kappa(x) \subset \kappa(y)$, are purely inseparable [EGA I 1971, Section 3.7] for all $y \in Y$ and $x = f(y)$. The following geometric property will play a crucial role later:

Lemma 2.3. *Let X be an everywhere strictly local scheme and Y be an irreducible scheme. Then every integral morphism $f : Y \to X$ is radical and, in particular, an injective map.*

Proof. Since the residue fields of X are separably closed and f is integral, it suffices to check that f is injective. Let $y, y' \in Y$ with the same image $x = f(y) = f(y')$. Our task is to show $y = y'$. Replace X by the spectrum of the local ring $R = \mathbb{O}_{X,x}$ and Y by the corresponding fiber product. Then $Y = \mathrm{Spec}(A)$ becomes affine, and the morphism of schemes $f : Y \to X$ corresponds to a homomorphism of rings $R \to A$. Write $A = \bigcup A_\iota$ as the filtered union of finite R-subalgebras. Since R is henselian and the A_ι are integral domains, the A_ι are local. Hence A is local too. It follows that the closed points $y, y' \in \mathrm{Spec}(A)$ coincide □

Proposition 2.4. *Let X be everywhere strictly local. Then every étale morphism $f : U \to X$ is a local isomorphism with respect to the Zariski topology.*

Proof. Fix a point $u \in U$. We must find an open neighborhood on which f is an open embedding. Set $x = f(u)$. Consider first the special case that f admits a section $s : X \to U$ through u. Since $U \to X$ is unramified, such a section must be an open embedding by [EGA IV$_4$ 1967, Corollary 17.4.2]. Replacing U by the image of this section, we reduce to the situation that f admits a right inverse s that is an isomorphism. Multiplying $f \circ s = \mathrm{id}_X$ with s^{-1} from the right yields $f = s^{-1}$, which is an isomorphism.

We now come to the general case. Since $\mathbb{O}_{X,x}$ is strictly local, the morphism $U \otimes_X \mathrm{Spec}(\mathbb{O}_{X,x}) \to \mathrm{Spec}(\mathbb{O}_{X,x})$ admits a section through $u \in U$, see [EGA IV$_4$ 1967, Proposition 18.5.11]. According to [EGA IV$_3$ 1966, Theorem 8.8.2] such a section comes from a section defined on some open neighborhood of $x \in X$, and the assertion follows. $\square$

Given a scheme X, we denote by (Et/X) the site whose objects are the étale morphisms $U \to X$, and (Zar/X) the site whose objects are the open subschemes $U \subset X$. In both cases, the covering families $(U_\alpha \to X)_\alpha$ are those where the map $\coprod U_\alpha \to X$ are surjective. In turn, we denote by X_{et} and X_{Zar} the corresponding topoi of sheaves within a fixed universe. The inclusion functor $i : (\mathrm{Zar}/X) \to (\mathrm{Et}/X)$ is cocontinuous, which means that the adjoint i^* on presheaves of the restriction functor i_* preserves the sheaf property. We thus obtain a morphism of topoi $i : X_{\mathrm{et}} \to X_{\mathrm{Zar}}$. The following is a direct consequence of the comparison lemma [SGA 4_1 1972, Exposé III, Theorem 4.1]:

Corollary 2.5. *Let X be everywhere strictly local. Then the canonical morphism $i : X_{\mathrm{et}} \to X_{\mathrm{Zar}}$ of topoi is an equivalence. In particular, sheaves and cohomology groups for the étale site of X is essentially the same as for the Zariski site.*

We are mainly interested in irreducible schemes. Recall that a scheme X is called *unibranch* if it is irreducible, and the normalization map $X'_{\mathrm{red}} \to X_{\mathrm{red}}$ is bijective see ([EGA IV_1 1964, Section 23.2.1]). Equivalently, the henselian local schemes $\mathrm{Spec}(\mathscr{O}^h_{X,x})$ are irreducible, for all $x \in X$, by [EGA IV_4 1967, Corollary 18.8.16]. We have the following characterization, which is close to Artin's original definition [1971].

Proposition 2.6. *Let X be irreducible. Then X is everywhere strictly local if and only if it is unibranch and its function field $k(X) = \kappa(\eta)$ is separably closed.*

Proof. In light of [EGA IV_4 1967, Corollary 18.6.13] the condition is necessary. To see that it is sufficient, we may assume that $X = \mathrm{Spec}(R)$ is a local integral scheme and have to check that R is henselian with separably closed residue field $k = R/\mathfrak{m}_R$.

Let us start with the latter. Seeking a contradiction, we assume that k is not separably closed. Then there is a finite separable field extension $k \subset L$ of degree $d \geq 2$. By the primitive element theorem, we have $L = k[T]/(f)$ for some irreducible separable monic polynomial $f \in k[T]$. Choose a monic polynomial $F(T)$ with coefficients in R reducing to $f(T)$. Then $\partial F/\partial T \in R$ is a unit, and it follows from [Milne 1980, Chapter I, Corollary 3.6] that the finite R-algebra $A = R[T]/(F)$ is étale. Since the field of fractions $\Omega = \mathrm{Frac}(R)$ is separably closed, we have a decomposition $A \otimes_R \Omega = \prod_{i=1}^d \Omega$. Set $S = \mathrm{Spec}(A)$, and fix one generic point $\eta_0 \in S$. Note that its residue field must be Ω. Its closure $S_0 \subset S$ is a local scheme, finite over R, and thus with residue field L.

Now choose a separable closure $k \subset k'$, consider the resulting strict henselization $R' = R^s$, and write $S'_0 = S_0 \otimes_R R'$ and $S' = S \otimes_R R'$ for the ensuing faithfully flat base-change. According to [SGA 1 1971, Exposé VIII, Theorem 4.1] the preimage

$S'_0 \subset S'$ is the closure of the point $\eta'_0 \in S'$ lying over $\eta_0 \in S$, and thus S'_0 is irreducible, in particular connected. The closed fiber $S'_0 \otimes_{R'} k' = \operatorname{Spec}(L) \otimes_k k'$ is a disjoint union of $d \geq 2$ closed points, which lie in the same connected component inside S'. But, since R' is henselian, [EGA IV$_4$ 1967, Theorem 18.5.11 (a)] ensures that no two points in the closed fiber $S' \otimes_{R'} k'$ lie in the same connected component of S', contradiction. Summing up, the residue field $k = R/\mathfrak{m}_R$ is separably closed.

It remains to see that R is henselian. Let $F \in R[T]$ be a monic polynomial. In light of [EGA IV$_4$ 1967, Theorem 18.5.11 (a′)], it suffices to see that the finite flat R-algebra $A = R[T]/(F)$ is a product of local rings. Let $d = \deg(F) = \deg(A)$ be its degree, and consider the R-scheme $Y = \operatorname{Spec}(A)$. Let $Y_1, \ldots, Y_r \subset Y$ be the closures of the connected components of the generic fiber $Y \otimes \Omega$. Using that R is unibranch, we infer that the closed fibers $Y_i \otimes k$ are local. By the going-down theorem for the integral ring extension $R \subset A$, every point $y \in Y \otimes k$ lies in some Y_i. It follows that $Y = \bigcup Y_i$. For each $y \in Y_k$, let $C_y \subset Y$ by the union of all those Y_i containing y. Clearly, the C_y are connected and local. Since the closed fibers of Y_i are local, the C_y, $y \in Y_k$, are pairwise disjoint. In turn, the C_y are the connected components of $Y = \operatorname{Spec}(A)$. It follows that A is a product of local rings. □

3. Total separable closure

Let X be an integral scheme. We say that X is *totally separably closed* if it is normal, and the function field $k(X)$ is separably closed. According to Proposition 2.6, these are precisely the integral normal schemes that are everywhere strictly local. From this we deduce:

Proposition 3.1. *If X is totally separably closed, so is every normal closed subscheme $X' \subset X$.*

Now let X_0 be an integral scheme, and choose a separable closure F^s of the function field $F = k(X_0)$. We define the *total separable closure*

$$X = \operatorname{TSC}(X_0)$$

to be the integral closure of X_0 inside F^s. We may regard it as the filtered inverse limit: Let $F \subset F_\lambda \subset F^{\text{alg}}$, $\lambda \in L$, be the set of intermediate fields that are finite over $F = k(X_0)$, ordered by the inclusion relation, and let $X_\lambda \to X_0$ be the corresponding integral closures. These form a filtered inverse system of schemes with finite surjective transition maps, and we get a canonical identification

$$X \to \varprojlim X_\lambda,$$

We tacitly assume that the smallest element in the index set L is denoted by $\lambda = 0$. Note that the fibers of the map $X \to X_0$, viewed as a topological space, are *profinite*.

Recall that such spaces are precisely totally disconnected compacta, which are also called *Stone spaces*.

Let me introduce the following notation as a general convention for this paper: Suppose that $C \subset X$ is a closed subscheme. Then the schematic images $C_\lambda \subset X_\lambda$ are closed, and the underlying set is just the image set. These form a filtered inverse system of schemes, again with affine transition maps by [EGA II 1961, Proposition 1.6.2]. According to [Bourbaki 1971, Chapter I, §4, No. 4, Corollary to Proposition 9] we get a canonical identification $C = \varprojlim C_\lambda$ as sets, and it is easy to see that this is an equality of schemes. The fiber products $X \times_{X_\lambda} C_\lambda \subset X$, $\lambda \in L$, are closed subschemes and each one contains C as a closed subscheme, but is usually much larger. In fact, one has $C = \bigcap_{\lambda \in L}(X \times_{X_\lambda} C_\lambda)$ as closed subschemes inside X. Now if C is integral, then its function field $k(C)$ is separably closed. This yields:

Proposition 3.2. *If the closed subscheme $C \subset X$ is normal, then $C = \mathrm{TSC}(C_\lambda)$ for each index $\lambda \in L$.*

Now suppose that we have a ground field k. Let X_0 be an integral k-scheme and $X = \mathrm{TSC}(X_0)$ its total separable closure, as above. The following observation reduces the situation to the case that the ground field is separably closed: Let k' be the relative algebraic closure of k inside F^s, where $F = k(X_0)$. The scheme $X \otimes_k k'$ is not necessarily integral, but the schematic image $X' \subset X \otimes_k k'$ of the canonical morphism $X \to X \otimes_k k'$ is. Note that if the k-scheme X_0 is algebraic, quasiprojective, or proper, the respective properties hold for the k'-scheme X'. Clearly, the morphism $X \to X'$ is integral and dominant. Regarding $k(X)$ as an separable closure of $k(X')$, we get a canonical morphism $X \to \mathrm{TSC}(X')$.

Proposition 3.3. *The canonical morphism $X \to \mathrm{TSC}(X')$ is an isomorphism.*

Proof. The scheme X is integral, the morphism in question is integral and birational, and the scheme $\mathrm{TSC}(X')$ is normal, and the result follows. □

Finally, suppose that X_0 is an integral *algebraic space* rather than a scheme. Then one may define its total separable closure $X = \mathrm{TSC}(X_0) = \varprojlim X_\lambda$ in the analogous way. But here nothing interesting happens: indeed, then some X_λ is a scheme ([Laumon and Moret-Bailly 2000, Corollary 16.6.2], when X_0 is noetherian, and [Rydh 2010, Theorem B], for X_0 quasicompact and quasiseparated), such that $\mathrm{TSC}(X_0)$ is a scheme.

4. Acyclic schemes

In this section we study acyclicity for quasicompact schemes with respect to the Zariski topology, the Nisnevich topology, and the étale topology. We thus take up the question in [SGA 4_2 1972, Exposé V, Problem 4.14] to study topoi for which every abelian sheaf has trivial higher cohomology.

First recall that a topological space is called *at most zero-dimensional* if its topology admits a basis consisting of subsets that are open-and-closed. Note that this is in the sense of dimension theory in general topology (see, for example, [Pears 1975]), rather than dimension theory in commutative algebra and algebraic geometry.

Given a ring R, we write $\operatorname{Max}(R) \subset \operatorname{Spec}(R)$ for the subspace of points corresponding to maximal ideals. Similarly, we write $\operatorname{Max}(X) \subset X$ for the set of closed points of a scheme X, endowed with the subspace topology.

Theorem 4.1. *Let X be a quasicompact scheme. Then the following are equivalent*:

(i) *We have $H^p(X, F) = 0$ for every abelian sheaf F and every $p \geq 1$, where cohomology is taken with respect to the Zariski topology.*

(ii) *Every surjective local isomorphism $U \to X$ admits a section.*

(iii) *The scheme X is affine, and each connected component is local.*

(iv) *The scheme X is affine, each irreducible component is local, and the space of closed points* $\operatorname{Max}(X)$ *is at most zero-dimensional.*

(v) *The scheme X is affine, and every element in $R = \Gamma(X, \mathscr{O}_X)$ is the sum of an idempotent and a unit.*

Proof. (i)$\Rightarrow$(ii) Seeking a contradiction, suppose that some surjective local isomorphism $U \to X$ does not admit a section. Since X is quasicompact, there are finitely many affine open subsets $U_1, \dots, U_n \subset U$ so that each $U_i \to X$ is an open embedding, and $\coprod U_i \to X$ is surjective. Replace U by the direct sum $\coprod U_i$. Let F be the product of the extension-by-zero sheaves $(f_i)_!(\mathbb{Z}_{U_i})$, where $f_i : U_i \to X$ are the inclusion maps. Then $\Gamma(X, F) = 0$ for all $1 \leq i \leq n$. Consider the Čech complex

$$\Gamma(X, F) \to \prod_{i=1}^{n} \Gamma(U_i, F) \to \prod_{i,j=1}^{n} \Gamma(U_i \cap U_j, F).$$

The constant section (1_{U_i}) in the middle is a cocycle, but not a coboundary, and this holds true for all refinements of the open covering $X = \bigcup U_i$. In turn, we have $\check{H}^1(X, F) \neq 0$. On the other hand, the canonical map $\check{H}^p(X, F) \to H^p(X, F)$ from Čech cohomology to sheaf cohomology is injective and actually bijective for $p = 1$, whence $H^1(X, F) \neq 0$, contradiction.

(ii)$\Rightarrow$(i) Let F be an abelian sheaf. Every F-torsor becomes trivial on some $U \to X$ as above. Since the latter has a section, the torsor is already trivial on X. It follows $H^1(X, F) = 0$. In turn, the global section functor $F \mapsto \Gamma(X, F)$ is exact, hence $H^p(X, F) = 0$ for all F and $p \geq 1$.

(ii)$\Rightarrow$(iii) Choose an affine open covering $X = U_1 \cup \dots \cup U_n$. Using that $\coprod U_i \to X$ has a section, we infer that X is quasiaffine, in particular separated. By the previous implication, $H^1(X, \mathscr{F}) = 0$ for each quasicoherent sheaf. According to Serre's criterion [EGA II 1961, Theorem 5.2.1] the scheme X is affine. Next, let $C \subset X$ be a connected component, and suppose C is not local. Choose two different closed points $a_1 \neq a_2$ in C, and let U be the sum of $U_1 = X \setminus \{a_2\}$ and $U_2 = X \setminus \{a_1\}$. The ensuing local isomorphism $U \to X$ allows a section $s : X \to U$. Then $s(C) \subset U$ is connected, and intersects both U_1 and U_2. In turn, $s(C) \subset U_i$ for $i = 1, 2$. But U_1 and U_2 have empty intersection when regarded as subsets of U, contradiction.

(iii)⇒(iv) Let $A \subset X$ be an irreducible component, and $C \subset X$ be its connected component. Then A is local, because the subset $A \subset C$ is closed and nonempty. To see that $\mathrm{Max}(X)$ is at most zero-dimensional, write $X = \mathrm{Spec}(R)$. Let $x \in X$ be a closed point, $x \in U \subset X$ an open neighborhood of the form $U = \mathrm{Spec}(R_f)$, and $C \subset X$ be the connected component of x. Let $S \subset R$ be the multiplicative system of all idempotents $e \in R$ that are units on C. Then $C = \mathrm{Spec}(S^{-1}R)$, and the localization map $R \to S^{-1}R$ factors over R_f. In turn, there is some $e \in S$ so that the localization map $R \to R_e$ factors over R_f. In other words, there is an open-and-closed neighborhood $x \in V$ contained in U. It follows that the subspace $\mathrm{Max}(X) \subset X$ is at most zero-dimensional.

(iv)⇔(v) This equivalence is due to [Johnstone 1977, Chapter V, Proposition in 3.9].

(v)⇒(iii) Write $X = \mathrm{Spec}(R)$, and let $C = \mathrm{Spec}(A)$ be a connected component, $A = R/\mathfrak{a}$. Then every nonunit $f \in A$ is of the form $f = 1 + u$ for some unit $u \in A^\times$. In turn, the subset $A \setminus A^\times \subset A$ comprises an ideal, because it coincides with the Jacobson radical

$$J = \{f \mid fg - 1 \in A^\times \text{ for all } g \in A\}.$$

Thus A is local.

(iii)⇒(ii) Let $f : U \to X$ be a surjective local isomorphism. To produce a section, we may assume that U is a finite sum of affine open subschemes of X, in particular of finite presentation over X. Let $x \in X$ be a closed point. Obviously, there is a section after base-changing to $C = \mathrm{Spec}(\mathscr{O}_{X,x})$. Since this is a connected component, we may write $C = \bigcap V_\lambda$, where the V_λ are the open-and-closed neighborhoods of x, see Appendix A. According to [EGA IV$_3$ 1966, Theorem 8.8.2] a section already exists over some V_λ. Using quasicompactness of X, we infer that there is an open-and-closed covering $X = X_1 \cup \cdots \cup X_n$ so that sections exists over each X_i. By induction on $n \geq 1$, one easily infers that a section exists over X. □

Let us call a scheme X *acyclic with respect to the Zariski topology* if it is quasicompact and satisfies the equivalent conditions of Theorem 4.1. Note that some conditions in Theorem 4.1 already occurred in various other contexts:

Rings satisfying condition (v), that is, every element is a sum of an idempotent and a unit are also known as *clean rings*. Such rings have been extensively studied in the realm of commutative algebra (we refer to [Nicholson and Zhou 2005] for an overview).

Rings for which every prime ideal is contained in only one maximal ideal, one of the conditions occurring in (iv), are called pm-*rings* or *Gelfand rings* (they were introduced in [De Marco and Orsatti 1971]).

Also note that in condition (iv) one cannot remove the assumption that $\mathrm{Max}(X)$ is at most zero-dimensional. In fact, if K is an arbitrary compact topological space, then the ring $R = \mathscr{C}(K)$ is pm, such that its spectrum has local irreducible components; on the other hand, $\mathrm{Max}(X) = K$ (see [Gillman and Jerison 1960], Section 4 for the latter, and Theorem 2.11 on p. 29 for the former).

We now turn to the *Nisnevich topology* on the category (Et/X). Recall that a morphism $U \to V$ of schemes is called *completely decomposed* if, for each $v \in V$, there is a point $u \in U$ with $f(u) = v$ and $\kappa(v) = \kappa(u)$. The covering families $(U_\alpha \to U)$ for the Nisnevich topology are those families for which each $U_\alpha \to U$ is completely decomposed, and $\bigcup U_\alpha \to U$ is surjective. Sheaves on the ensuing site are referred to as *Nisnevich sheaves*, and we write $H^p(X_{\mathrm{Nis}}, F)$ for their cohomology groups, see [Nisnevich 1989]. Note that each point $x \in X$ yields a point in the sense of topos-theory, and that the corresponding local ring is the henselization $\mathscr{O}^h_{X,x}$.

Theorem 4.2. *Let X be a quasicompact scheme. Then the following are equivalent*:

(i) $H^p(X_{\mathrm{Nis}}, F) = 0$ *for every abelian Nisnevich sheaf F and every $p \geq 1$.*

(ii) *Every completely decomposed étale surjection $U \to X$ admits a section.*

(iii) *The scheme X is affine, and each connected component is local henselian.*

(iv) *The scheme X is affine, each irreducible component is local henselian, and the space* $\mathrm{Max}(X)$ *is at most zero-dimensional.*

The arguments are parallel to the ones for Theorem 4.1, and left to the reader. For the implication (iv)⇒(iii), one needs the following:

Lemma 4.3. *Let Y be a local scheme. Then Y is henselian if and only if each irreducible component $Y_i \subset Y$, $i \in I$, is henselian.*

Proof. The condition is necessary, because $X_i \subset X$ are closed subsets. Conversely, suppose that each X_i is henselian. Let $f : X \to Y$ be a finite morphism, and let $a_1, \dots, a_n \in X$ be the closed points. Each of them maps to the closed point $b \in Y$. The closed subsets $X_i = f^{-1}(Y_i)$ contain $a_1, \dots, a_n$ and are finite over X_i, whence the canonical map $\coprod_{j=1}^n \mathrm{Spec}(\mathscr{O}_{X_i,a_j}) \to X_i$ is an isomorphism. For each $1 \leq j \leq n$, consider the subset

$$C_j = \bigcup_{i \in I} \mathrm{Spec}(\mathscr{O}_{X_i,a_j}) \subset X.$$

Then the $C_j \subset X$ are stable under generization and contain a single closed point a_j, thus $C_j = \mathrm{Spec}(\mathscr{O}_{X,a_j})$. Furthermore, the C_j form a partition of X. It is not a priori clear that $C_j \subset X$ is closed if the index set I is infinite. But this nevertheless holds: Write $X = \mathrm{Spec}(A)$, where A is a semilocal ring. Let $\mathfrak{m}_j \subset A$ be the maximal ideal corresponding to $a_j \in X$. Since the C_j are pairwise disjoint, the ideals $\mathfrak{m}_j$ are pairwise coprime. By the Chinese reminder theorem, the canonical map $A \to \prod_{j=1}^n A_{\mathfrak{m}_j}$ is bijective. Hence $C_j \subset X$ are closed, and X is the sum of local schemes. □

We finally come to the étale topology. This is the Grothendieck topology on the category (Et/X) whose covering families $(U_\alpha \to U)$ are those families with $\bigcup U_\alpha \to U$ surjective. Sheaves on this site are called *étale sheaves*, and we write $H^p(X_{\mathrm{et}}, F)$ for their cohomology groups. In the following, the equivalence (ii)⇔(iii) is due to Artin [1971]:

Theorem 4.4. *Let X be a quasicompact scheme. Then the following are equivalent*:

(i) *We have* $H^p(X_{\mathrm{et}}, F) = 0$ *for every abelian étale sheaf* F *and every* $p \geq 1$.

(ii) *Every étale surjective* $U \to X$ *admits a section.*

(iii) *The scheme* X *is affine*, *and each connected component is strictly local.*

(iv) *The scheme* X *is acyclic*, *each irreducible component is strictly local*, *and the space* $\mathrm{Max}(X)$ *is at most zero-dimensional.*

The remaining arguments are parallel to the ones for Theorem 4.1, and left to the reader.

I conjecture that the class of schemes that are acyclic with respect to either the Zariski or the Nisnevich topology are stable under images of integral surjections. Indeed, if $f : X \to Y$ is integral and surjective, with X acyclic, then Y is affine by Theorem 1.1, and it follows from Propositions 1.2 and 1.3 that each irreducible component is local or local henselian, respectively. It only remains to check that $\mathrm{Max}(Y)$ is at most zero-dimensional, and here lies the problem: We have a closed continuous surjection $\mathrm{Max}(X) \to \mathrm{Max}(Y)$, but cannot conclude from this that the image is at most zero-dimensional, in light of the existence of dimension-raising maps (see [Pears 1975, Chapter 7, Theorem 1.8]). The following weaker statement will suffice for our applications:

Proposition 4.5. *Let* Y *be a scheme*, *and* $Y = Y_1 \cup \cdots \cup Y_n$ *be a closed covering. If each* Y_i *is acyclic with respect to the Zariski topology*, *so is* Y.

Proof. As discussed above, it remains to show that the space of closed points $\mathrm{Max}(Y)$ is at most zero-dimensional. The canonical morphism $Y_1 \amalg \cdots \amalg Y_n \to Y$ is surjective and integral, whence Y is a pm-scheme by Proposition 1.5. Note that this ensures that the topological space Y is *normal*, that is, disjoint closed subsets can be separated by disjoint open neighborhoods, according to [Carral and Coste 1983, Proposition 2]. Furthermore, the subspace $\mathrm{Max}(Y)$ is *compact*, which means quasicompact and

Hausdorff [Schwartz 2011, Corollary 4.7 together with Proposition 4.1]. Since our covering is closed, we have $\mathrm{Max}(Y) = \mathrm{Max}(Y_1) \cup \cdots \cup \mathrm{Max}(Y_n)$, and this is again a closed covering.

To proceed, we now use the notion of small and large inductive dimension from general topology. Since this material is perhaps not so well-known outside general topology, I have included a brief discussion in Appendix C. The relevant results are tricky, but rely only on basic notations of topology. Proposition C.3, which is a consequence of a general sum theorem on the large inductive dimension, ensures that the compact space $\mathrm{Max}(Y) = \mathrm{Max}(Y_1) \cup \cdots \cup \mathrm{Max}(Y_n)$ is at most zero-dimensional. $\square$

5. The Cartan–Artin properties

Let X be a topological space and F be an abelian sheaf. Then there is a spectral sequence

$$E_2^{pq} = \check{H}^p(X, \underline{H}^q(F)) \Longrightarrow H^{p+q}(X, F)$$

computing sheaf cohomology in terms of Čech cohomology, where $\underline{H}^q(F)$ denotes the presheaf $U \mapsto H^p(U, F)$. A result of H. Cartan using this spectral sequence tells us that Čech and sheaf cohomology coincide on spaces admitting a basis $\mathscr{B}$ for the topology that is stable under finite intersections and satisfies $H^p(U, F) = 0$ for all $U \in \mathscr{B}$, $p \geq 1$, and F.

The same holds, cum grano salis, for arbitrary sites (see the discussion in Appendix B). For the étale site of a scheme X, there seems to be no candidate for such a basis $\mathscr{B}$ of open sets. However, there is the following substitute: Let $a = (a_1, \ldots, a_n)$ be a sequence of geometric points $a_i : \mathrm{Spec}(\Omega_i) \to X$. Following Artin, we define

$$X_a = X_{a_1,\ldots,a_n} = \mathrm{Spec}(\mathscr{O}^s_{X,a_1}) \times_X \cdots \times_X \mathrm{Spec}(\mathscr{O}^s_{X,a_n}).$$

The schemes X_a can be viewed as inverse limits of finite intersections of smaller and smaller neighborhoods, and therefore appears to be a good substitute for the $U \in \mathscr{B}$. It was M. Artin's insight [1971] that for the collapsing of the Cartan–Leray spectral sequence in the étale topology it indeed suffices to verify that the inverse limits X_a are acyclic.

In order to treat the Nisnevich topology, we use the following: If $x = (x_1, \ldots, x_n)$ is a sequence of ordinary points $x_i \in X$, we likewise set

$$X_x = X_{x_1,\ldots,x_n} = \mathrm{Spec}(\mathscr{O}^h_{X,x_1}) \times_X \cdots \times_X \mathrm{Spec}(\mathscr{O}^h_{X,x_n}),$$

using the henselizations rather then the strict localizations. Usually, one writes $\bar{x}_i : \mathrm{Spec}(\Omega_i) \to X$ for geometric points with image point $x_i \in X$. In order to keep notation simple, and since we are mainly interested in geometric points, I prefer to write a_i for geometric points. Although this is slightly ambiguous, it should cause no confusion. Let us now introduce the following terminology:

Definition 5.1. We say that the scheme X has the *weak Cartan–Artin property* if for each sequence $x = (x_1, \dots, x_n)$ of points on X, the scheme X_x is acyclic with respect to the Nisnevich topology. We say that X has the *strong Cartan–Artin property* if for each sequence $a = (a_1, \dots, a_n)$ of geometric points on X, the scheme X_a is acyclic with respect to the étale topology.

In the latter case, this means that the schemes X_a are affine, their irreducible components are strictly local, and the subspace $\operatorname{Max}(X_a) \subset X_a$ of closed points is at most zero-dimensional. In the former case, however, there is no condition on the residue fields of the closed points. Our interest in this property comes from the following result, which was proved by Artin [1971] for the strong Cartan–Artin property. For the weak Cartan–Artin property, the arguments are virtually the same.

Theorem 5.2 (M. Artin). *Let X be a noetherian scheme. If it has the strong Cartan–Artin property, then $\check{H}^p(X_{\text{et}}, \underline{H}^q(F)) = 0$ for all abelian étale sheaves F and all $p \geq 0$, $q \geq 1$. If it has the weak Cartan–Artin property, then $\check{H}^p(X_{\text{Nis}}, \underline{H}^q(F)) = 0$ for all abelian Nisnevich sheaves F and all $p \geq 0$, $q \geq 1$. In particular, Čech cohomology coincides with sheaf cohomology.*

To avoid tedious repetitions, we will work in both cases with the schemes X_a constructed with the strict localizations. This is permissible, in light of the following observation:

Proposition 5.3. *The scheme X has the weak Cartan–Artin property if and only if for each sequence $a = (a_1, \dots, a_n)$ of geometric points on X, the scheme X_a is acyclic with respect to the Nisnevich topology.*

Proof. Let $x_i \in X$ be the image of the geometric point a_i, and set $x = (x_1, \dots, x_n)$. Consider the canonical morphism $X_a \to X_x$. The rings $\mathcal{O}^s_{X,a}$ are filtered direct limits of finite étale $\mathcal{O}^h_{X,x}$-algebras, whence the scheme X_a is a filtered inverse limit $X_a = \varprojlim V_\lambda$ of X_x-schemes whose structure morphism $V_\lambda \to X_x$ are étale coverings. In particular, the morphism $f : X_a \to X_x$ and the projections $X_a \to V_\lambda$ are a flat integral surjections.

We now make the following observation: Suppose $B \subset X_a$ and $C \subset X_x$ are connected components, with $f(B) \subset C$. We claim that the induced map $f : B \to C$ is surjective. To see this, let $B_\lambda \subset V_\lambda$ be the connected components containing the images of B. These B_λ form a filtered inverse system of connected affine schemes with $B = \varprojlim B_\lambda$, and the structure maps $B_\lambda \to X_x$ factor over C. Since C is connected and the induced projections $V_\lambda \times_{X_x} C \to C$ are étale coverings, the preimages $V_\lambda \times_{X_x} C$ are sums of finitely many connected components, each being an étale covering of C. This follows, for example, from the fact that the category of finite Galois coverings of a connected scheme is a Galois category admitting a fiber functor. This was established for locally noetherian schemes in [SGA 1 1971, Exposé V]; for the general case see [Lenstra 1985, Theorem 5.24]. One of the connected components must be B_λ, and we conclude that the $B_\lambda \to C$ are surjective. Likewise, the transition maps $B_\lambda \to B_\mu$, $\lambda \geq \mu$ must be surjective, and we infer that $B = \varprojlim B_\lambda \to C$ is surjective.

Now suppose that X_a is acyclic with respect to the Nisnevich topology. Then it is affine by Theorem 4.2. It follows from Theorem 1.1 that X_x is affine. Let $C \subset X_x$ be a connected component. We have to show that C is local. Since $f : X_a \to X_x$ is surjective, there is a connected component $B \subset X_a$ with $f(B) \subset C$, and we saw in the preceding paragraph that $f : B \to C$ is an integral surjection. Since X_a is acyclic, the scheme B is local henselian. The same then holds for C, by Proposition 1.3.

Conversely, suppose that X_x is acyclic. Then X_x is affine, and the same holds for X_a, because f is an affine morphism. Let $B \subset X_a$ be a connected component, and $C \subset X_x$ the connected component containing $f(B)$. Then B is local, and we saw above that $f : B \to C$ is an integral surjection. Moreover, the B_λ are local, and the transition maps are local, whence $B = \varprojlim B_\lambda$ is local. □

6. Reduction to normal schemes

The goal of this section is to reduce checking the Cartan–Artin properties to the case of irreducible, or even normal schemes. Artin [1971] bypassed this by assuming the AF-property, together with the fact that any affine scheme is a subscheme of some integral scheme. Without assuming the AF-property, or knowing that any scheme is a subscheme of some integral scheme, a different approach is necessary.

Let X be a scheme. We now take a closer look at the functoriality of the X_a with respect to X. Let $f : X \to Y$ be a morphism of schemes, and $a = (a_1, \dots, a_n)$ a sequence of geometric points on X, and $b = (b_1, \dots, b_n)$ the sequence of geometric image points on Y. We then obtain an induced map between strictly local rings, and thus a morphism $X_a \to Y_b$.

Lemma 6.1. *Let $\mathcal{P}$ be a class of morphisms that is stable under fiber products, and that contains all closed embeddings. If $f : X \to Y$ is separated, and the induced morphisms $\operatorname{Spec}(\mathcal{O}^s_{X,a_i}) \to \operatorname{Spec}(\mathcal{O}^s_{Y,b_i})$ belong to $\mathcal{P}$, then $X_a \to Y_b$ belongs to $\mathcal{P}$.*

Proof. By assumption, the induced morphism

$$\prod_{i=1}^n \operatorname{Spec}(\mathcal{O}^s_{X,a_i}) \to \prod_{i=1}^n \operatorname{Spec}(\mathcal{O}^s_{Y,b_i})$$

belongs to $\mathcal{P}$, where the products designate fiber products over Y. It remains to check the following: If U, V are two X-schemes, then the morphism $U \times_X V \to U \times_Y V$

is a closed embedding. This indeed holds by [EGA I 1971, Proposition 5.2], because $f : X \to Y$ is separated. $\square$

Proposition 6.2. *With the notation above, if $f : X \to Y$ is integral then the same holds for $X_a \to Y_b$. Moreover, if $f : X \to Y$ is a finite, closed embedding, radical or with radical field extensions, the respective property holds for $X_a \to Y_b$.*

Proof. In light of Lemma 6.1, it suffices to treat the case that the sequence $a = (a_1, \ldots, a_n)$ consists of a single geometric point, which by abuse of notation we also denote by a. Write $X = \varprojlim X_\lambda$ as a filtered inverse limit of finite Y-schemes X_λ, and let a_λ be the geometric point on X_λ induced by a. Let $x \in X$ and $x_\lambda \in X_\lambda$ be their respective images. Then $\mathscr{O}_{X,x} = \varinjlim \mathscr{O}_{X_\lambda, x_\lambda}$, and it follows from [EGA IV$_4$ 1967, Corollary 18.6.14] that $\mathscr{O}^s_{X,a} = \varinjlim \mathscr{O}^s_{X_\lambda, a_\lambda}$. On the other hand, the base change $X_\lambda \times_Y \operatorname{Spec}(\mathscr{O}^s_{Y,b})$ is a finite sum of strictly local schemes. Let C_λ be the connected component corresponding to $a'_\lambda = (a_\lambda, b)$. It follows from [EGA IV$_4$ 1967, Proposition 18.6.8] that $\mathscr{O}^s_{X_\lambda, a_\lambda} = \mathscr{O}_{C_\lambda, a'_\lambda}$. In turn, $\mathscr{O}^s_{X,a}$ is a filtered inverse limit of finite $\mathscr{O}^s_{Y,b}$-algebras, thus integral. From this description, the additional assertions follow as well. $\square$

Our goal now is to reduce checking the Cartan–Artin properties to the case of integral normal schemes. Recall that $a = (a_1, \ldots, a_n)$ is a fixed sequence of geometric points on X.

Proposition 6.3. *Suppose there are only finitely many irreducible components $X_j \subset X$, $1 \leq j \leq r$ that contain all geometric points $a_1, \ldots, a_n$. Then the closed subschemes*

$$X_{j,a} = (X_j)_a = \prod_{i=1}^{n} \operatorname{Spec}(\mathscr{O}^s_{X_j, a_i}), \quad 1 \leq j \leq r,$$

form a closed covering of X_a. Here the product means fiber product over X.

Proof. Let $x_i \in X$ be the images of the geometric points a_i. Clearly, the structure morphism $X_a \to X$ factors through $\bigcap_{i=1}^n \operatorname{Spec}(\mathscr{O}_{X,x_i})$, and this intersection is the set of points $x \in X$ containing all $x_1, \ldots, x_n$ in their closure. Thus the inclusion $\bigcap_{i=1}^n \operatorname{Spec}(\mathscr{O}_{X',x_i}) \subset \bigcap_{i=1}^n \operatorname{Spec}(\mathscr{O}_{X,x_i})$ is an equality, where $X' = X_1 \cup \cdots \cup X_r$. It then follows that the closed embedding $X'_a \subset X_a$ is bijective. We thus may assume that $X = X'$.

Let $\eta \in X_a$ be a generic point. The structure morphism $X_a \to X$ is flat. By going-down, the point η maps to the generic point $\eta_j \in X_j$ for some $1 \leq j \leq r$. In turn, it is contained in $(X_j)_a = X_a \times_X X_j$. Consequently the $(X_j)_a \subset X$ form a covering, because they are contain every generic point and are closed subsets. $\square$

Proposition 6.4. *Suppose the irreducible components of X are locally finite, and that the diagonal $X \to X \times X$ is an affine morphism. Then X has the weak/strong*

Cartan–Artin property if and only if the respective property holds for each irreducible component, regarded as an integral scheme.

Proof. Let $a = (a_1, \dots, a_n)$ be a sequence of geometric points on X, and denote by $X_1, \dots, X_r \subset X$ the irreducible components over which all a_i factors. According to Proposition 6.3, we have a closed covering $X_a = (X_1)_a \cup \dots \cup (X_r)_a$. Now Proposition 4.5 ensures that the space of closed points $\mathrm{Max}(X_a)$ is at most zero-dimensional. Let $S = (X_1)_a \amalg \dots \amalg (X_r)_a$ be the sum. Then $S \to X_a$ is integral and surjective. By the assumption on the diagonal morphism of X, the scheme S is affine, and the scheme X_a must be affine as well, by Theorem 1.1. Finally, let $C \subset X_a$ be an irreducible component. Choose some $1 \leq j \leq r$ with $C \subset (X_j)_a$. Then C is an irreducible component of $(X_j)_a$. Since $(X_j)_a$ is acyclic, the scheme C is strictly local, according to Theorems 4.2 and 4.4, and it follows that X_a is acyclic. □

Concerning the Cartan–Artin properties, we now may restrict our attention to integral schemes. Our next task is to reduce to normal schemes. Suppose that $f : X \to Y$ is a finite birational morphism between integral schemes, and let $b = (b_1, \dots, b_n)$ be a sequence of geometric points in Y. As a shorthand, we write

$$f^{-1}(b) = f^{-1}(b_1) \times \dots \times f^{-1}(b_n)$$

for the finite set of sequences $a = (a_1, \dots, a_n)$ of geometric points with $f(a_i) = b_i$. Note that one may regard $f^{-1}(b)$ as the product of the points in the finite schemes $X \times_Y \operatorname{Spec} \kappa(b_i)$. For each $a \in f^{-1}(b)$, we obtain an induced morphism $X_a \to Y_b$, which is finite. Its image is denoted, for the sake of simplicity, by $f(X_a) \subset Y_b$, which is a closed subset. Now recall that an integral scheme X is called *geometrically unibranch* if the normalization map $X' \to X$ is radical. This obviously holds if X is already normal.

Proposition 6.5. *With the assumptions above, suppose that X is geometrically unibranch. Then the finite morphisms $X_a \to Y_b$ are radical, and the closed subsets*

$$f(X_a) \subset Y_b, \quad a \in f^{-1}(b),$$

form a closed covering of Y_b.

Proof. First, let us check that $X_a \to Y_b$ is radical. In light of Proposition 6.2, it suffices to treat the case that the sequence $a = (a_1, \dots, a_n)$ consists of a single geometric point. Let $x \in X$ and $y \in Y$ be the points corresponding to the geometric points $a = a_1$ and $b = b_1$, respectively. Set $R = \mathscr{O}_{Y,y}$ and write $X \times_Y \operatorname{Spec}(\mathscr{O}_{Y,y}) = \operatorname{Spec}(A)$. Then A is a semilocal, integral, and finite R-algebra. According to [EGA IV$_4$ 1967, Proposition 18.6.8] we have $A \otimes_R R^s = A^s$. This is a finite R^s-algebra, whence splits into $A^s = B_1 \times \dots \times B_m$, where the factors are local and correspond to the geometric points in $f^{-1}(y)$. In fact, the factors are the strict localizations of X at the these geometric points, and $\mathscr{O}_{X,a}$ is one of them. Since X is geometrically unibranch, all factors B_i are integral, according to [EGA IV$_4$ 1967, Proposition 18.6.12].

As $f : X \to Y$ is birational, the inclusion $R \subset A$ becomes bijective after localizing with respect to the multiplicative system $S = R \setminus 0$. In turn, the inclusion on the left

$$S^{-1}R^s \subset S^{-1}A^s = S^{-1}(A \otimes_R R^s) = S^{-1}A \otimes_{S^{-1}R} S^{-1}R^s = S^{-1}R^s$$

is bijective, such that we have a canonical identification $S^{-1}A^s = S^{-1}R^s$. Let $R_i \subset B_i$ be the image of the composite map $R^s \to B_i$, which is also a residue class ring of R^s. Hence the $R_i \simeq R^s$ are integral strictly local rings, in particular geometrically unibranch, and $R_i \subset B_i$ induces a bijection on fields of fractions. It follows that

$$\operatorname{Spec}(B_i) \to \operatorname{Spec}(R_i) \quad \text{and} \quad \operatorname{Spec}(\mathscr{O}^s_{X,a}) \to \operatorname{Spec}(\mathscr{O}^s_{Y,b})$$

are radical.

It remains to check that $Y_b = \bigcup_a f(X_a)$. Let $\eta \in Y_b$ be a generic point. The images $\eta_i \in \operatorname{Spec}(\mathscr{O}_{Y,b_i})$ are generic points as well, because the projections $Y_b \to \operatorname{Spec}(\mathscr{O}^s_{Y,b_i})$ are flat, whence satisfy going-down. We saw in the preceding paragraph that the generic points in $\operatorname{Spec}(\mathscr{O}^s_{Y,b_i})$ correspond to the points in $f^{-1}(b_i)$. Let $a_i \in f^{-1}(b_i)$ be the point corresponding to η_i, and form the sequence $a = (a_1, \dots, a_n)$. Using that the $f : X \to Y$ is birational, we infer that η lies in the image of $X_a \to Y_b$. □

If Y is an integral scheme, its normalization map $f : X \to Y$ is integral, but not necessarily finite. However, the latter holds for the so-called *Japanese schemes*, see [EGA IV$_1$ 1964, Chapter 0, §23].

Proposition 6.6. *Suppose that Y is an integral scheme whose normalization morphism $f : X \to Y$ is finite. If X has the weak or strong Cartan–Artin property, then the respective property holds for Y.*

Proof. Let $b = (b_1, \dots, b_n)$ be a sequence of geometric points on Y. One argues as in the proof of Proposition 6.4. Using that $f : X_a \to Y_b$ are radical, one sees that the closed points on Y_b have residue fields that are separably closed. Hence Y_b is acyclic with respect to the Nisnevich or étale topology, respectively. □

7. Reduction to TSC schemes

I conjecture that the class of schemes having the weak Cartan–Artin property is stable under images of integral surjections. The goal of this section is to establish a somewhat weaker variant that suffices for our applications. Recall that a scheme X is called *geometrically unibranch* if it is irreducible, and the normalization morphism $Y \to X_{\mathrm{red}}$ is radical. Equivalently, the strictly local rings $\mathscr{O}^s_{X,a}$ are irreducible, for all geometric points a on X, according to [EGA IV$_4$ 1967, Proposition 18.8.15].

Theorem 7.1. *Let $f : X \to Y$ be an integral surjection between schemes that are geometrically unibranch and have affine diagonal. If X has the weak Cartan–Artin property, so does Y.*

This relies on a precise understanding of the connected components inside the X_a, for sequences $a = (a_1, \ldots, a_n)$ of geometric points. We start by collecting some useful facts.

Proposition 7.2. *Suppose X is geometrically unibranch and quasiseparated. Then the connected components of X_a are irreducible.*

Proof. Clearly, the scheme

$$X_a = X_{a_1,\ldots,a_n} = \operatorname{Spec}(\mathscr{O}^s_{X,a_1}) \times_X \cdots \times_X \operatorname{Spec}(\mathscr{O}^s_{X,a_n})$$

is a filtered inverse limit of étale X-schemes that are quasicompact and quasiseparated. Let $U \to X$ be an étale morphism, with U affine. Since $U \to X$ satisfies going-down, each generic point on U maps to the generic point in X. Thus U contains only finitely many irreducible components, because $U \to X$ is quasifinite. Hence for each connected component $U' \subset U$, there is a sequence of irreducible components $U_1, \ldots, U_n \subset U$, possibly with repetitions, so that $U' = \bigcup U_i$ and $U_i \cap U_{i+1} \neq \varnothing$. Choose such a sequence with $n \geq 1$ minimal. Seeking a contradiction, we suppose $n \geq 2$. Choose $u \in U_1 \cap U_2$. Let $x \in X$ be its image. Clearly, $\mathscr{O}_{X,x} \to \mathscr{O}_{U,u}$ becomes bijective upon passing to strict henselizations. Since $\mathscr{O}^s_{X,x}$ is irreducible and $\mathscr{O}_{U,u} \subset \mathscr{O}^s_{U,u}$ satisfies going-down, we conclude that $\mathscr{O}_{U,u}$ is irreducible, contradiction. The upshot is that the connected components of U are irreducible. It remains to apply Lemma 7.3 below with $X_a = V$. $\square$

Lemma 7.3. *Let $V = \varprojlim V_\lambda$ be a filtered inverse limit with flat transition maps of schemes V_λ that are quasicompact and quasiseparated, and whose connected components irreducible. Then each connected component of V is irreducible.*

Proof. Let $C \subset V$ be a connected component. Then the closure of the images $\mathrm{pr}_\lambda(C) \subset V_\lambda$ are connected, whence are contained in a connected component $C_\lambda \subset V_\lambda$. These connected components form an inverse system, and we have canonical maps $C \to \varprojlim C_\lambda \subset \varprojlim V_\lambda = V$. The inclusion is a closed embedding, and $\varprojlim C_\lambda$ is connected, since the C_λ are quasicompact and quasiseparated. In turn $C = \varprojlim C_\lambda$. By assumption, the closed subsets $C_\lambda \subset V_\lambda$ are irreducible. According to a result of Lazard discussed in Appendix A, the connected component $C_\lambda \subset V_\lambda$ is the intersection of its neighborhoods that are open-and-closed. This ensures that the transition maps $C_\mu \to C_\lambda$ remain flat. If follows that the projections $C \to C_\lambda$ are flat as well. By going-down, each generic point $\eta \in C$ maps to the generic point $\eta_\lambda \in C_\lambda$, whence $\eta = (\eta_\lambda)$ is unique, and C must be irreducible. $\square$

Suppose that X is geometrically unibranch and quasiseparated, such that the connected components of X_a coincide with the irreducible components. We may describe the subspace $\operatorname{Min}(X_a) \subset X_a$ of generic points as follows: Let $K = k(X)$ be the function field, and $\mathcal{O}^s_{X,a_i} \subset L_i$ be the field of fractions of the strictly local ring. Since $f : X_a \to X$ is flat and thus satisfies going-down, each generic point of X_a maps to the generic point $\eta \in X$. Since $X_a \to X$ is an inverse limit of quasifinite X-scheme, each point in $f^{-1}(\eta)$ is indeed a generic point of X_a. Thus:

Proposition 7.4. *With the assumptions above, the subspace of generic points* $\operatorname{Min}(X_a) \subset X_a$ *is canonically identified with* $\operatorname{Spec}(L_1 \otimes_K \cdots \otimes_K L_n)$.

In particular, the space $\operatorname{Min}(X_a)$ is profinite. We may describe it in terms of Galois theory as follows: Choose a separable closure $K \subset K^s$ and embeddings $L_i \subset K^s$. Let $G = \operatorname{Gal}(K^s/K)$ be the Galois group, and set $H_i = \operatorname{Gal}(K^s/L_i)$. Then G is a profinite group, the $H_i \subset G$ are closed subgroups, and the quotients G/H_i are profinite. Thus:

Proposition 7.5. *With the assumptions above, the space of generic points* $\operatorname{Min}(X_a)$ *is homeomorphic to the orbit space for the canonical left G-action on the space* $G/H_1 \times \cdots \times G/H_n$.

Proof of Theorem 7.1. Suppose that X, Y are geometrically unibranch, with affine diagonal, and let $f : X \to Y$ be an integral surjection. Assume X satisfies the weak Cartan–Artin property. We have to show that the same holds for Y. Without restriction, it suffices to treat the case that Y, X are integral.

Let $b = (b_1, \dots, b_n)$ be a sequence of geometric points on Y, and $C \subset Y_b$ be a connected component. Our task is to check that C is local henselian. According to Proposition 7.2, the scheme C is integral. Let $\eta \in C$ be the generic point. Since f is surjective and integral, we may lift b to a sequence of geometric points a on X. Since X, Y are geometrically unibranch, the strictly local rings $\mathcal{O}^s_{X,a_i}$ and $\mathcal{O}^s_{Y,b_i}$ remain integral, and we have a commutative diagram

$$\begin{array}{ccc} \mathcal{O}_{X,a_i} & \longrightarrow & K_i \\ \uparrow & & \uparrow \\ \mathcal{O}_{Y,b_i} & \longrightarrow & L_i \end{array}$$

where the terms on the right are the fields of fractions. Let $K = k(X)$ and $L = k(Y)$. Then the induced morphism

$$K_1 \otimes_K \cdots \otimes_K K_n \subset L_1 \otimes_L \cdots \otimes_L L_n$$

is faithfully flat and integral. It follows that each generic point $\eta \in Y_b$ is in the image of $X_a \to Y_b$. The latter morphism is integral by Proposition 6.2, whence also surjective. In particular, there is an irreducible closed subscheme $B \subset X_a$ surjecting onto $C \subset Y_b$. By assumption, B is local henselian, and the morphism $B \to C$ is surjective and integral. Hence C is local henselian, according to Proposition 1.3. □

In particular, an integral normal scheme X_0 has the weak Cartan–Artin property if and only if this holds for the total separable closure $X = \mathrm{TSC}(X_0)$. The latter is a scheme that is everywhere strictly local. For such schemes, the Cartan–Artin properties reduce to a striking geometric property with respect to the Zariski topology:

Theorem 7.6. *Let X be an irreducible quasiseparated scheme that is everywhere strictly local. Then the following are equivalent*:

(i) *The scheme X has the Cartan–Artin property.*

(ii) *The scheme X has the weak Cartan–Artin property.*

(iii) *For every pair of irreducible closed subset $A, B \subset X$ the intersection $A \cap B$ is irreducible.*

Proof. Given a sequence $x = (x_1, \dots, x_n)$ of geometric points, which we may regard as ordinary points $x_i \in X$, we have $\mathscr{O}_{X,x_i} = \mathscr{O}^s_{X,x_i}$, whence

$$X_x = \mathrm{Spec}(\mathscr{O}_{X,x_1}) \cap \cdots \cap \mathrm{Spec}(\mathscr{O}_{X,x_n}),$$

and this is the set of points in X containing each x_i in their closure. Since X is quasiseparated, the scheme X_x is quasicompact, so every point specializes to a closed point. It follows that

$$X_x = \bigcup_z \mathrm{Spec}(\mathscr{O}_{X,z}),$$

where the union runs over all closed points $z \in X_x$.

The implication (i)⇒(ii) is trivial.

To see (iii)⇒(i), it suffices, in light of Theorem 4.4, to check that each X_x contains at most one closed point, that is, X_x is either empty or local. By induction on $n \geq 1$, it is enough to treat the case $n = 2$, with $x_1 \neq x_2$. Now suppose there are two closed points $a \neq b$ in $X_x = \mathrm{Spec}(\mathscr{O}_{X,x_1}) \cap \mathrm{Spec}(\mathscr{O}_{X,x_2})$, and let $A, B \subset X$ be their closure. Then $x_1, x_2 \in A \cap B$. The intersection is irreducible by assumption. Hence the generic point $\eta \in A \cap B$ lies in X_x. Since it is in the closure of $a, b \in X_x$, and these points are closed in X_x, we must have $a = \eta = b$, contradiction.

It remains to verify (ii)⇒(iii). Seeking a contradiction, we suppose that there are irreducible closed subsets $A, B \subset X$ with $A \cap B$ reducible. Choose generic points $x_1 \neq x_2$ in this intersection, and consider the pair $x = (x_1, x_2)$. The resulting scheme $X_x = \mathrm{Spec}(\mathscr{O}_{X,x_1}) \cap \mathrm{Spec}(\mathscr{O}_{X,x_2})$ is irreducible, because this holds for X. In light of Theorem 4.2, X_x must be local. Let $z \in X_x$ be the closed point. Write η_A and η_B for the generic points of A and B, respectively. By construction, $\eta_A, \eta_B \in X_x$, whence $z \in A \cap B$. Using that $x_1, x_2 \in A \cap B$ are generic and contained in the closure of $\{z\} \subset X$, we infer $x_1 = z = x_2$, contradiction. □

8. Algebraic surfaces

Let k be a ground field, and S_0 a separated scheme of finite type, assumed to be integral and 2-dimensional. Note that we do not suppose that S_0 is quasiprojective. Examples of normal surfaces that are proper but not projective appear in [Schröer 1999]. In this section we investigate geometric properties of the absolute integral closure $S = \mathrm{TSC}(S_0)$. Our result is:

Theorem 8.1. *With the assumptions above, let $A, B \subset S$ be two integral, 1-dimensional closed subschemes with $A \neq B$. Then $A \cap B$ contains at most one point.*

Proof. Clearly, we may assume that S_0 is proper and normal, and that the ground field k is algebraically closed. Employing the notation from Section 3, we write

$$S = \varprojlim S_\lambda \quad \text{and} \quad A = \varprojlim A_\lambda \quad \text{and} \quad B = \varprojlim B_\lambda, \qquad \lambda \in L.$$

Seeking a contradiction, we assume that there are two closed points $x \neq y$ in $A \cap B$. Let $x_\lambda, y_\lambda \in A_\lambda \cap B_\lambda$ be their images on S_λ. Enlarging $\lambda = 0$, we may assume that $A_0 \neq B_0$ and $x_0 \neq y_0$.

The integral Weil divisor $B_0 \subset S_0$ has a rational self-intersection number $(B_0)^2 \in \mathbb{Q}$, defined in the sense of Mumford [1961, §II (b)]. Now choose finitely many closed points $z_1, \ldots, z_n \in B_0$ neither contained in A_0 nor the singular locus $\mathrm{Sing}(S_0)$, and let $S_0' \to S_0$ be the blowing-up with reduced center given by these points. If we choose $n \gg 0$ large enough, the self-intersection number of the strict transform $B_0' \subset S_0'$ of B_0 becomes strictly negative. According to Artin [1970, Corollary 6.12] one may contract it to a point: there exist a proper birational morphism $g : S_0' \to Y_0$ with $\mathcal{O}_{Y_0} = g_*(\mathcal{O}_{S_0'})$ sending the integral curve B_0' to a closed point, and being an open embedding on the complement. Note that Y_0 is a proper 2-dimensional algebraic space over k, which is usually not a scheme. We also write $A_0 \subset S_0'$ for the curve A_0, considered as a closed subscheme of S_0'. The projection $A_0 \to Y_0$ is a finite morphism, which is not injective because the two points $x_0, y_0 \in A_0 \cap B_0'$ are mapped to a common image.

We now make an analogous construction in the inverse system. Consider the normalization $S_\lambda' \to S_\lambda \times_{S_0} S_0'$ of the integral component dominating S_0'. The universal properties of normalization and fiber product ensure that the S_λ', $\lambda \in L$ form an inverse system, and the transition maps are obviously affine. In turn, we get an integral scheme

$$S' = \varprojlim S_\lambda',$$

coming with a birational morphism $S' \to S$. This scheme S' is totally separably closed. Clearly, the projection $S_\lambda' \to S_\lambda$ are isomorphisms over a neighborhood of the preimage of $A_0 \subset S_0$, which contains $A_\lambda \subset S'$. Therefore, we may regard A_λ and the points x_λ, y_λ also as a closed subscheme of S_λ'.

Let $S'_\lambda \to Y_\lambda$ be the Stein factorization of the composition $S'_\lambda \to S'_0 \to Y_0$. This map contracts the connected components of the preimage of $B'_0 \subset S'_0$ to closed points, and is an open embedding on the complement. Using that this preimage contains the strict transform $B'_\lambda \subset S'_\lambda$ of $B_\lambda \subset S_\lambda$, which is irreducible and thus connected, we infer that $x_\lambda, y_\lambda \in S'_\lambda$ map to a common image in Y_λ. By the universal property of Stein factorizations, the Y_λ, $\lambda \in L$ form a filtered inverse system. The transition morphisms $Y_\lambda \to Y_j$ are finite, because these schemes are finite over Y_0. In turn, we get another totally separably closed scheme $Y = \varprojlim Y_\lambda$, endowed with a birational morphism $S' \to Y$. By construction, the finite morphism $A_\lambda \to Y_\lambda$ maps x_λ and y_λ to a common image. Passing to the inverse limit, we get an integral morphism

$$A = \varprojlim A_\lambda \to \varprojlim Y_\lambda = Y,$$

which maps $x \neq y$ to a common image. By assumptions, A is an integral scheme. According to Lemma 2.3, the map $A \to Y$ must be injective, contradiction. $\square$

9. Two facts on noetherian schemes

Before we come to higher-dimensional generalizations of Theorem 8.1, we have to establish two properties pertaining to quasiprojectivity and connectedness that might be of independent interest.

Proposition 9.1. *Let R be a noetherian ground ring, X a separated scheme of finite type, and $U \subset X$ be a quasiprojective open subset. Then there is a closed subscheme $Z \subset X$ disjoint from U and containing no point $x \in X$ of codimension* $\dim(\mathcal{O}_{X,x}) = 1$ *such that the blowing up $f : X' \to X$ with center Z yields a quasiprojective scheme X'.*

Proof. By a result of Gross [2012, proof of Theorem 1.5] there exists an open subscheme $V \subset X$ containing U and all points of codimension ≤ 1 that is quasiprojective over R. (Note that Gross's overall assumption that the ground ring R is excellent does not enter in this result.) By Chow's lemma (in the refined form of [Deligne 2010, Corollary 1.4]), there is a blowing-up $X' \to X$ with center $Z \subset X$ disjoint from V so that X' becomes quasiprojective over R. $\square$

Proposition 9.2. *Let X be a noetherian scheme satisfying Serre's Condition (S_2), and $D \subset X$ a connected Cartier divisor. Suppose that the local rings $\mathcal{O}_{X,x}$, $x \in X$ are catenary and homomorphic images of Gorenstein local rings. Then there is a sequence of irreducible components $D_0, \dots, D_r \subset D$, possibly with repetitions and covering D such that the successive intersections $D_{i-1} \cap D_i$, $1 \leq i \leq r$ have codimension ≤ 2 in X.*

Proof. Since D is connected and noetherian, there is a sequence of irreducible components $D_0, \dots, D_s \subset D$, possibly with repetitions and covering D so that the successive intersections are nonempty [EGA I 1971, Corollary 2.1.10]. Fix an index $1 \leq i \leq r$, choose a closed point $x \in D_{i-1} \cap D_i$, let $R = \mathcal{O}_{X,x}$ be the corresponding local ring on X, and $f \in R$ the regular element defining the Cartier divisor D at x.

To finish the proof, we insert between D_{i-1} and D_i a sequence of further irreducible components satisfying the codimension condition of the assertion on the local ring R. This can be achieved with Grothendieck's connectedness theorem [SGA 2 1968, Exposé XIII, Theorem 2.1] given in the form of Flenner, O'Carroll and Vogel [Flenner et al. 1999, §3.1] as follows:

Recall that a noetherian scheme S is called *connected in dimension d* if we have $\dim(S) > d$, and the complement $S \setminus T$ is connected for all closed subsets of dimension $\dim(T) < d$. Set $n = \dim(R)$. Since the local ring R is catenary and satisfies Serre's condition (S_2), it must be connected in dimension $n-1$ [Flenner et al. 1999, Corollary 3.1.13]. It follows that R/fR is connected in dimension $n-2$ by Lemma 3.1.11 of [loc. cit.], which is essentially Grothendieck's connectedness theorem. Consequently, there is a sequence $H_0, \dots, H_q$ of irreducible components of $\mathrm{Spec}(R/fR)$ so that H_0 and H_q are the local schemes attached to $x \in D_{i-1}, D_i$, and whose successive intersections have dimension $\geq n-2$ in $\mathrm{Spec}(R)$, whence codimension ≤ 2. The closures of the $H_0, \dots, H_q$ inside D yield the desired sequence of further irreducible components. $\square$

Let me state this result in a simplified form suitable for most applications:

Corollary 9.3. *Let X be a normal irreducible scheme that is of finite type over a regular noetherian ring R, and $D \subset X$ a connected Cartier divisor. Then there is a sequence of irreducible components $D_0, \dots, D_r \subset D$, possibly with repetitions, covering D such that the successive intersections $D_{i-1} \cap D_i$, $1 \leq i \leq r$ have codimension ≤ 2 in X.*

Proof. For every affine open subset $U \subset X$, the ring $A = \Gamma(U, \mathcal{O}_X)$ is the homomorphic image of some polynomial ring $R[T_1, \dots, T_n]$. Since the latter is regular, in particular Gorenstein and catenary, it follows that all local rings $\mathcal{O}_{X,x}$ are catenary, and homomorphic images of Gorenstein local rings. $\square$

10. Contractions to points

Let k be a ground field, and X a proper scheme assumed to be normal and irreducible. Set $n = \dim(X)$. We say that a connected closed subscheme $D \subset X$ is *contractible to a point* if there is a proper algebraic space Y and a morphism $g : X \to Y$ with $\mathcal{O}_Y = f_*(\mathcal{O}_X)$ sending D to a closed point and being an open embedding on the complement. Note that such a morphism is automatically proper, because X is proper and Y is separable. Moreover, it is unique up to unique isomorphism.

Let $C \subsetneq X$ be a connected closed subscheme. The goal of this section is to construct a projective birational morphism $f : X'' \to X$ so that some connected Cartier divisor $D'' \subset X''$ with $f(D'') = C$ becomes contractible to a point, generalizing the procedure in dimension $n = 2$ used in the proof for Theorem 8.1. The desired morphism f will be a composition $f = g \circ h$ of two birational morphisms.

The first morphism $g : X' \to X$ will replace the connected closed subscheme $C \subset X$ by some connected Cartier divisor $D' \subset X'$ that is a *projective* scheme. It will be itself a composition of three projective birational morphisms

$$X \xleftarrow{g_1} X_1 \xleftarrow{g_2} X_2 \xleftarrow{g_3} X_3 = X'.$$

Here $g_1 : X_1 \to X$ is the normalized blowing-up with center $C \subset Y$. Such a map has connected fibers by Zariski's main theorem. Let $D_1 = g_1^{-1}(C)$. According to Proposition 9.1, there is a closed subscheme $Z \subset D_1$ containing no point $d \in D_1$ of codimension $\dim(\mathscr{O}_{D_1,d}) \leq 1$ so that the blowing-up $\widetilde{D}_1 \to D_1$ becomes projective. Let $g_2 : X_2 \to X_1$ be the normalized blowing-up with the same center regarded as a closed subscheme $Z \subset X$. Viewing $\widetilde{D}_1$ as a closed subscheme on the blowing-up $\widetilde{X}_1 \to X_1$ with center Z, we denote by $D_2 \subset X_2$ its preimage on the normalized blowing-up, which is a Weil divisor. Finally, let $g_3 : X = X_3 \to X_2$ be the normalized blowing-up with center $D_2 \subset X_2$, and $D' = D_3$ be the preimage of D_2. Then $D' \subset X'$ is a Cartier divisor by the universal property of blowing-ups. The morphism $X' \to X$ has connected fibers, by Zariski's main theorem, and induces a surjection $D' \to C$. Since C is connected, it follows that D' is connected. The proper scheme D' is projective, because $\widetilde{D}$ is projective, and finite maps and blowing-ups are projective morphisms. Setting $g = g_1 \circ g_2 \circ g_3 : X' \to X$, we record:

Proposition 10.1. *The scheme X' is proper and normal, the morphism $g : X' \to X$ is projective and birational, the closed subscheme $D' \subset X'$ is a connected Cartier divisor that is a projective scheme, and $g(D') = C$.*

In a second step, we now construct the proper birational morphism $h : X'' \to X'$, so that X'' contains the desired contractible effective Cartier divisor. Choose an ample Cartier divisor $A \subset D'$ containing no generic point from the intersection of two irreducible component of D', and so that the invertible sheaf $\mathscr{O}_{D'}(A - D')$ is ample. Let $\varphi : \widetilde{X}' \to X'$ be the blowing-up with center $A \subset X'$. By the universal property of blowing-ups, there is a partial section $\sigma : D' \to \widetilde{X}'$. Let $\nu : X'' \to \widetilde{X}'$ be the normalization map, such that the composition

$$h : X'' \xrightarrow{\nu} \widetilde{X}' \xrightarrow{\varphi} X'$$

is the normalized blowing-up with center A. Set $D'' = \nu^{-1}(\sigma(D')) \subset X''$, and let $f = g \circ h : X'' \to X$ be the composite morphism.

Theorem 10.2. *The scheme X'' is proper and normal, the morphism $f : X'' \to X$ is projective and birational, and $D'' \subset X''$ is a connected Cartier divisor that is a* *projective scheme with $f(D'') = C$. Moreover, the closed subscheme $D'' \subset X''$ is contractible to a point.*

Proof. By construction, X'' is normal and proper, f is projective and birational, and $f(D'') = C$. According to [Perling and Schröer 2014, Lemma 4.4] the subschemes

$$\sigma(D'), \varphi^{-1}(A), \varphi^{-1}(D') \subset \widetilde{X}'$$

are Cartier, with $\varphi^{-1}(D) = \sigma(D') + \varphi^{-1}(A)$ in the group of Cartier divisors on $\widetilde{X}'$. Pulling back along the finite surjective morphism between integral schemes $\nu : X'' \to \widetilde{X}'$, we get Cartier divisors $D'' = \nu^{-1}(\sigma(D'))$, $E = h^{-1}(A)$, and $\widehat{D} = h^{-1}(D')$ on X'' satisfying $\widehat{D} = D'' + E$. In turn, $\mathscr{O}_{D''}(-D'') \simeq \mathscr{O}_{D''}(E - \widehat{D})$. The latter is isomorphic to the preimage with respect to the normalization map $\nu : X'' \to \widetilde{X}'$ of

$$\mathscr{O}_{\sigma(D')}(-1) \otimes \varphi^{*}\mathscr{O}_{D'}(-D').$$

The first tensor factor becomes, under the canonical identification $\sigma(D') \to D'$, the invertible sheaf $\mathscr{O}_D(A)$. Summing up, $\mathscr{O}_{D''}(-D'')$ is isomorphic to the pull back of the ample invertible sheaf $\mathscr{O}_{D'}(A - D')$ under a finite morphism, whence is ample. By Artin's contractibility criterion [Artin 1970, Corollary 6.12] the connected components of D'' are contractible to points.

It remains to verify that D'' is connected. For this it suffices to check that the dense open subscheme $D'' \setminus E$ is connected, which is isomorphic to $D' \setminus A$. By Proposition 9.2, there is a sequence of irreducible components $D'_1, \dots, D'_r \subset D'$, possibly with repetitions, covering D'' so that the successive intersections $D'_{i-1} \cap D'_i$ have codimension ≤ 2 in X', whence codimension ≤ 1 in D'_{i-1} and D'_i. By the choice of $A \subset D'$, the complements $(D'_{i-1} \cap D'_i) \setminus A$ are nonempty, and it follows that $D' \setminus A$ is connected. $\square$

11. Cyclic systems

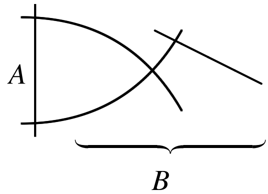
Let X be a scheme. Let us call a pair of closed subschemes $A, B \subset X$ a *cyclic system* if the following hold:

(i) The space A is irreducible.

(ii) The space B is connected.

(iii) The intersection $A \cap B$ is disconnected.

This ad hoc definition will be useful when dealing with totally separably closed schemes, and is somewhat more flexible than the "polygons" used in [Artin 1971], see also [Sergio 1982]. Note that this notion is entirely topological in nature, and that the conditions ensures that A and B are nonempty and $A \not\subset B$. One should bear in mind a picture like the following:



In this section, we establish some elementary but technical permanence properties for cyclic systems, which will be used later.

Proposition 11.1. *Let $A, B \subset X$ be a cyclic system, and $f : X' \to X$ closed surjection with connected fibers. Let $A' \subset X'$ be a closed irreducible subset with $f(A') = A$ and $B' = f^{-1}(B)$. Then the pair $A', B' \subset X'$ is a cyclic system.*

Proof. By assumption, A' is irreducible. Since f is closed, surjective and continuous, the space B carries the quotient topology with respect to $f : B' \to B$. Thus B' is connected, because B is connected and f has connected fibers [EGA I 1971, Proposition 2.1.14]. The set-theoretic "projection formula" gives

$$f(A' \cap B') = f(A' \cap f^{-1}(B)) = f(A') \cap B = A \cap B,$$

whence $A' \cap B'$ is disconnected. □

Proposition 11.2. *Let $A, B \subset X$ be a cyclic system, and $f : X' \to X$ be a morphism. Let $U \subset X$ be an open subset over which f becomes an isomorphism, and let $A', B' \subset X'$ be the closures of $f^{-1}(A \cap U)$, $f^{-1}(B \cap U)$, respectively. Suppose the following:*

(i) *$B \cap U$ is connected.*

(ii) *U intersects at least two connected components of $A \cap B$.*

Then the pair $A', B' \subset X'$ is a cyclic system.

Proof. Since A is irreducible, so is the nonempty open subset $A \cap U$, and thus the closure A'. Since $B \cap U$ is connected by Condition (i), the same holds for its closure B'. It remains to check that $A' \cap B'$ is disconnected. Since f is continuous and $A \subset X$ is closed, we have

$$f(A') = f(\overline{f^{-1}(A \cap U)}) \subset \overline{A \cap U} \subset A,$$

and similarly $f(B') \subset B$. Hence $f(A' \cap B') \subset A \cap B$. Seeking a contradiction, we assume that $A' \cap B'$ is connected. Choose points $u, v \in A \cap B$ from two different connected components with $u, v \in U$. Regarding u, v as elements from X' via $U = f^{-1}(U)$, we see that they are contained in the connected subset $A' \cap B' \subset X'$, whence the points $u, v \in A \cap B$ are contained in a connected subset of $f(A' \cap B')$, contradiction. □

Proposition 11.3. *Let $X = \varprojlim X_\lambda$ be a filtered inverse system of quasicompact quasiseparated schemes with integral transition maps, and $A, B \subset X$ be a cyclic system. Then their images $A_\lambda, B_\lambda \subset X_\lambda$ form cyclic systems for a cofinal subset of indices λ.*

Proof. The subsets $A_\lambda, B_\lambda \subset X_\lambda$ are irreducible and connected, because they are images of such spaces. They are closed, because the projections $X \to X_\lambda$ are closed maps. It remains to check that $A_\lambda \cap B_\lambda$ are disconnected for a cofinal subset of indices. Since inverse limits commute with fiber products, we have

$$\varprojlim(A_\lambda) \cap \varprojlim(B_\lambda) = \varprojlim(A_\lambda \cap B_\lambda).$$

The canonical inclusion $A \subset \varprojlim(A_\lambda)$ of closed subsets in X is bijective, since the underlying set of the letter is the inverse limit of the underlying sets of the X_λ. Similarly we have $B = \varprojlim(B_\lambda)$ as closed subsets. Thus $A \cap B = \varprojlim(A_\lambda \cap B_\lambda)$. By assumption, $A \cap B$ is disconnected and the $A_\lambda \cap B_\lambda$ are quasicompact. It follows from Proposition A.2 that $A_\lambda \cap B_\lambda$ must be disconnected for a cofinal subset of indices λ. □

12. Algebraic schemes

We now come to the main result of this paper, which generalizes Theorem 8.1 from surfaces to arbitrary dimensions. It takes the following form:

Theorem 12.1. *Let k be a ground field and X_0 an integral scheme that is separated, connected and of finite type. Then its total separable closure $X = \mathrm{TSC}(X_0)$ contains no cyclic system. In particular, for any pair $A, B \subset X$ of irreducible closed subsets, the intersection $A \cap B$ remains irreducible.*

Proof. Seeking a contradiction, we assume that there is a cyclic system $A, B \subset X$. Passing to normalization and compactification, we easily reduce to the situation that X_0 is normal and proper. As usual, write the total separable closure as an inverse limit of finite X_0-schemes X_λ, $\lambda \in L$, that are connected normal, and let $A_\lambda, B_\lambda \subset X_\lambda$ be the images of $A, B \subset X$, respectively. Then we have

$$X = \varprojlim X_\lambda, \quad A = \varprojlim A_\lambda, \quad \text{and} \quad B = \varprojlim B_\lambda,$$

and all transition morphisms are finite and surjective. Clearly, the A_λ are irreducible and the B_λ are connected. In light of Proposition 11.3, we may replace L by a cofinal subset and assume that the $A_\lambda, B_\lambda \subset X_\lambda$ form cyclic systems for all $\lambda \in L$, including $\lambda = 0$.

To start with, we deal with a rather *special case*, namely that the connected subset $B_0 \subset X_0$ is *contractible to a point*. Let $X_0 \to Y_0$ be the contraction. Then the composite map $A_0 \to Y_0$ has a disconnected fiber, because the connected components of the disconnected subset $A_0 \cap B_0$ are mapped to a common image point. Next, consider the induced maps $X_\lambda \to Y_\lambda$ contracting the connected components of the preimage $X_\lambda \times_{X_0} B_0 \subset X_\lambda$ to points. Again the composite map $A_\lambda \to Y_\lambda$ has a disconnected fiber. This is because the irreducible subset A_λ is not contained in the preimage $X_\lambda \times_{X_0} B_0$, the connected subset B_λ is contained in $X_\lambda \times_{X_0} B_0$, thus contained in a connected component of the latter, whence maps to a point in the scheme Y_λ.

To proceed, consider the Stein factorization

$$A_\lambda \to A'_\lambda \to Y_\lambda.$$

Then $A'_\lambda \to Y_\lambda$ is finite, but not injective. Passing to the inverse limits $A' = \varprojlim A'_\lambda$ and $Y = \varprojlim Y_\lambda$, we obtain a totally separably closed scheme Y, an integral scheme A', and an integral morphism $A' \to Y$. The latter must be radical, by Lemma 2.3. In turn, the maps $A'_\lambda \to Y_\lambda$ are radical for sufficiently large indices $\lambda \in L$, according to [EGA IV$_3$ 1966, Theorem 8.10.5]. In particular, these maps are injective, contradiction.

We now come to the *general case*. We shall proceed in six steps, some of which are repeated. In each step, a given proper normal connected scheme X_0 whose absolute separable closure $X = \mathrm{TSC}(X_0)$ contains a cyclic system $A, B \subset X$ will be replaced by another such scheme with slightly modified properties, until we finally reach the special case discussed in the preceding paragraphs. As we saw, the latter is impossible. To start with, choose points $u, v \in A \cap B$ coming from different connected components. Let $u_\lambda, v_\lambda \in A_\lambda \cap B_\lambda$ be their images, respectively.

Step 1: *We first reduce to the case that $u_0, v_0 \in A_0 \cap B_0$ lie in different connected components.* Suppose that for all indices $\lambda \in L$, the points $u_\lambda, v_\lambda \in A_\lambda \cap B_\lambda$ are contained in the same connected component $C_\lambda \subset A_\lambda \cap B_\lambda$. Then the C_λ, $\lambda \in L$, form an inverse system with finite transition maps, and their inverse limit $C = \varprojlim C_\lambda$ is connected ([EGA IV$_3$ 1966, Proposition 8.4.1], see also Appendix A). Using $u, v \in C \subset A \cap B$, we reach a contradiction. Hence there exist an index $\lambda \in L$ such that $u_\lambda, v_\lambda \in A_\lambda \cap B_\lambda$ stem from different connected components. Replacing X_0 by such an X_λ we thus may assume that $u_0, v_0 \in A_0 \cap B_0$ lie in different connected components.

Step 2: *Next we reduce to the case that $u_0, v_0 \in A_0 \cap B_0$ are not contained in a common irreducible component of B_0.* Let $f : X'_0 \to X_0$ be the normalized blowing-up with center $A_0 \cap B_0 \subset X_0$, such that the strict transforms of A_0, B_0 on X'_0 become disjoint.

Let X'_λ be the normalization of $X_\lambda \times_{X_0} X'_0$ inside the function field $k(X_\lambda) = k(X'_\lambda)$. Then the X'_λ, $\lambda \in L$, form an inverse system of proper normal connected schemes whose transition morphisms are finite and dominant. Let X' be their inverse limit.

By construction, we have projective birational morphisms $X'_\lambda \to X_\lambda$ and a resulting birational morphism $X' \to X$, which is a proper surjective map with connected fibers, according to [EGA IV$_3$ 1966, Proposition 8.10.5 (xii), (vi) and (vii)]. Note that all these X-morphisms become isomorphisms when base-changed to $X_0 \setminus (A_0 \cap B_0)$.

Let $A' \subset X'$ be the strict transform of $A \subset X$, and $B' \subset X'$ the reduced scheme-theoretic preimage of $B \subset X$. These form a cyclic system on X', by Proposition 11.1. Consider their images $A'_\lambda, B'_\lambda \subset X'_\lambda$, which are closed subsets. The A'_λ are irreducible and the B'_λ are connected. Clearly, the A'_λ can be viewed as the strict transforms of $A_\lambda \subset X_\lambda$. Moreover, the inclusion $B'_\lambda \subset X'_\lambda \times_{X_\lambda} B_\lambda$ is an equality of sets. This is because the morphisms $B \to B_\lambda$ and $X' \to X$ are surjective, by [EGA IV$_3$ 1966, Theorem 8.10.5 (vi)]. Again with Proposition 11.1, we infer that all $A_\lambda, B_\lambda \subset X_\lambda$ form a cyclic system.

Let $\varphi : X' \to X$ be the canonical morphism. The set-theoretic projection formula gives

$$\varphi(A' \cap B') = A \cap B.$$

Choose points $u', v' \in A' \cap B'$ mapping to u and v, respectively, and consider their images $u'_0, v'_0 \in A'_0 \cap B'_0$. The latter are not contained in the same connected component, because they map to $u_0, v_0 \in A_0 \cap B_0$. Moreover, we claim that the $u'_0, v'_0 \in A'_0 \cap B'_0$ are not contained in a common irreducible component of B'_0: Suppose they would lie in a common irreducible component $W'_0 \subset B'_0$. Since the strict transforms of A_0, B_0 on X'_0 are disjoint, the image $f(W'_0) \subset X_0$ under the blowing-up $f : X'_0 \to X_0$ must lie in the center $A_0 \cap B_0$. Then $u_0, v_0 \in f(W_0) \subset A_0 \cap B_0$ are contained in some connected subset, contradiction.

Summing up, after replacing X_0 by X'_0, we may additionally assume that $u_0, v_0 \in B_0$ are not contained in a common irreducible component of B_0.

Step 3: *The next goal is to turn the closed subscheme $B_0 \subset X_0$ into a Cartier divisor.* Let $X'_0 \to X_0$ be the normalized blowing-up with center $B_0 \subset X_0$, such that its preimage on X'_0 becomes a Cartier divisor. Define $X'_\lambda, X', A', B', A'_\lambda, B'_\lambda$ as in Step 2. Then $A', B' \subset X'$ and the $A'_\lambda, B'_\lambda \subset X'_\lambda$ form cyclic systems.

Again let $\varphi : X' \to X$ be the canonical morphism. Then $\varphi(A' \cap B') = A \cap B$ by the projection formula. Let $u', v' \in A' \cap B'$ be in the preimage of the points $u, v \in A \cap B$. Since their images $u_0, v_0 \in A_0 \cap B_0$ do not lie in a common connected component and are not contained in a common irreducible component of B, the same necessarily holds for their images $u'_0, v'_0 \in A'_0 \cap B'_0$.

Replacing X_0 by X'_0, we thus may additionally assume that $B_0 \subset X_0$ is the support of an effective Cartier divisor.

Step 4: *We now reduce to the case that all $B_\lambda \subset X_\lambda$ are supports of connected effective Cartier divisors.* The preimages $X_\lambda \times_{X_0} B_0 \subset X_\lambda$ of the Cartier divisor $B_0 \subset X_0$ remain Cartier divisors, because $X_\lambda \to X$ is a dominant morphism between normal schemes. Let

$$C_\lambda \subset X_\lambda \times_{X_0} B_0$$

be the connected component containing the connected subscheme

$$B_\lambda \subset X_\lambda \times_{X_0} B_0.$$

Clearly, the $C_\lambda \subset X_\lambda$ are connected Cartier divisors, which form an inverse system such that $C = \varprojlim C_\lambda$ contains $B = \varprojlim B_\lambda$. Note that the transition maps $C_j \to C_\lambda$, $i \leq j$ and the projections $C \to C_\lambda$ are not necessarily surjective. Nevertheless, we conclude with [EGA IV$_3$ 1966, Proposition 8.4.1] that C is connected. Clearly, the points $u, v \in A \cap C$ lie in different connected components, and are not contained in a common irreducible component of C, because this hold for their images $u_0, v_0 \in A_0 \cap B_0$. Thus $A, C \subset X$ form a cyclic system.

Replacing B by C and B_λ by C_λ, we thus may additionally assume that all $B_\lambda \subset X_\lambda$ are supports of connected Cartier divisors. Note that at this stage we have sacrificed our initial assumption that the projections $B \to B_\lambda$ and the transition maps $B_j \to B_\lambda$ are surjective.

Step 5: *Here we reduce to the case that the proper scheme B_0 is projective.* Since $u_0, v_0 \in B_0$ are not contained in a common irreducible component, they admit a common affine open neighborhood inside B_0: To see this, let $B_0' \subset B_0$ be the finite union of the irreducible components that do not contain u_0, and let $B_0'' \subset B_0$ be the finite union of the irreducible components that do not contain v_0. Then $B_0 = B_0' \cup B_0''$, so any affine open neighborhoods of $u_0 \in B_0 \setminus B_0'$ and $v_0 \in B_0 \setminus B_0''$ are disjoint, and their union is the desired common affine open neighborhood.

It follows that there is a closed subscheme $Z_0 \subset B_0$ containing neither u_0, v_0 nor any point $b \in B_0$ of codimension $\dim(\mathscr{O}_{B_0,b}) = 1$ so that the blowing-up $\widetilde{B}_0 \to B_0$ with center Z_0 yields a projective scheme $\widetilde{B}_0$. This holds by Proposition 9.1. Note that $Z_0 \subset B_0$ has codimension ≥ 2.

Let $X_0' \to X_0$ be the normalized blowing-up with the same center Z_0, regarded as a closed subscheme $Z_0 \subset X_0$. Define X_λ' and $X' = \varprojlim X_\lambda'$ as in Step 2. Let $U \subset X$ be the complement of the closed subscheme $X \times_{X_0} Z_0 \subset X$. This open subscheme intersects at least two connected components of $A \cap B$, because $u, v \in U$. Clearly, the canonical morphism $f : X' \to X$ becomes an isomorphism over U. Furthermore, the intersection $B \cap U$ is connected. To see this, set $Z_\lambda = X_\lambda \times_{X_0} Z_0$ and $U_\lambda = X_\lambda \setminus Z_\lambda$, such that $B \cap U = \varprojlim (B_\lambda \cap U_\lambda)$. In light of [EGA IV$_3$ 1966, Proposition 8.4.1] it suffices to check that $B_\lambda \cap U_\lambda$ is connected for each $\lambda \in L$. Indeed, by the going-down theorem applied to the finite morphism $B_\lambda \to B_0$, all $Z_\lambda \subset B_\lambda$ have codimension ≥ 2. The $B_\lambda \subset X_\lambda$ are supports of connected Cartier divisors. In light of Corollary 9.3, the set $B_\lambda \cap U_\lambda = B_\lambda \setminus Z_\lambda$ must remain connected.

According to Proposition 11.2, the closures $A', B' \subset X'$ of $A \cap U$ and $B \cap U$ form a cyclic system. Clearly, the image of the composite map $B' \to X_0$ is $B_0 \subset X_0$. Replacing X by X', we may assume that B_0 is a projective scheme. Note that with this step we have sacrificed the property that the $B_\lambda \subset X_\lambda$ are Cartier divisors.

Step 6: *We now repeat Step 4, and achieve again that $B_\lambda \subset X_\lambda$ are supports of connected Cartier divisors.* Moreover, the B_λ stay projective, because B_0 is projective and the $X_\lambda \to X_0$ are finite.

Step 7: *In this last step we achieve that $B_0 \subset X_0$ becomes contractible to a point, thus reaching a contradiction.* Choose an ample Cartier divisor $H_0 \subset B_0$ containing no generic point from the intersection of two irreducible components of B_0, and no generic point from the intersection $A_0 \cap B_0$, and no point from $\{u_0, v_0\}$. Furthermore, we demand that the invertible $\mathscr{O}_{B_0}$-module $\mathscr{O}_{B_0}(H_0 - B_0)$ is ample. Let $X'_0 \to X_0$ be the normalized blowing-up with center $H_0 \subset X_0$, and $A'_0, B'_0 \subset X'_0$ the strict transforms of $A_0, B_0 \subset X_0$, respectively. As in the proof for Theorem 10.2, the subscheme $B'_0 \subset X'_0$ is connected and contractible to a point. Define X'_λ and $X' = \varprojlim X'_\lambda$ as in Step 2.

Let $U \subset X$ be the complement of the closed subscheme $X \times_{X_0} H_0$. This open subscheme intersects at least two connected components of $A \cap B$, because $u, v \in U \cap V$. Clearly, the canonical morphism $f : X' \to X$ becomes an isomorphism over U. Furthermore, the intersection $B \cap U$ is connected. To see this, set $H_\lambda = X_\lambda \times_{X_0} H_0$ and $U_\lambda = X_\lambda \setminus H_\lambda$, such that

$$B \cap U = \varprojlim (B_\lambda \cap U_\lambda).$$

In light of [EGA IV$_3$ 1966, Proposition 8.4.1] it suffices to check that $B_\lambda \cap U_\lambda = B_\lambda \setminus H_\lambda$ is connected for each $\lambda \in L$. Since $B_\lambda \subset X_\lambda$ is the support of a connected Cartier divisor, there is a sequence of irreducible components $C_1, \dots, C_n$ covering B_λ so that each successive intersection $C_{j-1} \cap C_j$ has codimension ≤ 2 in X_λ. Let $f : X_\lambda \to X_0$ be the canonical morphisms, and consider the inclusion

$$f(C_{j-1} \cap C_j) \subset f(C_{j-1}) \cap f(C_j).$$

Since f is finite and dominant, the left hand side has codimension ≤ 2 in X_0, and the $f(C_{j-1})$ and $f(C_j)$ are irreducible components of B_0. By the choice of the ample Cartier divisor $H_0 \subset X_0$, the image $f(C_{j-1} \cap C_j)$ is not entirely contained in H_0. We conclude that $(C_{j-1} \cap C_j) \setminus H_\lambda$ is nonempty, thus $B_\lambda \setminus H_\lambda$ remains connected.

Let $A', B' \subset X'$ be the closure of $A \cap U$ and $B \cap U$. The former comprise a cyclic system, in light of Proposition 11.2. Clearly, the image of $B' \to X_0$ equals B'_0. Replacing X_0 by X'_0, we thus have achieved the situation that $B_0 \subset X_0$ is contractible to a point. We have seen in the beginning of this proof that such a cyclic system $A, B \subset X$ does not exist. □

13. Nisnevich Čech cohomology

We now apply our results to prove the following:

Theorem 13.1. *Let k be a ground field, and X be a quasicompact and separated k-scheme. Then $\check{H}^p(X_{\mathrm{Nis}}, \underline{H}^q(F)) = 0$ for all integers $p \geq 0$, $q \geq 1$ and all abelian Nisnevich sheaves F on X. In particular, the canonical maps*

$$\check{H}^p(X_{\mathrm{Nis}}, F) \to H^p(X_{\mathrm{Nis}}, F)$$

are bijective for all $p \geq 0$.

Proof. Recall that $\check{H}^p(X_{\mathrm{Nis}}, \underline{H}^q(F)) = \varinjlim_U \check{H}^p(U_{\mathrm{Nis}}, \underline{H}^q(F))$, where the direct limit runs over the system of all completely decomposed étale surjections $U \to X$. For an explanation of the transition maps, we refer to [Godement 1958, Chapter II, Section 5.7]. Since X is quasicompact and étale morphisms are open, one easily sees that the subsystem of all étale surjections $U \to X$ with U affine form a cofinal subsystem.

Let $[\alpha] \in \check{H}^p(X_{\mathrm{Nis}}, \underline{H}^q(F))$ be a Čech class with $q \geq 1$. Represent the class by some cocycle $\alpha \in H^q(U^p, F)$, where $U \to X$ is a completely decomposed étale surjection, and U^p denotes the p-fold self-product in the category of X-schemes. This self-product remains quasicompact, because X is quasiseparated. Since $q \geq 1$, there is a completely decomposed étale surjection $W \to U^p$ with W affine so that $\alpha | W = 0$. It suffices to show that there is a refinement $U' \to U$ so that the structure morphisms $U'^p \to U^p$ factors over $W \to U^p$, because then the class of α in the direct limit of the Čech complex vanishes. From this point on we merely work with the schemes W, U, X and may forget about the sheaf F and the cocycle α.

This allows us to reduce to the case of k-schemes of finite type: Write $X = \varprojlim X_\lambda$ as an inverse limit with affine transition maps of k-schemes X_λ that are of finite type. Passing to a cofinal subset of indices, we may assume that there is a smallest index $\lambda = 0$, so that X_0 is separated [Thomason and Trobaugh 1990, Appendix C, Proposition 7]. Since U is quasicompact, our morphisms $U \to X$ is of finite presentation, so we may assume that $U = U_0 \times_{X_0} X_0$ for some X_0-scheme U_0 of finite presentation. Similarly, we can assume $W = W_0 \times_{U_0^p} U^p$ for some U_0^p-scheme W_0 of finite presentation. We may assume that $U_0 \to X_0$ and $W_0 \to U_0^p$ are étale [EGA IV$_4$ 1967, Proposition 17.7.8] and surjective [EGA IV$_3$ 1966, Proposition 8.10.5]. According to Lemma 13.2 below, we can also impose that these morphisms are completely decomposed. Summing up, it suffices to produce a refinement $U_0' \to U_0$ so that $U_0'^p \to U_0^p$ factors over W_0. In other words, we may assume that X is of finite type over the ring $\mathbb{Z}$. In particular, X has only finitely many irreducible components, the normalization of the corresponding integral schemes are finite over X, and X is noetherian.

Next, we make use of the weak Cartan–Artin property. Indeed, according to Artin's induction argument [1971, Theorem 4.1] applied to the Nisnevich topology rather than the étale topology, we see that it suffices to check that X satisfies the weak Cartan–Artin property. Let X_0 be the normalization of some irreducible component of X, viewed as an integral scheme. According to Propositions 6.4 and 6.6, it is permissible to replace X by X_0. Changing notation, we write $X = \mathrm{TSC}(X_0)$ for the total separable closure of the integral scheme X_0. In light of Theorem 7.1, it is enough to verify the weak Cartan–Artin property for X.

Here, our problem reduces to a simple geometric statement: according to Theorem 7.6, it suffices to check that for each pair of irreducible closed subsets $A, B \subset X$, the intersection $A \cap B$ remains irreducible. Indeed, this holds by Theorem 12.1. Note that only in this very last step, we have used that X is separated, and the existence of a ground field k. $\square$

In the preceding proof, we used the following facts:

Lemma 13.2. *Suppose that $X = \varprojlim X_\lambda$ is a filtered inverse system of quasicompact schemes with affine transition maps, and $U \to X$ is a completely decomposed étale surjection, with U quasicompact. Then there is an index α and a completely decomposed étale surjection $U_\alpha \to X_\alpha$ with $U = X \times_{X_\alpha} U_\alpha$.*

Proof. The morphism $U \to X$ is of finite presentation, because it is étale and its domain is quasicompact. According to [EGA IV$_3$ 1966, Theorem 8.8.2] there is an X_α-scheme U_α of finite presentation with $U = X \times_{X_\alpha} U_\alpha$ for some index α. We may assume that $U_\alpha \to X_\alpha$ is surjective [EGA IV$_3$ 1966, Theorem 8.10.5] and étale [EGA IV$_4$ 1967, Proposition 17.7.8].

For each index $\lambda \geq \alpha$, let $F_\lambda \subset X_\lambda$ be the subset of all points $s \in X_\lambda$ so that the étale surjection $U_\lambda \otimes \kappa(s) \to \operatorname{Spec} \kappa(s)$ is completely decomposed, that is, admits a section. According to Lemma 13.3 below, this subset is indconstructible. In light of [EGA IV$_4$ 1967, Corollary 8.3.4] it remains to show that the inclusion $\bigcup_\lambda u_\lambda^{-1}(F_\lambda) \subset X$ is an equality, where $u_\lambda : X \to X_\lambda$ denote the projections.

Fix a point $x \in X$, and choose a point $u \in U$ above so that the inclusion $k(x) \subset \kappa(u)$ is an equality. Let $x_\lambda \in X_\lambda$ and $u_\lambda \in U_\lambda$ be the respective image points. Then $U_x = \varprojlim (U_\lambda)_{x_\lambda}$, because inverse limits commute with fiber products. Furthermore, the section of $U_x \to \operatorname{Spec} \kappa(x)$ corresponding to u comes from a section of $(U_\lambda)_{x_\lambda} \to \operatorname{Spec} \kappa(x_\lambda)$, for some index λ, by [EGA IV$_3$ 1966, Theorem 8.8.2]. $\square$

Recall that a subset $F \subset X$ is called *indconstructible* if each point $x \in X$ admits an open neighborhood $x \in U$ so that $F \cap U$ is the union of constructible subsets of U (see [EGA I 1971, Chapter I, Section 7.2 and Chapter 0, Section 2.2]).

Lemma 13.3. *Let $f : U \to X$ be a étale surjection. Then the subset $F \subset X$ of points $x \in X$ for which $U_x \to \operatorname{Spec} \kappa(x)$ admits a section is indconstructible.*

Proof. The problem is local, so we may assume that $X = \mathrm{Spec}(R)$ is affine. Fix a point $x \in X$ so that $U_x \to \mathrm{Spec}\,\kappa(x)$ admits a section, and let $A \subset X$ be its closure. The section extends to a section for $U_A \to A$ over some dense open subset $C \subset A$. Shrinking C if necessary, we may assume that $C = A \cap \mathrm{Spec}(R_g)$ for some $g \in R$. Then $C \subset X$ is constructible and contains x. □

Appendix A: Connected components of schemes

Let X be a topological space and $a \in X$ a point. The corresponding *connected component* $C = C_a$ is the union of all connected subsets containing a. This is then the largest connected subset containing a, and it is closed but not necessarily open. In contrast, the *quasicomponent* $Q = Q_a$ is defined as the intersection of all open-and-closed neighborhoods of a, which is also closed but not necessarily open. Clearly, we have $C \subset Q$. Equality holds if X has only finitely many quasicomponents, or if X is compact or locally connected, but in general there is a strict inclusion.

The simplest example with $C \subsetneq Q$ seems to be the following: Let U be an infinite discrete space, $Y_i = U \cup \{a_i\}$ be two copies of the Alexandroff compactification, and $X = Y_1 \cup Y_2$ the space obtained by gluing along the open subset U, which is quasicompact but not Hausdorff. Then all connected components are singletons, but the discrete space $Q = \{a_1, a_2\}$ is a quasicomponent.

Assume that X is a locally ringed space. Write $R = \Gamma(X, \mathscr{O}_X)$ for the ring of global sections and $\mathfrak{p} \subset R$ for the prime ideal of all global sections vanishing at our point $a \in X$. Obviously, the open-and-closed neighborhoods $U \subset X$ correspond to the idempotent elements $\epsilon \in R \setminus \mathfrak{p}$, via $U = X_\epsilon = \{x \in X \mid \epsilon(x) = 1\}$. Let $L \subset R \setminus \mathfrak{p}$ the multiplicative subsystem of all idempotent elements. We put an order relation by declaring $\epsilon \leq \epsilon'$ if $\epsilon' \mid \epsilon$. Then $\epsilon \mapsto X_\epsilon$ is a filtered inverse system of subspaces, and the quasicomponent is

$$Q = \bigcap_{\epsilon \in L} X_\epsilon = \varprojlim_{\epsilon \in L} X_\epsilon.$$

Now suppose that X is a scheme. Then each inclusion morphism $\iota_\epsilon : X_\epsilon \to X$ is affine, hence $\mathscr{A}_\iota = \iota_{\epsilon *}(\mathscr{O}_{X_\epsilon})$ are quasicoherent, and the same holds for $\mathscr{A} = \varinjlim \mathscr{A}_\epsilon$. In turn, we have $Q = \varprojlim_{\epsilon \in L} X_\epsilon = \mathrm{Spec}(\mathscr{A})$, which puts a scheme structure on the quasicomponent. The quasicoherent ideal sheaf for the closed subscheme $Q \subset X$ is $\mathscr{I} = \bigcup_{\epsilon \in L}(1 - \epsilon)\mathscr{O}_X$. The following observation is due to Ferrand in the affine case, and Lazard in the general case (see [Lazard 1967, Proposition 6.1 and Corollary 8.5]):

Lemma A.1. *Let X be a quasicompact and quasiseparated scheme and $a \in X$ be a point. Then we have an equality $C = Q$ between the connected component and the quasicomponent containing $a \in X$.*

If $X = \mathrm{Spec}(R)$ is affine, and $a \in X$ is the point corresponding to a prime ideal $\mathfrak{p} \subset R$, then the corresponding $C = Q$ is defined by the ideal $\mathfrak{a} = \bigcup Re$, where e runs through the idempotent elements in $\mathfrak{p}$. Clearly, we have $R/\mathfrak{a} = \varinjlim_e A/eA$. The latter description was already given by Artin [1971, p. 292].

Let X_0 be a scheme, and $X = \varprojlim_{\lambda \in L} X_\lambda$ be an inverse system of affine X_0-schemes. The argument for Lemma A.1 essentially depends on [EGA IV$_3$ 1966, Proposition 8.4.1 (ii)]. There, however, it is incorrectly claimed that a sum decomposition of X comes from a sum decomposition of X_λ for sufficiently large $\lambda \in L$, provided that X_0 is merely quasicompact. This only becomes true only under the additional hypothesis of quasiseparatedness:

Proposition A.2. *Let X_0 be a quasicompact and quasiseparated scheme, and X_λ an filtered inverse system of affine X_0-schemes, and $X = \varprojlim X_\lambda$. If $X = X' \cup X''$ is a decomposition into disjoint open subsets, then there is some $\lambda \in L$ and a decomposition $X_\lambda = X'_\lambda \cup X''_\lambda$ into disjoint open subset so that $X', X'' \subset X$ are the respective preimages.*

Proof. This is a special case of [EGA IV$_3$ 1966, Theorem 8.3.11] but I would like to give an alternative proof relying on absolute noetherian approximation rather than ind- and proconstructible sets: Choose a $\mathbb{Z}$-scheme of finite type S so that there is an affine morphism $X_0 \to S$ [Thomason and Trobaugh 1990, Appendix C, Theorem 9]. Replacing X_0 by S, we may assume that X_0 is noetherian. Let $\mathscr{A}_\lambda$ be the quasicoherent $\mathscr{O}_{X_0}$-algebras corresponding to the X_λ, such that X is the relative spectrum of $\mathscr{A} = \varinjlim(\mathscr{A}_\lambda)$. Since the topological space of X_0 is noetherian, the canonical map

$$\varinjlim H^0(X_0, \mathscr{A}_{\mathscr{L}}) \to H^0(X_0, \varinjlim \mathscr{A}_\lambda)$$

is bijective, see [Hartshorne 1977, Chapter III, Proposition 2.9]. Thus each idempotent $e \in \Gamma(X_0, \mathscr{A})$ comes from an element $e_\lambda \in \Gamma(X_0, \mathscr{A}_\lambda)$. Passing to a larger index, we may assume that $e_\lambda^2 = e_\lambda$, such that our element e_λ becomes idempotent. The statement follows from the aforementioned correspondence between idempotent global sections and open-and-closed subsets. $\square$

Here is a counterexamples with X_0 not quasiseparated. Let U be an infinite discrete set, and $X = Y_1 \cup Y_2$ be the gluing of two copies $Y = Y_1 = Y_2$ of the Alexandroff compactification along the open subset $U \subset Y_i$ mentioned above. We may regard the Alexandroff compactification as a profinite space, by choosing a total order on U: Let L be the set of all finite subsets of U, ordered by inclusion. Given $\lambda \in L$, we write $F_\lambda \subset U$ for the corresponding finite subset. For every $\lambda \le \mu$, let $F_\mu \to F_\lambda$ by the retraction of the canonical inclusion $F_\lambda \subset F_\mu$ sending the complementary points to the largest element $f_\lambda \in F_\lambda$. Then one easily sees that $Y = \varprojlim F_\lambda$, where the point at infinity becomes the tuple of largest elements $a = (f_\lambda)_{\lambda \in L}$.

To proceed, fix a ground field k. Then we may endow the profinite space Y with the structure of an affine k-scheme, by regarding F_λ as the spectrum of the k-algebra $R_\lambda = \mathrm{Hom}_{\mathrm{Set}}(F_\lambda, k)$. In turn, we regard the gluing $X = Y_1 \cup Y_2$ as a k-scheme. This scheme is quasicompact, but not quasiseparated, because the intersection $Y_1 \cap Y_2 = U$ is infinite and discrete. Consider the subset $V_\lambda = X \setminus F_\lambda$, which is open-and-closed. Moreover, the inclusion morphism $V_\lambda \to X$ is affine, and $\varprojlim_{\lambda \in L} V_\lambda = Q = \{a_1, a_2\}$ is the disconnected quasicomponent. The sum decomposition of the quasicomponent Q, however, does not come from a sum decomposition of any V_λ.

Here is a counterexample with X_0 not quasicompact. Let X_0 be the disjoint union of two infinite chains $C' = \bigcup C'_n$ and $C'' = \bigcup C''_n$, $n \in \mathbb{Z}$, of copies of the projective line over the ground field k, together with further copies B_n of the projective line joining the intersection $C'_{n-1} \cap C'_n$ with $C''_{n-1} \cap C''_n$. Let L be the collection of all finite subset of $\mathbb{Z}$, ordered by inclusion. Consider the inverse system of closed subsets $X_\lambda \subset X$ consisting of $C' \cup C''$ and the union $\bigcup_{n \notin \lambda} B_n$. Then all X_λ are connected, but $X = \varprojlim(X_\lambda) = \bigcap X_\lambda = C' \amalg C''$ becomes disconnected.

Appendix B: Čech and sheaf cohomology

Here we briefly recall H. Cartan's criterion for equality for Čech and sheaf cohomology, which is described in Godement's monograph [1958, Chapter II, Theorem 5.9.2]. Throughout we work in the context of topoi and sites. The general reference is [SGA 4_1 1972]. Suppose $\mathscr{E}$ is a topos. Choose a site $\mathscr{C}$ and an equivalence $\mathrm{Sh}(\mathscr{C}) = \mathscr{E}$, and make the choice so that the category $\mathscr{C}$ has a terminal object $X \in \mathscr{C}$ and all fiber products, and so that the Grothendieck topology is given in terms of a pretopology of coverings. Suppose we additionally have a subcategory $\mathscr{B} \subset \mathscr{C}$ such that every object $U \in \mathscr{C}$ admits a covering $(U_\alpha \to U)_\alpha$ with $U_\alpha \in \mathscr{B}$, and that $\mathscr{B}$ has all fiber products. Then the subcategory inherits a pretopology, and the restriction functor of sheaves on $\mathscr{C}$ to sheaves on $\mathscr{B}$ is an equivalence.

For each abelian sheaf F on $\mathscr{C}$ and each object $U \in \mathscr{C}$, we write $H^p(U, F)$ for the cohomology groups in the sense of derived functors, and

$$\check{H}^p(U, F) = \varinjlim \check{H}^p(\mathfrak{U}, \check{H}^q(F)),$$

for the Čech cohomology groups, where the direct limit runs over all coverings $\mathfrak{U} = (U_\alpha \to U)_\alpha$, and the transition maps are defined as in [Godement 1958, Chapter II, Section 5.7]. Note that me may restrict to those coverings with all $U_\alpha \in \mathscr{B}$.

Let F be an abelian sheaf on the site $\mathscr{C}$. We denote by $\underline{H}^q(F)$ the presheaf $U \mapsto H^q(U, F)$. The functors $F \mapsto \underline{H}^p(F)$, $p \geq 0$, form a δ-functor from the category of abelian sheaves to the category of abelian presheaves. This δ-functor is universal because it vanishes on injective objects for $p \geq 1$.

The inclusion functor $F \mapsto \underline{H}^0(F)$ from the category of abelian sheaves to the category of abelian presheaves is right adjoint to the sheafification functor. This adjointness ensures that $F \mapsto \underline{H}^0(F)$ sends injective objects to injective objects. Furthermore, $F \mapsto \check{H}^p(U, F)$, $p \geq 0$, form a universal δ-functor from the category of abelian presheaves on U to the category of abelian groups, see [Tamme 1994, Theorem 2.2.6]. In turn, we get a Grothendieck spectral sequence

$$\check{H}^p(U, \underline{H}^q(F)) \Rightarrow H^{p+q}(U, F), \tag{1}$$

which we call a *Čech-to-sheaf-cohomology spectral sequence*.

Theorem B.1 (H. Cartan). *The following are equivalent*:

(i) $H^q(U, F) = 0$ *for all* $U \in \mathscr{B}$ *and* $q \geq 1$.

(ii) $\check{H}^q(U, F) = 0$ *for all* $U \in \mathscr{B}$ *and* $q \geq 1$.

If these equivalent conditions are satisfied, then $\check{H}^p(X, \underline{H}^q(F)) = 0$ *for all* $p \geq 0$ *and* $q \geq 1$, *and the edge maps* $\check{H}^p(X, F) \to H^p(X, F)$ *are bijective for all* $p \geq 0$.

Proof. Suppose (i) holds. Then $\underline{H}^q(F) = 0$ for all $q \geq 1$. This ensures that $\check{H}^p(V, \underline{H}^q(F)) = 0$ for all $V \in \mathscr{C}$. In turn, the edge map $\check{H}^p(V, F) \to H^p(U, F)$ in the Čech-to-sheaf-cohomology spectral sequence (1) is bijective. For $V = U \in \mathscr{B}$, this gives (ii), while for $V = X$ we get the amendment of the statement.

Conversely, suppose that (ii) holds. We inductively show that $H^p(U, F) = 0$, or equivalently that the edge maps $\check{H}^p(U, F) \to H^p(U, F)$ are bijective, for all $p \geq 1$ and all $U \in \mathscr{B}$. First note that $\check{H}^0(U, \underline{H}^r(F)) = 0$ for all $r \geq 1$, because the sheafification of $\underline{H}^r(F)$ vanishes. Thus the case $p = 1$ in the induction follows from the short exact sequence

$$0 \to \check{H}^1(U, F) \to H^1(U, F) \to \check{H}^0(U, \underline{H}^1(F)).$$

Now let $p \geq 1$ be arbitrary, and suppose that $H^p(U, F) = 0$ for $1, \ldots, p-1$ and all $U \in \mathscr{B}$. The argument in the previous paragraph gives $\check{H}^r(U, \underline{H}^s(F)) = 0$ for all $0 < s < p$ and $r \geq 1$, and we already noted that $\check{H}^0(U, \underline{H}^p(F)) = 0$. Consequently, the edge map $\check{H}^p(U, F) \to H^p(U, F)$ must be bijective. □

Appendix C: Inductive dimension in general topology

The proof of the crucial Proposition 4.5 depends on results from dimension theory that are perhaps not so well-known outside general topology. Here we briefly review the relevant material. For a comprehensive treatment see, for example, Pears' monograph [1975, Chapter 4].

Let X be a topological space. There are several ways to obtain suitable notions of dimension that are meaningful for compact spaces. One of them is the *large inductive dimension* $\operatorname{Ind}(X) \geq -1$, which goes back to Brouwer [1913]. Here one first defines $\operatorname{Ind}(X) \leq n$ as a property of topological spaces by induction as follows: The induction starts with $n = -1$, where $\operatorname{Ind}(X) \leq -1$ simply means that X is empty. Now suppose that $n \geq 0$, and that the property is already defined for $n - 1$. Then $\operatorname{Ind}(X) \leq n$ means that for all closed subsets $A \subset X$ and all open neighborhoods $A \subset V$, there is a smaller open neighborhood $A \subset U \subset V$ whose boundary $\Phi = \overline{U} \setminus U$ has $\operatorname{Ind}(\Phi) \leq n-1$. Now one defines $\operatorname{Ind}(X) = n$ as the least integer $n \geq -1$ for which $\operatorname{Ind}(X) \leq n$ holds, or $n = \infty$ if no such integer exists.

A variant is the *small inductive dimension* $\operatorname{ind}(X) \geq -1$, which is defined in an analogous way, but with points $a \in X$ instead of closed subsets $A \subset X$. If every point $a \in X$ is closed, an easy induction gives

$$\operatorname{ind}(X) \leq \operatorname{Ind}(X).$$

There are, however, compact spaces with strict inequality [Pears 1975, Chapter 8, §3]. Moreover, little of use can be said if the space contains nonclosed points. For example, a finite linearly ordered space $X = \{x_0, \dots, x_n\}$ obviously has $\operatorname{Ind}(S) = 0$ and $\operatorname{ind}(S) = n$. Note that such spaces arise as spectra of valuation rings R with Krull dimension $\dim(R) = n$.

Recall that a space X is called *at most zero-dimensional* if its topology admits a basis consisting of open-and-closed subsets $U \subset X$. It is immediate that this is equivalent to $\operatorname{ind}(X) \leq 0$. In fact, both small and large inductive dimension can be seen as generalizations of this concept [Pears 1975, Chapter 4, Proposition 1.1 and Corollary 2.2], at least for compact spaces:

Proposition C.1. *If X is compact, then the following conditions are equivalent*:

(i) *X is at most zero-dimensional.*

(ii) $\operatorname{ind}(X) \leq 0$.

(iii) $\operatorname{Ind}(X) \leq 0$.

There are several *sum theorems* which express the large inductive dimension of a covering $X = \bigcup A_\lambda$ in terms of the large inductive dimension of the subspaces $A_\lambda \subset X$. The following sum theorem is very useful for us; its proof is tricky, but only relies only basic notations of set-theoretical topology [Pears 1975, Chapter 4, Proposition 4.13].

Proposition C.2. *Suppose that X is compact, and $X = A \cup B$ is a closed covering. If the intersection $A \cap B$ is at most zero-dimensional, and the subspaces have* $\operatorname{Ind}(A), \operatorname{Ind}(B) \leq n$, *then* $\operatorname{Ind}(X) \leq n$.

From this we deduce the following fact, which, applied to $X_i = \operatorname{Max}(Y_i)$, enters in the proof of Proposition 4.5.

Proposition C.3. *Suppose that X is compact, and that $X = X_1 \cup \cdots \cup X_n$ is a closed covering, where the X_i are at most zero-dimensional. Then X is at most zero-dimensional.*

Proof. By induction it suffices to treat the case $n = 2$. Since X_i are at most zero-dimensional, the same holds for the subspace $X_1 \cap X_2 \subset X_i$. The spaces X_i are compact, because they are closed inside the compact space X. According to Proposition C.1, we have $\mathrm{Ind}(X_i) \leq 0$. Proposition C.2 yields $\mathrm{Ind}(X) \leq 0$, and Proposition C.1 again tells us that X is at most zero-dimensional. $\square$

Acknowledgments

I wish to thank Johan de Jong for inspiring conversations, and also the referee for very useful comments, which helped to improve the exposition, clarify the paper and remove mistakes.

References

[Aberbach 2005] I. M. Aberbach, "The vanishing of $\mathrm{Tor}_1^R(R^+, k)$ implies that R is regular", *Proc. Amer. Math. Soc.* **133**:1 (2005), 27–29. MR Zbl

[Artin 1970] M. Artin, "Algebraization of formal moduli. II. Existence of modifications", *Ann. of Math.* (2) **91** (1970), 88–135. MR Zbl

[Artin 1971] M. Artin, "On the joins of Hensel rings", *Advances in Math.* **7** (1971), 282–296. MR Zbl

[Benoist 2013] O. Benoist, "Quasi-projectivity of normal varieties", *Int. Math. Res. Not.* **2013**:17 (2013), 3878–3885. MR Zbl

[Bhatt 2014] B. Bhatt, "A flat map that is not a directed limit of finitely presented flat maps", preprint, University of Michigan, 2014, Available at http://www-personal.umich.edu/~bhattb/math/flat-maps-limits-fp.pdf.

[Bhatt and de Jong 2014] B. Bhatt and A. J. de Jong, "Lefschetz for local Picard groups", *Ann. Sci. Éc. Norm. Supér.* (4) **47**:4 (2014), 833–849. MR Zbl

[Bhatt and Scholze 2015] B. Bhatt and P. Scholze, "The pro-étale topology for schemes", pp. 99–201 in *De la géométrie algébrique aux formes automorphes*, edited by J.-B. Bost et al., Astérisque **369**, Société Mathématique de France, Paris, 2015. MR Zbl

[Borho and Weber 1971] W. Borho and V. Weber, "Zur Existenz total ganz abgeschlossener Ringerweiterungen", *Math. Z.* **121** (1971), 55–57. MR Zbl

[Bourbaki 1971] N. Bourbaki, *Éléments de mathématique. Topologie générale. Chapitres 1 à 4*, Hermann, Paris, 1971. MR Zbl

[Brouwer 1913] L. E. J. Brouwer, "Über den natürlichen Dimensionsbegriff", *J. Reine Angew. Math.* **142** (1913), 146–152. MR Zbl

[Carral and Coste 1983] M. Carral and M. Coste, "Normal spectral spaces and their dimensions", *J. Pure Appl. Algebra* **30**:3 (1983), 227–235. MR Zbl

[Conrad 2007] B. Conrad, "Deligne's notes on Nagata compactifications", *J. Ramanujan Math. Soc.* **22**:3 (2007), 205–257. Erratum in J. Ramanujan Math. Soc. 24 (2009), 427–428. MR Zbl

[De Marco and Orsatti 1971] G. De Marco and A. Orsatti, "Commutative rings in which every prime ideal is contained in a unique maximal ideal", *Proc. Amer. Math. Soc.* **30** (1971), 459–466. MR Zbl

[Deligne 2010] P. Deligne, "Le théorème de plongement de Nagata", *Kyoto J. Math.* **50**:4 (2010), 661–670. MR Zbl

[EGA I 1971] A. Grothendieck and J. A. Dieudonné, *Eléments de géométrie algébrique, I: Le lange des schémas*, Grundlehren der Math. Wissenschaften **166**, Springer, Berlin, 1971. MR Zbl

[EGA II 1961] A. Grothendieck, "Eléments de géométrie algébrique, II: Étude globale élémentaire de quelques classes de morphismes", *Inst. Hautes Études Sci. Publ. Math.* **8** (1961), 5–222. MR Zbl

[EGA IV_1 1964] A. Grothendieck and J. A. Dieudonné, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, I", *Inst. Hautes Études Sci. Publ. Math.* **20** (1964), 5–259. MR Zbl

[EGA IV_3 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

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

[Endler 1972] O. Endler, *Valuation theory*, Springer, 1972. MR Zbl

[Enochs 1968] E. Enochs, "Totally integrally closed rings", *Proc. Amer. Math. Soc.* **19** (1968), 701–706. MR Zbl

[Ferrand 2003] D. Ferrand, "Conducteur, descente et pincement", *Bull. Soc. Math. France* **131**:4 (2003), 553–585. MR Zbl

[Flenner et al. 1999] H. Flenner, L. O'Carroll, and W. Vogel, *Joins and intersections*, Springer Monographs in Mathematics, Springer, 1999. MR Zbl

[Gabber 1994] O. Gabber, "Affine analog of the proper base change theorem", *Israel J. Math.* **87**:1-3 (1994), 325–335. MR Zbl

[Gillman and Jerison 1960] L. Gillman and M. Jerison, *Rings of continuous functions*, The University Series in Higher Mathematics, D. Van Nostrand Co., Princeton, N.J., 1960. MR Zbl

[Godement 1958] R. Godement, *Topologie algébrique et théorie des faisceaux*, Actualit'es Sci. Ind. No. 1252. Publ. Math. Univ. Strasbourg. No. 13, Hermann, Paris, 1958. MR Zbl

[Gross 2012] P. Gross, "The resolution property of algebraic surfaces", *Compos. Math.* **148**:1 (2012), 209–226. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977. MR Zbl

[Hochster 1970] M. Hochster, "Totally integrally closed rings and extremal spaces", *Pacific J. Math.* **32** (1970), 767–779. MR Zbl

[Hochster and Huneke 1992] M. Hochster and C. Huneke, "Infinite integral extensions and big Cohen–Macaulay algebras", *Ann. of Math.* (2) **135**:1 (1992), 53–89. MR Zbl

[Horrocks 1971] G. Horrocks, "Birationally ruled surfaces without embeddings in regular schemes", *Bull. London Math. Soc.* **3** (1971), 57–60. MR Zbl

[Huneke 2011] C. Huneke, "Absolute integral closure", pp. 119–135 in *Commutative algebra and its connections to geometry*, edited by A. Corso and C. Polini, Contemp. Math. **555**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Huneke and Lyubeznik 2007] C. Huneke and G. Lyubeznik, "Absolute integral closure in positive characteristic", *Adv. Math.* **210**:2 (2007), 498–504. MR Zbl

[Johnstone 1977] P. T. Johnstone, *Topos theory*, Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1977. MR Zbl

[Kleiman 1966] S. L. Kleiman, "Toward a numerical theory of ampleness", *Ann. of Math.* (2) **84** (1966), 293–344. MR Zbl

[Laumon and Moret-Bailly 2000] G. Laumon and L. Moret-Bailly, *Champs algébriques*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics] **39**, Springer, 2000. MR

[Lazard 1967] D. Lazard, "Disconnexités des spectres d'anneaux et des préschémas", *Bull. Soc. Math. France* **95** (1967), 95–108. MR Zbl

[Lenstra 1985] H. Lenstra, "Galois theory for schemes", Lecture notes, Universiteit Leiden, 1985, Available at http://websites.math.leidenuniv.nl/algebra/GSchemes.pdf. Zbl

[Mazza et al. 2006] C. Mazza, V. Voevodsky, and C. Weibel, *Lecture notes on motivic cohomology*, Clay Mathematics Monographs **2**, American Mathematical Society, Providence, RI, 2006. MR Zbl

[Milne 1980] J. S. Milne, *Étale cohomology*, Princeton Mathematical Series **33**, Princeton University Press, 1980. MR Zbl

[Mumford 1961] D. Mumford, "The topology of normal singularities of an algebraic surface and a criterion for simplicity", *Inst. Hautes Études Sci. Publ. Math.* **9**:9 (1961), 5–22. MR Zbl

[Nicholson and Zhou 2005] W. K. Nicholson and Y. Zhou, "Clean rings: a survey", pp. 181–198 in *Advances in ring theory*, edited by J. Chen et al., World Sci. Publ., Hackensack, NJ, 2005. MR Zbl

[Nisnevich 1989] Y. A. Nisnevich, "The completely decomposed topology on schemes and associated descent spectral sequences in algebraic K-theory", pp. 241–342 in *Algebraic K-theory: connections with geometry and topology* (Lake Louise, AB, 1987), edited by J. F. Jardine and V. P. Snaith, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **279**, Kluwer Acad. Publ., Dordrecht, 1989. MR Zbl

[Pears 1975] A. R. Pears, *Dimension theory of general spaces*, Cambridge University Press, 1975. MR Zbl

[Perling and Schröer 2014] M. Perling and S. Schröer, "Vector bundles on proper toric 3-folds and certain other schemes", preprint, 2014. To appear in *Trans. Amer. Math. Soc.* arXiv

[Rydh 2010] D. Rydh, "Submersions and effective descent of étale morphisms", *Bull. Soc. Math. France* **138**:2 (2010), 181–230. MR Zbl

[Rydh 2015] D. Rydh, "Noetherian approximation of algebraic spaces and stacks", *J. Algebra* **422** (2015), 105–147. MR Zbl

[Schoutens 2004] H. Schoutens, "On the vanishing of Tor of the absolute integral closure", *J. Algebra* **275**:2 (2004), 567–574. MR Zbl

[Schröer 1999] S. Schröer, "On non-projective normal surfaces", *Manuscripta Math.* **100**:3 (1999), 317–321. MR Zbl

[Schröer 2014] S. Schröer, "Points in the fppf topology", preprint, 2014. To appear in *Ann. Sc. Norm. Super. Pisa Cl. Sci.* arXiv

[Schwartz 2011] N. Schwartz, "Graph components of prime spectra", *J. Algebra* **337** (2011), 13–49. MR Zbl

[Sergio 1982] I. Sergio, "Rings whose spectra do not have polygons", *Matematiche* (*Catania*) **37**:1 (1982), 62–71. MR Zbl

[SGA 1 1971] A. Grothendieck, *Revêtements étales et groupe fondamental* (Séminaire de Géométrie Algébrique du Bois Marie 1960–1961), Lecture Notes in Math. **224**, Springer, Berlin, 1971. MR Zbl

[SGA 2 1968] A. Grothendieck, *Cohomologie locale des faisceaux cohérents et théorèmes de Lefschetz locaux et globaux* (Séminaire de Géométrie Algébrique du Bois Marie 1962), Advanced Stud. in Pure Math. **2**, North-Holland, Amsterdam, 1968. MR Zbl

[SGA 4_1 1972] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 1: Théorie des topos, Exposés I–IV* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **269**, Springer, Berlin, 1972. MR Zbl

[SGA 4_2 1972] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 2: Exposés V–VIII* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **270**, Springer, Berlin, 1972. MR Zbl

[Stack Project Authors 2005–] T. Stack Project Authors, "The Stacks project", electronic reference, 2005–, Available at http://stacks.math.columbia.edu.

[Tamme 1994] G. Tamme, *Introduction to étale cohomology*, Universitext, Springer, 1994. MR Zbl

[Thomason and Trobaugh 1990] R. W. Thomason and T. Trobaugh, "Higher algebraic K-theory of schemes and of derived categories", pp. 247–435 in *The Grothendieck Festschrift, Vol. III*, edited by P. Cartier et al., Progr. Math. **88**, Birkhäuser, Boston, 1990. MR Zbl

[Vakil and Wickelgren 2011] R. Vakil and K. Wickelgren, "Universal covering spaces and fundamental groups in algebraic geometry as schemes", *J. Théor. Nombres Bordeaux* **23**:2 (2011), 489–526. MR Zbl

[Voevodsky et al. 2000] V. Voevodsky, A. Suslin, and E. M. Friedlander, *Cycles, transfers, and motivic homology theories*, Annals of Mathematics Studies **143**, Princeton University Press, 2000. MR Zbl

Communicated by Ravi Vakil
Received 2015-03-24 Revised 2016-10-18 Accepted 2016-11-18

schroeer@math.uni-duesseldorf.de
Mathematisches Institut, Heinrich-Heine-Universität, 40204, Düsseldorf, Germany,

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.583

The sixth moment of automorphic L-functions

Vorrapan Chandee and Xiannan Li

We consider the L-functions $L(s, f)$, where f is an eigenform for the congruence subgroup $\Gamma_1(q)$. We prove an asymptotic formula for the sixth moment of this family of automorphic L-functions.

1. Introduction

Moments of L-functions are of great interest to analytic number theorists. For instance, for $\zeta(s)$ denoting the Riemann zeta function and

$$I_k(T) := \int_0^T \left|\zeta\left(\tfrac{1}{2}+it\right)\right|^{2k} dt,$$

asymptotic formulae were proven for $k = 1$ by Hardy and Littlewood and for $k = 2$ by Ingham; see [Titchmarsh 1986, Chapter VII]. This work is closely related to zero density results and the distribution of primes in short intervals. More recently, moments of other families of L-functions have been studied for their numerous applications, including nonvanishing and subconvexity results. In many applications, it is important to develop technology which can understand such moments for larger k.

The behavior of moments for larger k remain mysterious. However, recently there has been great progress in our understanding. First, good heuristics and conjectures on the behavior of $I_k(T)$ have appeared in the literature. To be precise, a folklore conjecture states that

$$I_k(T) \sim c_k T (\log T)^{k^2}$$

for constants c_k depending on k, but the values of c_k were unknown for general k until the work of Keating and Snaith [2000], which related these moments to circular unitary ensembles and provided precise conjectures for c_k. The choice of group is consistent with the Katz–Sarnak philosophy [Katz and Sarnak 1999], which indicates that the symmetry group associated to this family should be unitary.

MSC2010: primary 11M41; secondary 11F11, 11F67.
Keywords: moment of L-functions, automorphic L-functions, $\Gamma_1(q)$.

Based on heuristics for shifted divisor sums, Conrey and Ghosh [1998] derived a conjecture in the case $k = 3$ and Conrey and Gonek [2001] derived a conjecture in the case $k = 4$. In particular, the conjecture for the sixth moment is

$$I_3(T) \sim 42a_3 \frac{T(\log T)^9}{9!}$$

for some arithmetic factor a_3. Further conjectures, including lower order terms and cases of other symmetry groups, are available from the work of Conrey, Farmer, Keating, Rubinstein and Snaith [Conrey et al. 2005] as well as from the work of Diaconu, Goldfeld and Hoffstein [Diaconu et al. 2003].

In support of these conjectures, lower bounds of the right order of magnitude are available due to Rudnick and Soundararajan [2005], while good upper bounds of the right order of magnitude are available, conditionally on the Riemann hypothesis, due to Soundararajan [2009] and later improved by Harper [2013].

Despite this, verifications of the moment conjectures for high moments remain elusive. Typically, even going slightly beyong the fourth moment to obtain a twisted fourth moment is quite difficult, and there are few families for which this is known.

Quite recently, Conrey, Iwaniec and Soundararajan [Conrey et al. 2012] derived an asymptotic formula for the sixth moment of Dirichlet L-functions with a power saving error term. Instead of fixing the modulus q and only averaging over primitive characters $\chi \pmod q$, they also averaged over the modulus $q \le Q$, giving them a larger family of size Q^2. Further, they included a short average on the critical line. In particular, they showed that

$$\sum_{q\le Q} \sideset{}{^*}\sum_{\chi \pmod q} \int_{-\infty}^{\infty} \big|L\big(\tfrac12+it,\chi\big)\big|^6 \big|\Gamma\big(\tfrac14+\tfrac12 it\big)\big|^6\,dt \sim 42b_3 \frac{Q^2(\log Q)^9}{9!} \int_{-\infty}^{\infty} \big|\Gamma\big(\tfrac14+\tfrac12 it\big)\big|^6\,dt$$

for some constant b_3. This is consistent with the analogous conjecture for the Riemann zeta function above.

The authors of this paper subsequently derived an asymptotic formula for the eight moment of this family of L-functions, conditionally on the generalized Riemann hypothesis [Chandee and Li 2014], namely

$$\sum_{q\le Q} \sideset{}{^*}\sum_{\chi \pmod q} \int_{-\infty}^{\infty} \big|L\big(\tfrac12+it,\chi\big)\big|^8 \big|\Gamma\big(\tfrac14+\tfrac12 it\big)\big|^8\,dt \sim 24024b_4 \frac{Q^2(\log Q)^{16}}{16!} \int_{-\infty}^{\infty} \big|\Gamma\big(\tfrac14+\tfrac12 it\big)\big|^8\,dt$$

for some constant b_4.

In this paper, we study a family of L-functions attached to automorphic forms on GL(2). To be more precise, let $S_k(\Gamma_0(q), \chi)$ be the space of cusp forms of weight

$k \geq 2$ for the group $\Gamma_0(q)$ and the nebentypus character $\chi \pmod q$, where

$$\Gamma_0(q) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, ad - bc = 1,\ c \equiv 0 \pmod q \right\}.$$

Also, let $S_k(\Gamma_1(q))$ be the space of holomorphic cusp forms for the group

$$\Gamma_1(q) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, ad - bc = 1,\ c \equiv 0 \pmod q,\ a \equiv d \equiv 1 \pmod q \right\}.$$

Note that $S_k(\Gamma_1(q))$ is a Hilbert space with the Petersson's inner product

$$\langle f, g \rangle = \int_{\Gamma_1(q)\backslash\mathbb{H}} f(z)\bar{g}(z) y^{k-2}\, dx\, dy,$$

and

$$S_k(\Gamma_1(q)) = \bigoplus_{\chi \pmod q} S_k(\Gamma_0(q), \chi).$$

Let $\mathcal{H}_\chi \subset S_k(\Gamma_0(q), \chi)$ be an orthogonal basis of $S_k(\Gamma_0(q), \chi)$ consisting of Hecke cusp forms, normalized so that the first Fourier coefficient is 1. For each $f \in \mathcal{H}_\chi$, we let $L(f, s)$ be the L-function associated to f, defined for $\mathrm{Re}(s) > 1$ as

$$L(f, s) = \sum_{n \geq 1} \frac{\lambda_f(n)}{n^s} = \prod_p \left(1 - \frac{\lambda_f(p)}{p^s} + \frac{\chi(p)}{p^{2s}}\right)^{-1}, \tag{1-1}$$

where $\{\lambda_f(n)\}$ are the Hecke eigenvalues of f. With our normalization, $\lambda_f(1) = 1$. In general, the Hecke eigenvalues satisfy the Hecke relation

$$\lambda_f(m)\lambda_f(n) = \sum_{d \mid (m,n)} \chi(d) \lambda_f\left(\frac{mn}{d^2}\right), \tag{1-2}$$

for all $m, n \geq 1$. We define the completed L-function as

$$\Lambda\left(f, \tfrac{1}{2} + s\right) = \left(\frac{q}{4\pi^2}\right)^{s/2} \Gamma\left(s + \tfrac{1}{2}k\right) L\left(f, \tfrac{1}{2} + s\right), \tag{1-3}$$

which satisfies the functional equation

$$\Lambda\left(f, \tfrac{1}{2} + s\right) = i^k \bar{\eta}_f \Lambda\left(\bar{f}, \tfrac{1}{2} - s\right),$$

where $|\eta_f| = 1$ when f is a newform.

Suppose for each $f \in \mathcal{H}_\chi$ we have an associated number α_f. Then we define the harmonic average of α_f over $\mathcal{H}_\chi$ to be

$$\sum_{f \in \mathcal{H}_\chi}^{h} \alpha_f = \frac{\Gamma(k-1)}{(4\pi)^{k-1}} \sum_{f \in \mathcal{H}_\chi} \frac{\alpha_f}{\|f\|^2}.$$

Note that when the first coefficient $\lambda_f(1) = 1$, $\|f\|^2$ is essentially the value of a certain L-function at 1, and so on average, $\|f\|^2$ is constant. As in other works, it is possible to remove the weighting by $\|f\|^2$ through what is now a standard argument.

We shall be interested in moments of the form

$$\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\big|L\big(f,\tfrac12\big)\big|^{2k}.$$

We note that the size of the family is around q^2. For prime level, η_f can be expressed in terms of Gauss sums, and in particular we expect η_f to equidistribute on the circle as f varies over an orthogonal basis of $S_k(\Gamma_1(q))$. Thus, we expect our family of L-functions to be unitary.

In this paper, we prove an asymptotic formula for the sixth moment; this will be the first time that the sixth moment of a family of L-functions over GL(2) has been understood. Following [Conrey et al. 2005], we have the following conjecture for the sixth moment of our family. We refer the reader to Appendix A for a brief derivation of the arithmetic factor in the conjecture.

Conjecture 1.1. Let q be a prime number. As $q\to\infty$, we have

$$\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\big|L\big(f,\tfrac12\big)\big|^{6}\sim 42\mathscr{C}_3\Big(1-\frac1q\Big)^4\Big(1+\frac4q+\frac1{q^2}\Big)C_q^{-1}\frac{(\log q)^9}{9!},$$

where

$$\begin{aligned}\mathscr{C}_3&:=\prod_p C_p,\\ C_p&:=\Big(1+\frac4p+\frac7{p^2}-\frac2{p^3}+\frac9{p^4}-\frac{16}{p^5}+\frac1{p^6}-\frac4{p^7}\Big)\Big(1+\frac1p\Big)^{-4}.\end{aligned}\tag{1-4}$$

Iwaniec and Xiaoqing Li [2007] proved a large sieve result for this family, and Djankovic [2011] used their result to prove, for an odd integer $k\ge 3$ and prime q, that

$$\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=-1}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\big|L\big(f,\tfrac12\big)\big|^{6}\ll q^{\varepsilon}.$$

as $q\to\infty$. In this paper, we shall prove the following.

Theorem 1.2. *Let q be a prime and $k\ge 5$ be odd. Then, as $q\to\infty$, letting $\mathscr{C}_3$ and C_p be as defined in (1-4), we have*

$$\begin{aligned}\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=-1}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\int_{-\infty}^{\infty}\big|\Lambda\big(f,\tfrac12+it\big)\big|^6\,dt\\ \sim 42\mathscr{C}_3\Big(1-\frac1q\Big)^4\Big(1+\frac4q+\frac1{q^2}\Big)C_q^{-1}\frac{(\log q)^9}{9!}\int_{-\infty}^{\infty}\big|\Gamma\big(\tfrac12k+it\big)\big|^6\,dt.\end{aligned}$$

In fact, we are able to prove this with an error term of $q^{-1/4}$, as opposed to the $q^{-1/10}$ error term in [Conrey et al. 2012]. The reason behind this superior error term is explained in the outline in Section 1A. In future work, we hope to extend our attention to the eighth moment.

The assumption that k is odd implies that all $f \in \mathcal{H}_\chi$ are newforms. This is for convenience only and is not difficult to remove. Indeed, when k is even, all $f \in \mathcal{H}_\chi$ are newforms except possibly when χ is the principal character and f is induced by a cusp form of full level. We avoid this case for the sake of brevity. Similarly, the assumption that $k \geq 5$ simplifies parts of the calculation; it is possible to prove Theorem 1.2 for smaller k.

Since the Γ function decays rapidly on vertical lines, the average over t is fairly short. It is included for the same reason as in [Conrey et al. 2012; Chandee and Li 2014]: it allows us to avoid certain unbalanced sums in the computation of the moment. Although this appears to be a small technical change in the main statement, evaluating such moments without the short integration over t is a significant challenge. Our theorem follows from the more general Theorem 2.5 for shifted moments in Section 2.

1A. ***Outline of the paper.*** To help orient the reader, we provide a sketch of the proof, and introduce the various sections of the paper. After applying the approximate functional equation developed in Section 3, the main object to be understood is roughly of the form

$$\frac{2}{\phi(q)} \sum_{\substack{\chi \ (\mathrm{mod}\ q) \\ \chi(-1)=(-1)^k}} \sum_{f \in \mathcal{H}_\chi}^{h} \sum_{m,n \asymp q^{3/2}} \frac{\sigma_3(m)\sigma_3(n)\lambda_f(n)\overline{\lambda_f}(m)}{\sqrt{mn}}.$$

In fact, since the coefficients $\lambda_f(n)$ are not completely multiplicative, the expression is significantly more complicated for the purpose of extracting main terms.

Applying Petersson's formula for the average over $f \in \mathcal{H}_\chi$ leads to diagonal terms $m = n$ which are evaluated fairly easily in Section 4A, as well as off-diagonal terms which involve sums of the form

$$\sum_{m,n \asymp q^{3/2}} \frac{\sigma_3(m)\sigma_3(n)}{\sqrt{mn}} \frac{2}{\phi(q)} \sum_{\substack{\chi \ (\mathrm{mod}\ q) \\ \chi(-1)=(-1)^k}} \sum_{c} S_\chi(m,n;cq) J_{k-1}\left(\frac{4\pi\sqrt{mn}}{cq}\right),$$

where $S_\chi(m,n;cq)$ is the Kloosterman sum defined in Lemma 2.3, and $J_{k-1}(x)$ is the J-Bessel function of order $k-1$.

Let us focus on the transition region for the Bessel function where $c \asymp q^{1/2}$, so that the conductor is a priori of size $qc \asymp q^{3/2}$. It is here that the addition average over

$\chi \pmod q$ comes into play. To be more precise, to understand the exponential sum

$$\frac{2}{\phi(q)} \sum_{\substack{\chi \pmod q \\ \chi(-1)=(-1)^k}} S_\chi(m, n; cq),$$

it suffices to understand

$$\sideset{}{^*}\sum_{\substack{a \pmod{cq} \\ a \equiv 1 \pmod q}} \mathrm{e}\left(\frac{am + \bar{a}n}{cq}\right),$$

which, assuming that $(c, q) = 1$ and using the Chinese remainder theorem and reciprocity, is

$$\mathrm{e}\left(\frac{m+n}{cq}\right) \sideset{}{^*}\sum_{a \pmod c} \mathrm{e}\left(\frac{\bar{q}(a-1)m + \bar{q}(\bar{a}-1)n}{c}\right).$$

Note that $\mathrm{e}(x) = \exp(2\pi i x)$. The factor $\mathrm{e}\left(\frac{m+n}{cq}\right)$ has small derivatives and may be treated as a smooth function, while the conductor of the rest of the exponential sum has decreased to $c \asymp q^{1/2}$. The details of these calculations are in Section 5.

This phenomenon of the drop in conductor appears in other examples. In the case of the sixth moment of Dirichlet L-functions in [Conrey et al. 2012], it occurs when replacing q with the complementary divisor $(m-n)/q \asymp q^{1/2}$. It is quite interesting that the same drop in conductor occurs by seemingly very different mechanisms. However, note that when the complementary divisor is small, the ordered pair (m, n) is forced to be in a narrow region. That this does not occur in our case is one of the reasons behind the superior error term in our result.

After the conductor drop, we apply Voronoi summation to the sum over m and n in Section 6. We need a version of Voronoi summation including shifts. The proof of this is essentially the same as the proof of the standard Voronoi summation formula for $\sigma_3(n)$ by Ivić [1997]. We state the result required in Appendix B.

After applying Voronoi, it is easy to guess which terms should contribute to the main terms and which terms should be error terms. The main terms are described in Proposition 6.1 and the error terms are bounded in Proposition 6.2. Essentially, we expect the main terms to be a sum of products of 9 factors of ζ, the same as the diagonal contribution but with permutations in the shifts, as in Theorem 2.5. This is by no means immediately visible from the expression in Proposition 6.1. Indeed, it takes some effort to see that we get the right number of ζ factors. Along the way, we use, among other things, a calculation of Iwaniec and Xiaoqing Li in [Iwaniec and Li 2007]. This is done in Section 7. In order to finish the verifications, we need to check that the local factors of two expressions agree. The details here are standard but intricate, and are provided in Appendix A.

Finally, the error terms from Voronoi summation are bounded in Section 8. Here, one needs to show that the dual sums from Voronoi summation are essentially quite short, which is related to the reduction in conductor from cq to c earlier.

2. Notation and the shifted sixth moment

We begin with some notation. Let $\boldsymbol{\alpha} := (\alpha_1, \alpha_2, \alpha_3)$ and $\boldsymbol{\beta} := (\beta_1, \beta_2, \beta_3)$. For a complex number s, we write $\boldsymbol{\alpha} + s := (\alpha_1 + s, \alpha_2 + s, \alpha_3 + s)$. We define

$$\delta(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \frac{1}{2} \sum_{j=1}^{3} (\alpha_j - \beta_j), \tag{2-1}$$

$$G(s; \boldsymbol{\alpha}, \boldsymbol{\beta}) := \prod_{j=1}^{3} \Gamma\big(s + \tfrac{1}{2}(k-1) + \alpha_j\big) \Gamma\big(s + \tfrac{1}{2}(k-1) - \beta_j\big), \tag{2-2}$$

and

$$\Lambda(f, s; \boldsymbol{\alpha}, \boldsymbol{\beta}) := \prod_{j=1}^{3} \Lambda(f, s + \alpha_j) \Lambda(\bar{f}, s - \beta_j). \tag{2-3}$$

Note that we have

$$\begin{aligned} \Lambda(f; \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \Lambda\big(f, \tfrac{1}{2}; \boldsymbol{\alpha}, \boldsymbol{\beta}\big) \\ &= \Big(\frac{q}{4\pi^2}\Big)^{\delta(\boldsymbol{\alpha}, \boldsymbol{\beta})} G\big(\tfrac{1}{2}; \boldsymbol{\alpha}, \boldsymbol{\beta}\big) \prod_{j=1}^{3} L\big(f, \tfrac{1}{2} + \alpha_j\big) L\big(\bar{f}, \tfrac{1}{2} - \beta_j\big). \end{aligned} \tag{2-4}$$

We define the shifted k-divisor function by

$$\sigma_k(n; \alpha_1, \ldots, \alpha_k) = \sum_{n_1 n_2 \cdots n_k = n} n_1^{-\alpha_1} n_2^{-\alpha_2} \cdots n_k^{-\alpha_k}. \tag{2-5}$$

Let

$$\mathscr{B}(a, b; \boldsymbol{\alpha}) := \frac{\mu(a) \sigma_3(b; \alpha_1 + \alpha_2, \alpha_2 + \alpha_3, \alpha_3 + \alpha_1)}{a^{\alpha_1 + \alpha_2 + \alpha_3}}. \tag{2-6}$$

Next we need the following lemmas, which help us generate the conjecture of the sixth moment, namely

$$\frac{2}{\phi(q)} \sum_{\substack{\chi \pmod q \\ \chi(-1) = (-1)^k}} \sideset{}{^h}\sum_{f \in \mathcal{H}_\chi} \Lambda(f; \boldsymbol{\alpha}, \boldsymbol{\beta}). \tag{2-7}$$

Lemma 2.1. *We have*

$$\sigma_2(n_1 n_2; \alpha_1, \alpha_2) = \sum_{d \mid (n_1, n_2)} \mu(d) d^{-\alpha_1 - \alpha_2} \sigma_2\Big(\frac{n_1}{d}; \alpha_1, \alpha_2\Big) \sigma_2\Big(\frac{n_2}{d}; \alpha_1, \alpha_2\Big).$$

Proof. Since both sides are multiplicative functions, it is enough to prove the lemma when n_1n_2 is a prime power. We set $n_1 = p^a$ and $n_2 = p^b$, where $1 \le a \le b$. Then

$$\begin{aligned}\sigma_2(p^a p^b;\alpha_1,\alpha_2) &= \sum_{0\le k\le a} \mu(p^k)p^{-k(\alpha_1+\alpha_2)}\sigma_2(p^{a-k};\alpha_1,\alpha_2)\sigma_2(p^{b-k};\alpha_1,\alpha_2)\\ &= \sigma_2(p^a;\alpha_1,\alpha_2)\sigma_2(p^b;\alpha_1,\alpha_2)\\ &\qquad - p^{-(\alpha_1+\alpha_2)}\sigma_2(p^{a-1};\alpha_1,\alpha_2)\sigma_2(p^{b-1};\alpha_1,\alpha_2). \end{aligned}\tag{2-8}$$

On the other hand,

$$\sigma_2(p^n;\alpha_1,\alpha_2) = \sum_{\ell=0}^{n} p^{-\ell\alpha_1}p^{-(n-\ell)\alpha_2} = p^{-n\alpha_2}\frac{1-1/p^{(n+1)(\alpha_1-\alpha_2)}}{1-1/p^{\alpha_1-\alpha_2}},$$

and the lemma follows by substituting the above formula into (2-8). □

We write the product of L-functions in terms of Dirichlet series in the next lemma.

Lemma 2.2. *Let $L(f,w)$ be an L-function in $\mathcal{H}_\chi$. For $\mathrm{Re}(s+\alpha_i) > 1$, we have*

$$L(f,s+\alpha_1)L(f,s+\alpha_2)L(f,s+\alpha_3) = \sum_{a,b\ge1}\frac{\chi(ab)\mathscr{B}(a,b;\boldsymbol{\alpha})}{(ab)^{2s}}\sum_{n\ge1}\frac{\lambda_f(an)\sigma_3(n;\boldsymbol{\alpha})}{(an)^s}.$$

Proof. From the Hecke relation (1-2) and Lemma 2.1, we have

$$\begin{aligned}&L(f,s+\alpha_1)L(f,s+\alpha_2)L(f,s+\alpha_3)\\ &= \sum_{n_1\ge1}\frac{\lambda_f(n_1)}{n_1^{s+\alpha_1}}\sum_{d\ge1}\frac{\chi(d)}{d^{2s+\alpha_2+\alpha_3}}\sum_{j\ge1}\frac{\lambda_f(j)\sigma_2(j;\alpha_2,\alpha_3)}{j^s}\\ &= \sum_{d\ge1}\frac{\chi(d)}{d^{2s+\alpha_2+\alpha_3}}\sum\sum_{n_1,j\ge1}\frac{\sigma_2(j;\alpha_2,\alpha_3)}{n_1^{s+\alpha_1}j^s}\sum_{e\mid(n_1,j)}\chi(e)\lambda_f\Big(\frac{jn_1}{e^2}\Big)\\ &= \sum_{a\ge1}\frac{\mu(a)\chi(a)}{a^{2s+\alpha_1+\alpha_2+\alpha_3}}\sum\sum_{d,e\ge1}\frac{\chi(de)\sigma_2(e;\alpha_2,\alpha_3)}{(de)^{2s}d^{\alpha_2+\alpha_3}e^{\alpha_1}}\\ &\qquad\times\sum\sum_{j,n_1\ge1}\frac{\lambda_f(n_1ja)\sigma_2(j;\alpha_2,\alpha_3)}{(ajn_1)^s n_1^{\alpha_1}}\\ &= \sum_{a\ge1}\frac{\mu(a)\chi(a)}{a^{2s+\alpha_1+\alpha_2+\alpha_3}}\sum_{b\ge1}\frac{\chi(b)\sigma_3(b;\alpha_1+\alpha_2,\alpha_2+\alpha_3,\alpha_3+\alpha_1)}{b^{2s}}\\ &\qquad\times\sum_{n\ge1}\frac{\lambda_f(an)\sigma_3(n;\alpha_1,\alpha_2,\alpha_3)}{(an)^s}. \end{aligned}\tag{2-9}$$

This completes the lemma. □

Lemma 2.3. *The orthogonality relation for Dirichlet characters is*

$$\frac{2}{\phi(q)} \sum_{\substack{\chi \pmod q \\ \chi(-1)=(-1)^k}} \chi(m)\overline{\chi}(n) = \begin{cases} 1 & \text{if } m \equiv n \pmod q,\ (mn,q)=1, \\ (-1)^k & \text{if } m \equiv -n \pmod q,\ (mn,q)=1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-10}$$

Petersson's formula gives

$$\sum_{f \in \mathcal{H}_\chi}^{h} \bar{\lambda}_f(m)\lambda_f(n) = \delta_{m=n} + \sigma_\chi(m,n), \tag{2-11}$$

where

$$\begin{aligned} \sigma_\chi(m,n) &= 2\pi i^{-k} \sum_{c \equiv 0 \pmod q} c^{-1} S_\chi(m,n;c) J_{k-1}\Big(\frac{4\pi}{c}\sqrt{mn}\Big) \\ &= 2\pi i^{-k} \sum_{c=1}^{\infty} (cq)^{-1} S_\chi(m,n;cq) J_{k-1}\Big(\frac{4\pi}{cq}\sqrt{mn}\Big), \end{aligned}$$

and S_χ is the Kloosterman sum defined by

$$S_\chi(m,n;cq) = \sum_{a\bar{a} \equiv 1 \pmod{cq}} \chi(a)\, \mathrm{e}\Big(\frac{am+\bar{a}n}{cq}\Big).$$

From Lemma 2.2, we have that

$$\begin{aligned} \prod_{i=1}^{3} & L(f, s+\alpha_i) L(f, s-\beta_i) \\ &= \sum\sum\sum\sum_{a_1,b_1,a_2,b_2 \ge 1} \frac{\chi(a_1b_1)\overline{\chi}(a_2b_2)\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{(a_1b_1a_2b_2)^{2s}} \\ &\qquad\qquad \times \sum\sum_{n,m \ge 1} \frac{\lambda_f(a_1n)\overline{\lambda_f}(a_2m)\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{(a_1na_2m)^s}. \end{aligned} \tag{2-12}$$

By the orthogonality relation of Dirichlet characters and Petersson's formula in Lemma 2.3, a naive guess might be that the main contribution comes from the diagonal terms $a_1b_1 = a_2b_2$ and $a_1n = a_2m$, where $(a_ib_i, q) = 1$, which is

$$\mathcal{C}(s,\boldsymbol{\alpha},\boldsymbol{\beta}) := \sum\sum\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2,m,n \ge 1 \\ a_1n = a_2m \\ a_1b_1 = a_2b_2 \\ (a_i,q)=(b_j,q)=1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{(a_1b_1)^{2s}} \frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{(a_2b_2)^{2s}} \frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{(a_1n)^s(a_2m)^s} \tag{2-13}$$

for $\mathrm{Re}(s)$ large enough. This can be written as the Euler product

$$\mathcal{C}(s,\boldsymbol{\alpha},\boldsymbol{\beta})=\prod_p \mathcal{C}_p(s,\boldsymbol{\alpha},\boldsymbol{\beta}),$$

where for $p\neq q$,

$$\begin{aligned}\mathcal{C}_p(s,\boldsymbol{\alpha},\boldsymbol{\beta}) &:= \\ &\sum\sum\sum\sum\sum\sum_{\substack{r_1,t_1,r_2,t_2,u_1,u_2\geq 0\\ r_1+u_1=r_2+u_2\\ r_1+t_1=r_2+t_2}} \frac{\mathscr{B}(p^{r_1},p^{t_1};\boldsymbol{\alpha})}{p^{2s(r_1+t_1)}}\frac{\mathscr{B}(p^{r_2},p^{t_2};-\boldsymbol{\beta})}{p^{2s(r_2+t_2)}} \\ &\qquad\times\frac{\sigma_3(p^{u_1};\boldsymbol{\alpha})\sigma_3(p^{u_2};-\boldsymbol{\beta})}{p^{s(r_1+u_1)}p^{s(r_2+u_2)}}, \end{aligned}\tag{2-14}$$

and for $p=q$,

$$\mathcal{C}_q(s,\boldsymbol{\alpha},\boldsymbol{\beta}):=\sum_{u\geq 0}\frac{\sigma_3(q^u;\boldsymbol{\alpha})\sigma_3(q^u;-\boldsymbol{\beta})}{q^{2us}}.\tag{2-15}$$

Next, for $\zeta_p(w):=(1-1/p^w)^{-1}$, let

$$\begin{aligned}\mathcal{Z}_p(s;\boldsymbol{\alpha},\boldsymbol{\beta}) &:= \prod_{i=1}^3\prod_{j=1}^3 \zeta_p(2s+\alpha_i-\beta_j),\\ \mathcal{Z}(s;\boldsymbol{\alpha},\boldsymbol{\beta}) &:= \prod_{i=1}^3\prod_{j=1}^3 \zeta(2s+\alpha_i-\beta_j).\end{aligned}\tag{2-16}$$

and

$$\mathcal{A}(s;\boldsymbol{\alpha},\boldsymbol{\beta}):=\mathcal{C}(s;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{Z}(s;\boldsymbol{\alpha},\boldsymbol{\beta})^{-1}=\prod_p \mathcal{C}_p(s;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{Z}_p(s;\boldsymbol{\alpha},\boldsymbol{\beta})^{-1}.\tag{2-17}$$

We define

$$\mathcal{M}(q;\boldsymbol{\alpha},\boldsymbol{\beta}):=\left(\frac{q}{4\pi^2}\right)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})}G\big(\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big)\mathcal{A}\mathcal{Z}\big(\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big).\tag{2-18}$$

When $\mathrm{Re}(\alpha_i),\mathrm{Re}(\beta_i)\ll 1/\log q$, the term $\mathcal{A}\big(\frac{1}{2},\boldsymbol{\alpha},\boldsymbol{\beta}\big)$ is absolutely convergent. Now, let S_j be the permutation group of j variables. Based on the analysis of the diagonal contribution, we expect $\mathcal{M}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ to be a part of the average in (2-7), and we also notice that the expression $\mathcal{M}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ is fixed by the action of $S_3\times S_3$. Since we expect our final answer to be symmetric under the full group S_6, we sum over the cosets $S_6/(S_3\times S_3)$. In fact, the method of Conrey, Farmer, Keating, Rubinstein and Snaith [Conrey et al. 2005] gives the following conjecture for the average of $\Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta})$.

Conjecture 2.4. Assume that $\boldsymbol{\alpha},\boldsymbol{\beta}$ satisfy

$$\begin{aligned}\mathrm{Re}(\alpha_i),\mathrm{Re}(\beta_i)&\ll\frac{1}{\log q},\\ \mathrm{Im}\,(\alpha_i),\mathrm{Im}\,(\beta_i)&\ll q^{1-\varepsilon}.\end{aligned}$$

We have

$$\frac{2}{\phi(q)} \sum_{\substack{\chi \pmod q \\ \chi(-1)=(-1)^k}} \sideset{}{^h}\sum_{f\in\mathcal{H}_\chi} \Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{\pi\in S_6/(S_3\times S_3)} \mathcal{M}(q;\pi(\boldsymbol{\alpha},\boldsymbol{\beta}))(1+O(q^{-1/2+\varepsilon})),$$

where we define $\pi(\boldsymbol{\alpha},\boldsymbol{\beta}) = \pi(\alpha_1,\dots,\alpha_k,\beta_1,\dots,\beta_k)$ for $\pi\in S_{2k}$, where π acts on the $2k$-tuple $(\alpha_1,\dots,\alpha_k,\beta_1,\dots,\beta_k)$ as usual.

We also write $\pi(\boldsymbol{\alpha},\boldsymbol{\beta}) = (\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))$ by an abuse of notation, where $\pi(\boldsymbol{\alpha},\boldsymbol{\beta})$ is as above. Our main goal is to find an asymptotic formula for

$$\mathcal{M}_6(q) := \frac{2}{\phi(q)} \sum_{\substack{\chi \pmod q \\ \chi(-1)=(-1)^k}} \sideset{}{^h}\sum_{f\in\mathcal{H}_\chi} \int_{-\infty}^{\infty} \Lambda(f;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it)\,dt, \tag{2-19}$$

and we will prove the following result.

Theorem 2.5. *Let q be prime and $k\ge 5$ be odd. For $\alpha_i,\beta_j \ll \frac{1}{\log q}$, we have that*

$$\mathcal{M}_6(q) = \int_{-\infty}^{\infty} \sum_{\pi\in S_6/(S_3\times S_3)} \mathcal{M}(q,\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it)\,dt + O(q^{-1/4+\varepsilon}).$$

We note that as the shifts go to 0, the main term of this moment is of the size $(\log q)^9$, and we derive Theorem 1.2. We refer the reader to [Conrey et al. 2005] for the details of this type of calculation.

3. Approximate functional equation

In this section, we prove an approximate functional equation for the product of L-functions. Let

$$H(s;\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_{j=1}^{3}\prod_{\ell=1}^{3}\Big(s^2-\Big(\frac{\alpha_j-\beta_\ell}{2}\Big)^2\Big)^3,$$

and define, for any $\xi>0$,

$$W(\xi;\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{2\pi i}\int_{(1)} G\big(\tfrac12+s;\boldsymbol{\alpha},\boldsymbol{\beta}\big)H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\xi^{-s}\,\frac{ds}{s}.$$

Moreover, let $\Lambda_0(f,\boldsymbol{\alpha},\boldsymbol{\beta})$ be

$$\Big(\frac{q}{4\pi^2}\Big)^{\delta(\alpha,\beta)} \sum\sum\sum\sum_{a_1,b_1,a_2,b_2\ge 1} \frac{\chi(a_1b_1)\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1b_1}\,\frac{\overline{\chi}(a_2b_2)\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2b_2}$$
$$\times\sum\sum_{n,m\ge 1}\frac{\lambda_f(a_1n)\sigma_3(n;\boldsymbol{\alpha})}{(a_1n)^{1/2}}\,\frac{\overline{\lambda_f}(a_2m)\sigma_3(m;-\boldsymbol{\beta})}{(a_2m)^{1/2}}$$
$$\times W\bigg(\frac{(2\pi)^6a_1^3b_1^2a_2^3b_2^2nm}{q^3};\boldsymbol{\alpha},\boldsymbol{\beta}\bigg). \tag{3-1}$$

Lemma 3.1. *We have*

$$H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta}) = \Lambda_0(f,\boldsymbol{\alpha},\boldsymbol{\beta}) + \Lambda_0(f,\boldsymbol{\beta},\boldsymbol{\alpha}).$$

Proof. We consider

$$I := \frac{1}{2\pi i}\int_{(1)} \Lambda\big(f, s+\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big) H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\,\frac{ds}{s}.$$

Moving the contour integral to (-1), we obtain that

$$\begin{aligned} I &= \Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta})H(0;\boldsymbol{\alpha},\boldsymbol{\beta}) + \frac{1}{2\pi i}\int_{(-1)} \Lambda\big(f, s+\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big) H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\,\frac{ds}{s} \\ &= \Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta})H(0;\boldsymbol{\alpha},\boldsymbol{\beta}) - \frac{1}{2\pi i}\int_{(1)} \Lambda\big(f, -s+\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big) H(-s;\boldsymbol{\alpha},\boldsymbol{\beta})\,\frac{ds}{s}. \end{aligned}$$

By the functional equation, we have $\Lambda\big(f, s+\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big) = \Lambda\big(f, -s+\tfrac{1}{2};\boldsymbol{\beta},\boldsymbol{\alpha}\big)$. Moreover, H is an even function, and $H(s;\boldsymbol{\alpha},\boldsymbol{\beta}) = H(s;\boldsymbol{\beta},\boldsymbol{\alpha})$. Therefore,

$$\begin{aligned} \Lambda(f;\boldsymbol{\alpha},\boldsymbol{\beta})H(0;\boldsymbol{\alpha},\boldsymbol{\beta}) &= \frac{1}{2\pi i}\int_{(1)} \Lambda\big(f, s+\tfrac{1}{2};\boldsymbol{\alpha},\boldsymbol{\beta}\big) H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\,\frac{ds}{s} \\ &\qquad + \frac{1}{2\pi i}\int_{(1)} \Lambda\big(f, s+\tfrac{1}{2};\boldsymbol{\beta},\boldsymbol{\alpha}\big) H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\,\frac{ds}{s}. \end{aligned}$$

The lemma follows after writing Λ as a product of L-functions and Gamma functions and using Lemma 2.2. □

Next, we let

$$V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\xi,\eta;\mu) = \Big(\frac{\mu}{4\pi^2}\Big)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})} \int_{-\infty}^{\infty} \Big(\frac{\eta}{\xi}\Big)^{it} W\Big(\frac{\xi\eta(4\pi^2)^3}{\mu^3};\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it\Big)\,dt,$$

and

$$\begin{aligned} \Lambda_1(f;\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum\sum\sum\sum_{a_1,b_1,a_2,b_2\ge 1} \frac{\chi(a_1b_1)\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1b_1}\,\frac{\overline{\chi}(a_2b_2)\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2b_2} \\ &\quad\times \sum\sum_{n,m\ge 1} \frac{\lambda_f(a_1n)\sigma_3(n;\boldsymbol{\alpha})}{(a_1n)^{1/2}}\,\frac{\overline{\lambda_f}(a_2m)\sigma_3(m;-\boldsymbol{\beta})}{(a_2m)^{1/2}} V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n, a_2^3b_2^2m; q). \end{aligned} \tag{3-2}$$

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

$$H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\int_{-\infty}^{\infty} \Lambda(f;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it)\,dt = \Lambda_1(f;\boldsymbol{\alpha},\boldsymbol{\beta}) + \Lambda_1(f;\boldsymbol{\beta},\boldsymbol{\alpha}).$$

The proof follows easily from Lemma 3.1.

Remark. The integration over t is added so that the main contribution comes from when $a_1^3b_1^2n \ll q^{3/2+\varepsilon}$ and $a_2^3b_2^2m \ll q^{3/2+\varepsilon}$, as we will see from Lemma 3.3 below. Without the integration over t, the ranges of a_i, b_j, m, n that we need to consider satisfy the weaker condition $a_1^3b_1^2na_2^3b_2^2m \ll q^{3+\varepsilon}$, and the proof presented here does not extend to this range.

Lemma 3.3. *If ξ or $\eta \gg q^{3/2+\varepsilon}$, then for any $A>1$, we have*

$$V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\xi,\eta;q) \ll q^{-A},$$

where the implied constant depends on ε and A.

Proof. From the definition of W and V and a change of variables ($s+it=w$, $s-it=z$), we can write $V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\xi,\eta;\mu)$ as

$$\left(\frac{q}{4\pi^2}\right)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})}\frac{4\pi}{(2\pi i)^2}\int_{(0)}\prod_{j=1}^{3}\Gamma\bigl(w+\tfrac{1}{2}(k-1)+\alpha_j\bigr)\left(\frac{q^{3/2}}{(4\pi^2)^{3/2}\xi}\right)^{w}$$
$$\times\int_{(1)}\prod_{j=1}^{3}\Gamma\bigl(z+\tfrac{1}{2}(k-1)-\beta_j\bigr)\left(\frac{q^{3/2}}{(4\pi^2)^{3/2}\eta}\right)^{z}H\bigl(\tfrac{1}{2}(z+w);\boldsymbol{\alpha},\boldsymbol{\beta}\bigr)\,\frac{dz\,dw}{z+w}.$$

When $\xi \gg q^{3/2+\varepsilon}$, we move the contour integral over w to the far right, and similarly, when $\eta \gg q^{3/2+\varepsilon}$, we move the contour integral over z to the far right. The lemma then follows. □

4. Setup for the proof of Theorem 2.5 and diagonal terms

From Lemma 3.2, we have that for $\alpha_i,\beta_i \ll 1/\log q$,

$$H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{M}_6(q)=\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\{\Lambda_1(f;\boldsymbol{\alpha},\boldsymbol{\beta})+\Lambda_1(f;\boldsymbol{\beta},\boldsymbol{\alpha})\}.$$

Therefore, to evaluate $\mathcal{M}_6(q)$, it is sufficient to compute asymptotically

$$\mathscr{M}_1(q;\boldsymbol{\alpha},\boldsymbol{\beta}):=\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\Lambda_1(f;\boldsymbol{\alpha},\boldsymbol{\beta}). \tag{4-1}$$

Applying Petersson's formula, we obtain

$$\sideset{}{^h}\sum_{f\in\mathcal{H}_\chi}\bar{\lambda}_f(a_2m)\lambda_f(a_1n)=\delta_{a_2m=a_1n}+\sigma_\chi(a_2m,a_1n),$$

where $\sigma_\chi(a_2m,a_1n)$ is defined as in Lemma 2.3. We then write

$$\mathscr{M}_1(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})+\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{4-2}$$

where $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ is the diagonal contribution from $\delta_{a_2m=a_1n}$, and $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ is the contribution from $\sigma_\chi(a_2m,a_1n)$.

In Section 4A, we show that the term $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ contributes one of the twenty terms in Conjecture 2.4, namely the term corresponding to $\mathcal{M}(q;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it)$. Moreover, $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ gives another nine terms in the conjecture, namely those transpositions in $S_6/S_3\times S_3$ which switch α_i and β_j for a fixed $i, j=1,2,3$. We explicitly work out one of these terms in Proposition 6.1. Similarly, $\mathscr{D}(q;\boldsymbol{\beta},\boldsymbol{\alpha})$ gives rise to the term corresponding to $\mathcal{M}(q;\boldsymbol{\beta}+it,\boldsymbol{\alpha}+it)$, and the last nine expressions arise from $\mathscr{K}(q;\boldsymbol{\beta},\boldsymbol{\alpha})$.

4A. ***Evaluating the diagonal terms*** $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$**.** Recall that

$$\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\sum\sum\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2,m,n\ge 1\\ a_1n=a_2m}}\chi(a_1b_1)\bar\chi(a_2b_2)$$
$$\times\frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2b_2}\frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{(a_1n)^{1/2}(a_2m)^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n,a_2^3b_2^2m;q).$$

We compute the diagonal contribution in the following lemma.

Lemma 4.1. *With the same notation as above, we have*

$$\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\int_{-\infty}^{\infty}\mathcal{M}(q;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it)\,dt+O(q^{-3/4+\varepsilon}).$$

Proof. We apply the orthogonality relation for Dirichlet characters in (2-10) and obtain that for $(a_ib_i,q)=1$,

$$\frac{2}{\phi(q)}\sum_{\substack{\chi \pmod q\\ \chi(-1)=(-1)^k}}\chi(a_1b_1)\bar\chi(a_2b_2)=\begin{cases}1 & \text{if } a_1b_1\equiv a_2b_2 \pmod q,\\ (-1)^k & \text{if } a_1b_1\equiv -a_2b_2 \pmod q,\\ 0 & \text{otherwise.}\end{cases}$$

When $a_1b_1\ge q/4$ or $a_2b_2\ge q/4$, we have that $a_1^3b_1^2n\ge q^2/16$ or $a_2^3b_2^2m\ge q^2/16$. From Lemma 3.3, $V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n,a_2^3b_2^2m;q)\ll q^{-A}$ in that range, so the contribution from these terms is negligible. Hence the main contribution from $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ comes from the terms with $a_1b_1=a_2b_2$ when $(a_ib_i,q)=1$, and

$$\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\sum\sum\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2,m,n\ge 1\\ a_1n=a_2m\\ a_1b_1=a_2b_2\\ (a_ib_i,q)=1}}\frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2b_2}$$
$$\times\frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{(a_1n)^{1/2}(a_2m)^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n,a_2^3b_2^2m;q)+O(q^{-A}).$$

Since $a_1b_1=a_2b_2$ and $a_1n=a_2m$, $a_1^3b_1^2n=a_2^3b_2^2m$. Therefore, $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ can be written as

$$\frac{1}{2\pi i}\Big(\frac{q}{4\pi^2}\Big)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})}\int_{-\infty}^{\infty}\int_{(1)} G\big(\tfrac12+s;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it\big) \times H(s;\boldsymbol{\alpha},\boldsymbol{\beta})\Big(\frac{q}{4\pi^2}\Big)^{3s}\mathcal{AZ}\big(\tfrac12+s,\boldsymbol{\alpha},\boldsymbol{\beta}\big)\,\frac{ds}{s}\,dt,$$

where we have used equations (2-13)–(2-17).

Note that $\mathcal{A}(s;\boldsymbol{\alpha},\boldsymbol{\beta})$ is absolutely convergent when $\mathrm{Re}(s)>\frac14+\varepsilon$. Furthermore, the pole at $s=-\frac12(\alpha_i-\beta_j)$ from the zeta factor $\mathcal{Z}\big(\frac12+s;\boldsymbol{\alpha},\boldsymbol{\beta}\big)$ is canceled by the zero at the same point from $H(s;\boldsymbol{\alpha},\boldsymbol{\beta})$. Thus, in the region $\mathrm{Re}(s)>-\frac14+\varepsilon$, the integrand is analytic except for a simple pole at $s=0$. Moving the line of integration to $\mathrm{Re}(s)=-\frac14+\varepsilon$, we obtain that $\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ is

$$\Big(\frac{q}{4\pi^2}\Big)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})}\int_{-\infty}^{\infty} G\big(\tfrac12;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it\big)H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{AZ}\big(\tfrac12;\boldsymbol{\alpha},\boldsymbol{\beta}\big)\,dt+O(q^{-3/4+\varepsilon}).$$

The lemma now follows from (2-18) and $\mathcal{AZ}\big(\frac12;\boldsymbol{\alpha},\boldsymbol{\beta}\big)=\mathcal{AZ}\big(\frac12;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it\big)$. □

5. Setup for the off-diagonal terms $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$

Define $\mathcal{K}f=i^{-k}f+i^k\bar{f}$. If g is a real function, then $g\mathcal{K}f=\mathcal{K}(gf)$. Applying the orthogonality relation for χ from (2-10) to $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$, we obtain that

$$\begin{aligned}\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=2\pi\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\ge1\\(a_1a_2b_1b_2,q)=1}}&\frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2b_2}\\ &\times\sum\sum_{n,m\ge1}\frac{\sigma_3(n;\boldsymbol{\alpha})}{(a_1n)^{1/2}}\frac{\sigma_3(m;-\boldsymbol{\beta})}{(a_2m)^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n,a_2^3b_2^2m;q)\\ &\times\sum_{c=1}^{\infty}\frac{1}{cq}J_{k-1}\Big(\frac{4\pi}{cq}\sqrt{a_2ma_1n}\Big)\mathcal{K}\sideset{}{^*}\sum_{\substack{a\ (\mathrm{mod}\ cq)\\ a\equiv\overline{a_1b_1}a_2b_2\ (\mathrm{mod}\ q)}}\mathrm{e}\Big(\frac{aa_2m+\bar{a}a_1n}{cq}\Big),\end{aligned}$$

where $\sum^*$ denotes a sum over reduced residues. Let f be a smooth partition of unity such that

$$\sum_M^d f\Big(\frac{m}{M}\Big)=1,$$

where f is supported in $\big[\frac12,3\big]$ and $\sum_M^d$ denotes a dyadic sum over $M=2^k$, $k\ge0$.

Rearranging the sum, we have

$$\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{2\pi}{q}\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\ge1\\(a_1a_2b_1b_2,q)=1}}\frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2}b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2}b_2} \times\sum_M^d\sum_N^d S(\boldsymbol{a},\boldsymbol{b},M,N;\boldsymbol{\alpha},\boldsymbol{\beta}),$$

where

$$S(\boldsymbol{a},\boldsymbol{b},M,N;\boldsymbol{\alpha},\boldsymbol{\beta}) := \sum_{c=1}^{\infty}\frac{1}{c}\sum\sum_{m,n\geq 1}\frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{n^{1/2}m^{1/2}}\mathcal{F}(\boldsymbol{a},\boldsymbol{b},m,n,c),$$

$$\begin{aligned}\mathcal{F}(\boldsymbol{a},\boldsymbol{b},m,n,c) &:= \mathcal{G}(\boldsymbol{a},\boldsymbol{b},m,n,c)\\ &\quad\times \mathcal{K}\sum^{*}_{\substack{a \ (\mathrm{mod}\ cq)\\ a\equiv\overline{a_1b_1}a_2b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\Big(\frac{aa_2m+\bar{a}a_1n}{cq}\Big)J_{k-1}\Big(\frac{4\pi}{cq}\sqrt{a_2ma_1n}\Big),\end{aligned} \tag{5-1}$$

$$\mathcal{G}(\boldsymbol{a},\boldsymbol{b},m,n,c) := V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n, a_2^3b_2^2m;q)f\Big(\frac{m}{M}\Big)f\Big(\frac{n}{N}\Big). \tag{5-2}$$

As described in the outline of the paper, we now endeavor to compute $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$.

We write

$$\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta}) = \mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta}) + \mathscr{K}_E(q;\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{5-3}$$

letting $\mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ represent the contribution from the sum over $c<C$, where $C=\sqrt{a_1a_2MN}/q^{2/3}$, and $\mathscr{K}_E(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ the rest. We show that the contribution from $\mathscr{K}_E(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ is small in Section 5A. This is possible by the decay of the Bessel functions, and such a truncation bounds the size of the conductor inside the exponential sum.

For $\mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta})$, we start by reducing the conductor inside the exponential sum from cq to c in Section 5B. This step takes advantage of the average over $\chi \pmod q$.

Before showing each step, we provide properties of Bessel functions that will be used later.

Lemma 5.1. *We have*

$$J_{k-1}(2\pi x) = \frac{1}{\pi\sqrt{x}}\big\{W(2\pi x)\,\mathrm{e}\big(x-\tfrac14 k+\tfrac18\big) + \overline{W}(2\pi x)\,\mathrm{e}\big(-x+\tfrac14 k-\tfrac18\big)\big\}, \tag{5-4}$$

where $W^{(j)}(x)\ll_{j,k} x^{-j}$. *Moreover,*

$$J_{k-1}(2x) = \sum_{\ell=0}^{\infty}(-1)^{\ell}\frac{x^{2\ell+k-1}}{\ell!(\ell+k-1)!}, \tag{5-5}$$

and

$$J_{k-1}(x) \ll \min(x^{-1/2}, x^{k-1}). \tag{5-6}$$

Finally, when calculating the main terms of $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$, *we use the fact that for* $\alpha,\beta,\gamma>0$,

$$\begin{aligned}\mathcal{K}\int_0^{\infty}\mathrm{e}((\alpha+\beta)x+\gamma x^{-1})J_{k-1}\big(4\pi\sqrt{\alpha\beta x}\big)\,\frac{dx}{x}\\ = 2\pi\, J_{k-1}\big(4\pi\sqrt{\alpha\gamma}\big)J_{k-1}\big(4\pi\sqrt{\beta\gamma}\big),\end{aligned} \tag{5-7}$$

and the integration is 0 *if* $\alpha,\beta>0$ *and* $\gamma\leq 0$.

These results are standard. We refer the reader to [Watson 1944] for the first three claims, and to [Oberhettinger 1972] for the last claim.

5A. ***Truncating the sum over c.*** In this section we show that we can truncate the sum over c in $S(\boldsymbol{a}, \boldsymbol{b}, M, N; \boldsymbol{\alpha}, \boldsymbol{\beta})$ with small error contribution.

Proposition 5.2. *Let $C = \sqrt{a_1 a_2 MN}/q^{2/3}$, $k \geq 5$, and $\mathcal{F}(\boldsymbol{a}, \boldsymbol{b}, m, n, c)$ be defined as in (5-1). Further, let*

$$\mathscr{K}_E(q; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1, b_1, a_2, b_2 \geq 1 \\ (a_1 a_2 b_1 b_2, q) = 1}} \frac{\mathscr{B}(a_1, b_1; \boldsymbol{\alpha})}{a_1^{3/2} b_1} \frac{\mathscr{B}(a_2, b_2; -\boldsymbol{\beta})}{a_2^{3/2} b_2} \times \sum_M^d \sum_N^d S_E(\boldsymbol{a}, \boldsymbol{b}, M, N; \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

where

$$S_E(\boldsymbol{a}, \boldsymbol{b}, M, N; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{c \geq C} \frac{1}{c} \sum\sum_{m, n \geq 1} \frac{\sigma_3(n; \boldsymbol{\alpha}) \sigma_3(m; -\boldsymbol{\beta})}{n^{1/2} m^{1/2}} \mathcal{F}(\boldsymbol{a}, \boldsymbol{b}, m, n, c).$$

Then $\mathscr{K}_E(q; \boldsymbol{\alpha}, \boldsymbol{\beta}) \ll q^{-5/12+\varepsilon}$.

Proof. Note that the contribution of terms when $a_1^3 b_1^2 N$ or $a_2^3 b_2^2 M \gg q^{3/2+\varepsilon}$ is $\ll_{\varepsilon, A} q^{-A}$, due to the fast decay rate of $\mathcal{G}(\boldsymbol{a}, \boldsymbol{b}, m, n, c)$ defined in (5-2). Thus we discount such terms in the rest of the proof. For $k \geq 5$, we let

$$\begin{aligned}
\tilde{\Delta}(m, n) &= \tilde{\Delta}(m, n, \boldsymbol{a}, \boldsymbol{b}; C) \\
&= \sum_{c \geq C} \frac{1}{c} \mathcal{K} \sideset{}{^*}\sum_{\substack{a \ (\mathrm{mod}\ cq) \\ a \equiv \overline{a_1 b_1} a_2 b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\Big(\frac{a a_2 m + \bar{a} a_1 n}{cq}\Big) J\Big(\frac{4\pi \sqrt{a_1 a_2 mn}}{cq}\Big) \\
&= \tilde{\Delta}_1(m, n) + \tilde{\Delta}_2(m, n),
\end{aligned}$$

where

$$\tilde{\Delta}_1(m, n) := \sum_{\substack{c \geq C \\ (c, q) = 1}} \frac{1}{c} \mathcal{K} \sideset{}{^*}\sum_{\substack{a \ (\mathrm{mod}\ cq) \\ a \equiv \overline{a_1 b_1} a_2 b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\Big(\frac{a a_2 m + \bar{a} a_1 n}{cq}\Big) J\Big(\frac{4\pi \sqrt{a_1 a_2 mn}}{cq}\Big),$$

and $\tilde{\Delta}_2(m, n)$ is the sum of the terms where $(c, q) > 1$. Now for $(c, q) = 1$, the Weil bound gives

$$\sideset{}{^*}\sum_{\substack{a \ (\mathrm{mod}\ cq) \\ a \equiv \overline{a_1 b_1} a_2 b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\Big(\frac{a a_2 m + \bar{a} a_1 n}{cq}\Big) \ll c^{1/2+\varepsilon} \sqrt{(a_1 m, a_2 n, c)},$$

and from the bound in (5-6), we have

$$J\Big(\frac{4\pi \sqrt{a_1 a_2 mn}}{cq}\Big) \ll \Big(\frac{\sqrt{a_1 a_2 mn}}{cq}\Big)^{k-1}.$$

When $a_1^3 b_1^2 N$ and $a_2^3 b_2^2 M \ll q^{3/2+\varepsilon}$, we obtain that for $k \geq 5$,

$$\sum_{m,n} \frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{\sqrt{mn}} \tilde{\Delta}_1(m,n)\mathcal{G}(\boldsymbol{a},\boldsymbol{b},m,n,c) \\ \ll q^{\varepsilon} M^{k/2} N^{k/2} \sum_{c \geq C} \frac{1}{c} c^{1/2+\varepsilon} \biggl(\frac{\sqrt{a_1 a_2}}{cq}\biggr)^{k-1} \ll q^{7/12+\varepsilon}.$$

In the above, we have used that $\max(a_1^3 b_1^2 N, a_2^3 b_2^2 M) \ll q^{3/2+\varepsilon}$. Then summing over a_1, b_1, a_2, b_2 gives the desired bound.

Now, for $(c,q) > 1$, we use the bound

$$\sideset{}{^*}\sum_{\substack{a \ (\mathrm{mod}\ cq) \\ a \equiv \overline{a_1 b_1} a_2 b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\biggl(\frac{a a_2 m + \bar{a} a_1 n}{cq}\biggr) \ll (cq)^{1/2+\varepsilon} \sqrt{(a_1 m, a_2 n, cq)}.$$

Hence, for $q \mid c$ and $k \geq 5$, we obtain

$$\sum_{m,n} \frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{\sqrt{mn}} \tilde{\Delta}_2(m,n)\mathcal{G}(\boldsymbol{a},\boldsymbol{b},m,n,c) \\ \ll q^{\varepsilon} M^{k/2} N^{k/2} \sum_{\substack{c \geq C \\ q \mid c}} \frac{1}{c} (qc)^{1/2} \biggl(\frac{\sqrt{a_1 a_2}}{cq}\biggr)^{k-1} \ll q^{1/12+\varepsilon}.$$

Then summing over a_1, b_1, a_2, b_2 gives the desired bound. □

From this proposition, we are left to consider only

$$\mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2 \geq 1 \\ (a_1 a_2 b_1 b_2, q) = 1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2} b_1} \frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2} b_2} \\ \times \sideset{}{^d}\sum_{M} \sideset{}{^d}\sum_{N} S_M(\boldsymbol{a},\boldsymbol{b},M,N;\boldsymbol{\alpha},\boldsymbol{\beta}), \tag{5-8}$$

where

$$S_M(\boldsymbol{a},\boldsymbol{b},M,N;\boldsymbol{\alpha},\boldsymbol{\beta}) := \sum_{c<C} \frac{1}{c} \sum_{m,n \geq 1} \frac{\sigma_3(n;\boldsymbol{\alpha})\sigma_3(m;-\boldsymbol{\beta})}{n^{1/2} m^{1/2}} \mathcal{F}(\boldsymbol{a},\boldsymbol{b},m,n,c). \tag{5-9}$$

5B. ***Treatment of the exponential sum.*** Next, we reduce the conductor in the exponential sum in $\mathcal{F}(\boldsymbol{a},\boldsymbol{b},m,n,c)$ before applying Voronoi summation.

Lemma 5.3 (treatment of the exponential sum). *Assume that* $(c,q) = 1$ *and let*

$$Y := \sideset{}{^*}\sum_{\substack{a \ (\mathrm{mod}\ cq) \\ a \equiv \overline{a_1 b_1} a_2 b_2 \ (\mathrm{mod}\ q)}} \mathrm{e}\biggl(\frac{au + \bar{a}v}{cq}\biggr).$$

Then we have

$$Y = \mathrm{e}\left(\frac{(a_2b_2)^2u+(a_1b_1)^2v}{cqa_1b_1a_2b_2}\right) \sideset{}{^*}\sum_{x \pmod c} \mathrm{e}\left(\frac{\bar{q}(a_1b_1x-a_2b_2)u}{a_1b_1c}\right)\mathrm{e}\left(\frac{\bar{q}(a_2b_2\bar{x}-a_1b_1)v}{a_2b_2c}\right).$$

Proof. By the Chinese remainder theorem, for each $a \pmod{cq}$, there exist unique $x \pmod c$ and $y \pmod q$ such that

$$a = xq\bar{q}+yc\bar{c}, \tag{5-10}$$

where $\bar{q}$ denotes the inverse of $q \pmod c$, and $\bar{c}$ denotes the inverse of $c \pmod q$. Using (5-10) and the reciprocity relation

$$\frac{\bar{a}}{b}+\frac{\bar{b}}{a} \equiv \frac{1}{ab} \pmod 1,$$

where $(a, b) = 1$, $\bar{a}$ is the inverse of $a \pmod b$, and $\bar{b}$ is the inverse of $b \pmod a$, we obtain that

$$Y = \mathrm{e}\left(\frac{\overline{a_1b_1}a_2b_2\bar{c}u+a_1b_1\overline{a_2b_2}\bar{c}v}{q}\right) \sideset{}{^*}\sum_{x \pmod c} \mathrm{e}\left(\frac{x\bar{q}u+\bar{x}\bar{q}v}{c}\right).$$

Thus

$$Y = \mathrm{e}\left(\frac{(a_2b_2)^2u+(a_1b_1)^2v}{cqa_1b_1a_2b_2}\right) \sideset{}{^*}\sum_{x \pmod c} \mathrm{e}\left(\frac{x\bar{q}u+\bar{x}\bar{q}v}{c}\right)\mathrm{e}\left(\frac{-\bar{q}a_2b_2u}{ca_1b_1}\right)\mathrm{e}\left(\frac{-\bar{q}a_1b_1v}{ca_2b_2}\right),$$

and the lemma follows. □

Note that when $c < C < q$, we automatically have $(c, q) = 1$. The point of this lemma is that we may treat

$$\mathrm{e}\left(\frac{(a_2b_2)^2u+(a_1b_1)^2v}{cqa_1b_1a_2b_2}\right)$$

as a smooth function with small derivatives, while the other exponentials have conductor at most $ca_ib_i \le q^{1+\varepsilon}$ after truncation. It should be noted, however, that we are most concerned with the contribution from the transition region of the Bessel function, where the conductor ca_ib_i should be thought of as around size $q^{1/2}$.

6. Applying Voronoi summation

To calculate $\mathscr{K}_M(q; \boldsymbol{\alpha}, \boldsymbol{\beta})$ from Section 5A, we first evaluate $S_M(\boldsymbol{a}, \boldsymbol{b}, M, N; \boldsymbol{\alpha}, \boldsymbol{\beta})$, defined in (5-9). We write

$$S_M(\boldsymbol{a}, \boldsymbol{b}, M, N; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{c<C} \sideset{}{^*}\sum_{x \pmod c} \frac{1}{c}(\mathcal{S}^+_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c, x)+\mathcal{S}^-_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c, x)),$$

in which

$$\mathcal{S}^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)=i^{\mp k}\sum_{m\geq 1}\frac{\sigma_3(m;-\boldsymbol{\beta})}{m^{1/2}}f\Big(\frac{m}{M}\Big)\,\mathrm{e}\Big(\pm\frac{a_2^2b_2m}{cqa_1b_1}\Big)\,\mathrm{e}\Big(\pm\frac{\bar{q}(a_1b_1x-a_2b_2)a_2m}{a_1b_1c}\Big)$$
$$\times\sum_{n\geq 1}\sigma_3(n;\boldsymbol{\alpha})F^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(m,n,c)\,\mathrm{e}\Big(\pm\frac{\bar{q}(a_2b_2\bar{x}-a_1b_1)a_1n}{a_2b_2c}\Big),$$

with

$$F^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(m,n,c)$$
$$=\frac{1}{n^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2n,a_2^3b_2^2m;q)J_{k-1}\Big(\frac{4\pi}{cq}\sqrt{a_2ma_1n}\Big)\,\mathrm{e}\Big(\pm\frac{a_1^2b_1n}{cqa_2b_2}\Big)f\Big(\frac{n}{N}\Big).$$

Let

$$\frac{\lambda_1}{\eta_1}=\frac{\bar{q}(a_2b_2\bar{x}-a_1b_1)a_1}{a_2b_2c}\tag{6-1}$$

and

$$\frac{\lambda_2}{\eta_2}=\frac{\bar{q}(a_1b_1x-a_2b_2)a_2}{a_1b_1c},\tag{6-2}$$

where $(\lambda_1,\eta_1)=(\lambda_2,\eta_2)=1$. Moreover, define

$$\mathcal{V}^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c;\boldsymbol{a},\boldsymbol{b},y,z)=\frac{1}{y^{1/2}}\frac{1}{z^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2y,a_2^3b_2^2z;q)J_{k-1}\Big(\frac{4\pi\sqrt{a_2za_1y}}{cq}\Big)$$
$$\times i^{\mp k}\,\mathrm{e}\Big(\pm\frac{a_1^2b_1y}{cqa_2b_2}\pm\frac{a_2^2b_2z}{cqa_1b_1}\Big)f\Big(\frac{y}{N}\Big)f\Big(\frac{z}{M}\Big).$$

We then apply Voronoi summation as in Theorem B.1 to the sum over n,m and obtain that $\mathcal{S}^{+}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)+\mathcal{S}^{-}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)$ is

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\operatorname*{Res}_{s_1=1-\alpha_i}\operatorname*{Res}_{s_2=1+\beta_j}(\mathcal{T}^{+}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)+\mathcal{T}^{-}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2))$$
$$+\sum_{i=1}^{8}(\mathcal{T}^{+}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)+\mathcal{T}^{-}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)),$$

where in the region of absolute convergence,

$$\mathcal{T}^{\pm}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2):=\mathcal{F}^{\pm}_{M}(c;\boldsymbol{\alpha},\boldsymbol{\beta})D_3\Big(s_1,\pm\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big)D_3\Big(s_2,\pm\frac{\lambda_2}{\eta_2},-\boldsymbol{\beta}\Big),\tag{6-3}$$
$$\mathcal{F}^{\pm}_{M}(c;\boldsymbol{\alpha},\boldsymbol{\beta}):=\mathcal{F}^{\pm}_{M}(c,s_1,s_2;\boldsymbol{\alpha},\boldsymbol{\beta})$$
$$:=\int_0^\infty\int_0^\infty y^{s_1-1}z^{s_2-1}\mathcal{V}^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c;\boldsymbol{a},\boldsymbol{b},y,z)\,dy\,dz,$$

$$\mathcal{T}^{\pm}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x) := \frac{\pi^{3/2+\alpha_1+\alpha_2+\alpha_3}}{\eta_1^{3+\alpha_1+\alpha_2+\alpha_3}} \sum_{i=1}^{3} \operatorname*{Res}_{s=1+\beta_i} D_3\Big(s_2, \pm\frac{\lambda_2}{\eta_2}, -\boldsymbol{\beta}\Big) \times \sum_{n=1}^{\infty} A_3\Big(n, \pm\frac{\lambda_1}{\eta_1}, \boldsymbol{\alpha}\Big) \mathcal{F}^{\pm}_1(c,n;\boldsymbol{\alpha},\boldsymbol{\beta}), \quad (6\text{-}4)$$

$$\mathcal{F}^{\pm}_1(c,n;\boldsymbol{\alpha},\boldsymbol{\beta}) := \mathcal{F}^{\pm}_1(c,n,s;\boldsymbol{\alpha},\boldsymbol{\beta}) := \int_0^\infty \int_0^\infty z^{s-1} \mathcal{V}^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c;\boldsymbol{a},\boldsymbol{b},y,z) U_3\Big(\frac{\pi^3 ny}{\eta_1^3};\boldsymbol{\alpha}\Big)\,dy\,dz,$$

and $\mathcal{T}^{\pm}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)$ is defined similarly for $i=2,3,4$. Further,

$$\mathcal{T}^{\pm}_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x) := \frac{\pi^{3+\alpha_1-\beta_1+\alpha_2-\beta_2+\alpha_3-\beta_3}}{\eta_1^{3+\alpha_1+\alpha_2+\alpha_3}\eta_2^{3-\beta_1-\beta_2-\beta_3}} \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} A_3\Big(n,\pm\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big) \times A_3\Big(m,\pm\frac{\lambda_2}{\eta_2},-\boldsymbol{\beta}\Big)\mathcal{F}^{\pm}_5(c,n,m;\boldsymbol{\alpha},\boldsymbol{\beta}), \quad (6\text{-}5)$$

$$\mathcal{F}^{\pm}_5(c,n,m;\boldsymbol{\alpha},\boldsymbol{\beta}) := \int_0^\infty\int_0^\infty \mathcal{V}^{\pm}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c;\boldsymbol{a},\boldsymbol{b},y,z) \times U_3\Big(\frac{\pi^3 mz}{\eta_2^3};-\boldsymbol{\beta}\Big)U_3\Big(\frac{\pi^3 ny}{\eta_1^3};\boldsymbol{\alpha}\Big)\,dy\,dz,$$

and $\mathcal{T}^{\pm}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)$ is defined similarly for $i=6,7,8$.

As mentioned in Section 4, there are nine terms from $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$. In particular, we will show that these terms arise from

$$\sum_{c<C}\ \sideset{}{^*}\sum_{x \pmod c} \frac{1}{c}(\mathcal{T}^{+}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)+\mathcal{T}^{-}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)),$$

and in fact each term comes from the residues at $s_1=1-\alpha_i$ and $s_2=1+\beta_j$ for $i=1,2,3$. We state the contribution from the residues $s_1=1-\alpha_1$ and $s_2=1+\beta_1$ in Proposition 6.1 below, and prove it in Section 7. By symmetry, the analogous result holds for the other residues. Then, we will show that the rest of $\mathcal{T}^{\pm}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)$ are negligible in Section 8 as stated in Proposition 6.2.

Proposition 6.1. *Let*

$$R_{\alpha_1,\beta_1} := \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\ge 1\\ (a_1a_2b_1b_2,q)=1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2}b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2}b_2} \times \sum_M^d \sum_N^d \sum_{c<C}\ \sideset{}{^*}\sum_{x \pmod c} \frac{1}{c} \operatorname*{Res}_{s_1=1-\alpha_1}\ \operatorname*{Res}_{s_2=1+\beta_1} (\mathcal{T}^{+}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2) + \mathcal{T}^{-}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)).$$

Then we have

$$R_{\alpha_1,\beta_1} = H(0;\boldsymbol{\alpha},\boldsymbol{\beta}) \int_{-\infty}^{\infty} \mathcal{M}(q;\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it)\,dt + O(q^{-1/2+\varepsilon}),$$

where $(\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})) = (\beta_1,\alpha_2,\alpha_3;\alpha_1,\beta_2,\beta_3)$.

Proposition 6.2. *For* $i = 1,\ldots,8$, *define*

$$\mathcal{E}_i^{\pm}(q;\boldsymbol{\alpha},\boldsymbol{\beta}) := \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\ge 1\\ (a_1a_2b_1b_2,q)=1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2}b_1} \frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2}b_2} \times \sum_{M}^{d} \sum_{N}^{d} \sum_{c<C} \sideset{}{^*}\sum_{x \pmod c} \frac{\mathcal{T}^{\pm}_{i,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)}{c}.$$

Then

$$\mathcal{E}_i^{\pm}(q;\boldsymbol{\alpha},\boldsymbol{\beta}) \ll q^{-1/4+\varepsilon}.$$

We will prove this proposition in Section 8.

7. Proof of Proposition 6.1

We begin by collecting some lemmas which will be used in this section.

7A. *Preliminary lemmas.*

Lemma 7.1. *Let* $(a,\ell) = 1$. *We have*

$$f(c,\ell) := \sideset{}{^*}\sum_{\substack{x \pmod{c\ell}\\ x\equiv a \pmod{\ell}}} 1 = c\prod_{\substack{p\mid c\\ p\nmid \ell}}\Big(1-\frac{1}{p}\Big) = \phi(c)\prod_{p\mid(\ell,c)}\Big(1-\frac{1}{p}\Big)^{-1}.$$

Proof. We first prove that if $(m,n\ell) = 1$, then

$$f(mn,\ell) = f(n,\ell)\phi(m). \tag{7-1}$$

For all x satisfying $(x,mn\ell) = 1$, we can write $x = um\bar{m} + vn\ell\overline{n\ell}$ such that $m\bar{m} \equiv 1 \pmod{n\ell}$, $n\ell\overline{n\ell} \equiv 1 \pmod m$, and $(u,n\ell) = (v,m) = 1$. Moreover, $x \equiv a \pmod \ell$ if and only if $u \equiv a \pmod \ell$. By the Chinese remainder theorem,

$$f(mn,\ell) = \sideset{}{^*}\sum_{\substack{x \pmod{mn\ell}\\ x\equiv a \pmod{\ell}}} 1 = \sideset{}{^*}\sum_{\substack{u \pmod{n\ell}\\ u\equiv a \pmod{\ell}}} \sideset{}{^*}\sum_{v \pmod m} 1 = f(n,\ell)\phi(m).$$

Let $c = c_1c_2$, where all prime factors of c_1 also divide ℓ, and $(c_2,\ell) = 1$. From (7-1), we have that $f(c,\ell) = f(c_1,\ell)\phi(c_2)$.

Now let x be any residue modulo $c_1\ell$ with $x \equiv a \pmod \ell$. Then $(x,c_1\ell) = 1$ since $(a,\ell) = 1$. Thus all such x can be uniquely written as $x = a + k\ell$, where $k = 0,\ldots,c_1-1$, so $f(c_1,\ell) = c_1$. We then have $f(c,\ell) = c_1\phi(c_2)$, and the statement follows from the identity $\phi(c_2) = c_2\prod_{p\mid c_2}(1-1/p)$. $\square$

Lemma 7.2. *Let α, β, y, z be nonnegative real numbers satisfying $\alpha y, \beta z \ll q^2$ and define*

$$T = T(y, z, \alpha, \beta) := \sum_{\delta=1}^{\infty} \frac{1}{\delta} J_{k-1}\Big(\frac{4\pi\sqrt{\alpha\beta yz}}{\delta}\Big) \mathcal{K}\, \mathrm{e}\Big(\frac{\alpha y}{\delta} + \frac{\beta z}{\delta}\Big).$$

Further, let $L = q^{100}$ and w be a smooth function on $\mathbb{R}^+$ with $w(x) = 1$ if $0 \le x \le 1$, and $w(x) = 0$ if $x > 2$. Then for any $A > 0$, we have

$$T = 2\pi \sum_{\ell=1}^{\infty} w\Big(\frac{\ell}{L}\Big) J_{k-1}\big(4\pi\sqrt{\alpha y \ell}\big) J_{k-1}\big(4\pi\sqrt{\beta z \ell}\big) \\ - 2\pi \int_0^{\infty} w\Big(\frac{\ell}{L}\Big) J_{k-1}\big(4\pi\sqrt{\alpha y \ell}\big) J_{k-1}\big(4\pi\sqrt{\beta z \ell}\big)\, d\ell + O_A(q^{-A}).$$

Proof. We follow the arguments of Section 3 of [Iwaniec and Li 2007] to evaluate T. Let $\eta(s)$ be a smooth function on $\mathbb{R}^+$ with $\eta(s) = 0$ if $0 \le s < \frac{1}{4}$, $0 \le \eta(s) \le 1$ if $\frac{1}{4} \le s \le \frac{1}{2}$, and $\eta(s) = 1$ if $s > \frac{1}{2}$. We then obtain that

$$T = \mathcal{K} \sum_{\delta} \frac{\eta(\delta)}{\delta} J_{k-1}\Big(\frac{4\pi\sqrt{\alpha\beta yz}}{\delta}\Big) \mathrm{e}\Big(\frac{\alpha y}{\delta} + \frac{\beta z}{\delta}\Big).$$

After inserting this smooth function we apply Poisson summation to obtain that

$$T = \sum_{\ell} \hat{F}(\ell) := \sum_{\ell} \mathcal{K} \int_0^{\infty} \frac{\eta(u)}{u} \mathrm{e}\Big(\ell u + \frac{\alpha y}{u} + \frac{\beta z}{u}\Big) J_{k-1}\Big(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\Big)\, du.$$

By (5-4), we can write the integral above in terms of two integrals with the phase

$$\ell u + \frac{\alpha y \pm \sqrt{\alpha\beta yz} + \beta z}{u}.$$

If $|\ell| > L$, the factor ℓu dominates. Then integrating by parts A times, we have that

$$\int_0^{\infty} \frac{\eta(u)}{u} \mathrm{e}\Big(\ell u + \frac{\alpha y}{u} + \frac{\beta z}{u}\Big) J_{k-1}\Big(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\Big)\, du \ll q^{-A}.$$

Therefore,

$$T = \sum_{\ell} \hat{F}(\ell) w\Big(\frac{|\ell|}{L}\Big) + O(q^{-A}).$$

Now, we write $\sum_{\ell} \hat{F}(\ell) w(|\ell|/L) = T_1 - T_2$, where

$$T_1 := \sum_{\ell} w\Big(\frac{|\ell|}{L}\Big) \mathcal{K} \int_0^{\infty} \frac{1}{u} \mathrm{e}\Big(\ell u + \frac{\alpha y}{u} + \frac{\beta z}{u}\Big) J_{k-1}\Big(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\Big)\, du,$$

and

$$T_2 := \sum_{\ell} w\left(\frac{|\ell|}{L}\right)\mathcal{K}\int_0^\infty \frac{1-\eta(u)}{u}\,\mathrm{e}\left(\ell u+\frac{\alpha y}{u}+\frac{\beta z}{u}\right)J_{k-1}\left(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\right)du.$$

We use (5-7) to evaluate T_1 and obtain

$$T_1 = 2\pi\sum_{\ell=1}^{\infty} w\left(\frac{\ell}{L}\right)J_{k-1}\big(4\pi\sqrt{\alpha y\ell}\big)J_{k-1}\big(4\pi\sqrt{\beta z\ell}\big). \tag{7-2}$$

For T_2, we note that $\xi(u)=1-\eta(u)=1$ if $0<s<\frac{1}{4}$, $0\le\xi(u)\le 1$ if $\frac14\le s\le\frac12$, and $\xi(u)=0$ if $s>\frac12$. Interchanging the sum over ℓ and the integration over u and applying the Poisson summation formula, we have

$$T_2 = \mathcal{K}\int_0^\infty \frac{\xi(u)}{u}\,\mathrm{e}\left(\frac{\alpha y}{\delta}+\frac{\beta z}{\delta}\right)J_{k-1}\left(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\right)\sum_{\ell} L\hat{w}(L(\ell+u))\,du.$$

Since $\hat{w}(y)\ll(1+|y|)^{-A}$, the main contribution comes from $\ell=0$ and $0\le u<\frac14$. Therefore

$$\begin{aligned}
T_2 &= \mathcal{K}\int_0^\infty \frac{\xi(u)}{u}\,\mathrm{e}\left(\frac{\alpha y}{\delta}+\frac{\beta z}{\delta}\right)J_{k-1}\left(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\right)L\hat{w}(Lu)\,du+O(q^{-A})\\
&= \mathcal{K}\int_0^\infty \frac{1}{u}\,\mathrm{e}\left(\frac{\alpha y}{\delta}+\frac{\beta z}{\delta}\right)J_{k-1}\left(\frac{4\pi\sqrt{\alpha\beta yz}}{u}\right)L\hat{w}(Lu)\,du+O(q^{-A})\\
&= 2\pi\int_0^\infty w\left(\frac{\ell}{L}\right)J_{k-1}\big(4\pi\sqrt{\alpha y\ell}\big)J_{k-1}\big(4\pi\sqrt{\beta z\ell}\big)\,d\ell+O(q^{-A}),
\end{aligned} \tag{7-3}$$

where the last equality comes from Plancherel's formula and (5-7). □

The next lemma deals with the sum and the integral involving ℓ.

Lemma 7.3. *Let w be a smooth function on $\mathbb{R}^+$ with $w(x)=1$ if $0\le x\le 1$, and $w(x)=0$ if $x>2$. Also let γ be a complex number with $\operatorname{Re}\gamma\ll 1/\log q$, $\operatorname{Re}\gamma<0$, and $L=q^{100}$. Then*

$$\sum_{\ell=1}^{\infty} w\left(\frac{\ell}{L}\right)\frac{1}{\ell^{1+\gamma}}-\int_0^\infty w\left(\frac{\ell}{L}\right)\frac{1}{\ell^{1+\gamma}}\,d\ell=\zeta(1+\gamma)+O(q^{-20}).$$

Proof. Let $\tilde{w}(z)$ be the Mellin transform of w, defined by

$$\tilde{w}(z)=\int_0^\infty w(t)\frac{t^z}{t}\,dt.$$

From the definition, $\tilde{w}(z)$ is analytic for $\operatorname{Re} z>0$, and integration by parts gives

$$\tilde{w}(z)=-\frac{1}{z}\int_0^\infty w'(t)t^z\,dt,$$

so $\tilde{w}(z)$ can be analytically continued to $\operatorname{Re} z > -1$ except at $z = 0$, where it has a simple pole with residue $w(0) = 1$. For $\sigma > \min\{0, -\operatorname{Re}\gamma\}$, we have

$$\sum_{\ell=1}^{\infty} w\Big(\frac{\ell}{L}\Big)\frac{1}{\ell^{1+\gamma}} = \sum_{\ell=1}^{\infty} \frac{1}{2\pi i}\int_{(\sigma)} \tilde{w}(z)\Big(\frac{\ell}{L}\Big)^{-z}\frac{1}{\ell^{1+\gamma}}\,dz$$
$$= \frac{1}{2\pi i}\int_{(\sigma)} \tilde{w}(z)L^z\zeta(1+\gamma+z)\,dz. \tag{7-4}$$

Shifting the contour to $\operatorname{Re}(z) = -\frac{1}{4}$, we have that (7-4) is

$$\zeta(1+\gamma)+\tilde{w}(-\gamma)L^{-\gamma}+O(q^{-20}).$$

The lemma follows from noting that

$$\tilde{w}(-\gamma)L^{-\gamma} = \int_0^{\infty} w\Big(\frac{\ell}{L}\Big)\frac{1}{\ell^{1+\gamma}}\,d\ell. \qquad \square$$

7B. ***Calculation of residues.*** In this section, we calculate

$$\operatorname*{Res}_{s_1=1-\alpha_1}\ \operatorname*{Res}_{s_2=1+\beta_1}\big(\mathcal{T}^{+}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)+\mathcal{T}^{-}_{M,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x,s_1,s_2)\big).$$

To do this, we essentially need to consider

$$\operatorname*{Res}_{s_1=1-\alpha_1} D_3\Big(s_1,\pm\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big)y^{s_1-1},$$

where λ_1/η_1 is defined in (6-1). Let $(a_1b_1, a_2b_2) = \lambda$, $a_1b_1 = u_1\lambda$, $a_2b_2 = u_2\lambda$, where $(u_1, u_2) = 1$. Note that $(u_1x-u_2, c) = (u_2\bar{x}-u_1, c) = \delta$. Hence

$$\delta_1 := ((a_2b_2\bar{x}-a_1b_1)a_1, a_2b_2c) = \lambda((u_2\bar{x}-u_1)a_1, u_2c) = \lambda\delta(a_1, u_2c/\delta),$$

and $\lambda_1 = \bar{q}(a_2b_2\bar{x}-a_1b_1)a_1/\delta_1$ and $\eta_1 = a_2b_2c/\delta_1$.

By (B-4), we obtain that

$$\mathcal{R}_1\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big) := \operatorname*{Res}_{s_1=1-\alpha_1} D_3\Big(s_1,\pm\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big)$$
$$= \frac{1}{\eta_1^{2-2\alpha_1+\alpha_2+\alpha_3}}\sum\sum_{\substack{1\le a_1,a_2\le\eta_1\\ \eta_1\mid a_1a_2}} \zeta\Big(1-\alpha_1+\alpha_2,\frac{a_2}{\eta_1}\Big)\zeta\Big(1-\alpha_1+\alpha_3,\frac{a_3}{\eta_1}\Big).$$

Hence

$$\operatorname*{Res}_{s_1=1-\alpha_1} D_3\Big(s_1,\pm\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big)y^{s_1-1} = \mathcal{R}_1\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big)y^{-\alpha}.$$

Similarly, we let $\mathcal{R}_2(c,\boldsymbol{a},\boldsymbol{b}) := \operatorname{Res}_{s_2=1+\beta_1} D_3(s_2, \pm\lambda_2/\eta_2, -\boldsymbol{\beta})$. Then

$$\mathcal{R}_2\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big) = \frac{1}{\eta_2^{2+2\beta_1-\beta_2-\beta_3}}\sum\sum_{\substack{1\le a_1,a_2\le\eta_2\\ \eta_2\mid a_1a_2}} \zeta\Big(1+\beta_1-\beta_2,\frac{a_2}{\eta_2}\Big)\zeta\Big(1+\beta_1-\beta_3,\frac{a_3}{\eta_2}\Big),$$

and

$$\operatorname*{Res}_{s_2=1+\beta_1} D_3\Big(s_1, \pm\frac{\lambda_2}{\eta_2}, \boldsymbol{\alpha}\Big) z^{s_2-1} = \mathcal{R}_2\Big(\frac{c}{\delta}, \boldsymbol{a}, \boldsymbol{b}\Big) z^{\beta}.$$

7C. *Computing R_{α_1,β_1}.* From the previous section, R_{α_1,β_1} can be written as

$$R_{\alpha_1,\beta_1} = \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\geq 1\\ (a_1a_2b_1b_2,q)=1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2}b_1} \frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2}b_2} \times \sum_M^d \sum_N^d \mathcal{A}(\boldsymbol{a},\boldsymbol{b},M,N) + O(q^{-1/2+\varepsilon}),$$

where $\mathcal{A}(\boldsymbol{a},\boldsymbol{b},M,N)$ is defined as

$$\sum_{c=1}^{\infty} \frac{\mathcal{F}(c)}{c} \sideset{}{^*}\sum_{x \pmod c} \mathcal{R}_1\Big(\frac{c}{\delta}, \boldsymbol{a}, \boldsymbol{b}\Big) \mathcal{R}_2\Big(\frac{c}{\delta}, \boldsymbol{a}, \boldsymbol{b}\Big), \tag{7-5}$$

and

$$\begin{aligned}\mathcal{F}(c) &:= \mathcal{F}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(c,\boldsymbol{a},\boldsymbol{b}) \\ &:= \int_0^\infty\int_0^\infty \frac{1}{y^{1/2+\alpha_1}} \frac{1}{z^{1/2-\beta_1}} V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2y, a_2^3b_2^2z; q) J_{k-1}\Big(\frac{4\pi\sqrt{a_2ya_1z}}{cq}\Big) \\ &\qquad \times f\Big(\frac{y}{N}\Big) f\Big(\frac{z}{M}\Big) \mathcal{K}\, \mathrm{e}\Big(\frac{a_1^2b_1y}{cqa_2b_2} + \frac{a_2^2b_2z}{cqa_1b_1}\Big)\, dy\, dz.\end{aligned}$$

We remark that we can extend the sum over c to all positive integers in a similar manner as in the truncation argument in Proposition 5.2. Now, we let

$$\frac{1}{c^2}\mathcal{G}(c,\boldsymbol{a},\boldsymbol{b}) := \mathcal{R}_1(c,\boldsymbol{a},\boldsymbol{b})\mathcal{R}_2(c,\boldsymbol{a},\boldsymbol{b}),$$

so that we can write the sum over c in (7-5) as

$$\begin{aligned}&\sum_{c=1}^{\infty} \frac{\mathcal{F}(c)}{c} \sum_{\delta\mid c} \sideset{}{^*}\sum_{\substack{x \pmod c\\ (u_1x-u_2,c)=\delta}} \frac{1}{(c/\delta)^2} \mathcal{G}\Big(\frac{c}{\delta}, \boldsymbol{a}, \boldsymbol{b}\Big) \\ &\qquad = \sum_{\delta=1}^{\infty} \frac{1}{\delta} \sum_{c=1}^{\infty} \frac{\mathcal{F}(c\delta)}{c} \sideset{}{^*}\sum_{\substack{x \pmod{c\delta}\\ (u_1x-u_2,c\delta)=\delta}} \frac{\mathcal{G}(c,\boldsymbol{a},\boldsymbol{b})}{c^2} \\ &\qquad = \sum_{\delta=1}^{\infty} \frac{1}{\delta} \sum_{c=1}^{\infty} \frac{\mathcal{F}(c\delta)}{c} \sideset{}{^*}\sum_{x \pmod{c\delta}} \sum_{b\mid((u_1x-u_2)/\delta,c)} \frac{\mu(b)\mathcal{G}(c,\boldsymbol{a},\boldsymbol{b})}{c^2} \\ &\qquad = \sum_{\substack{\delta\geq 1\\ (\delta,u_1u_2)=1}} \frac{1}{\delta} \sum_{\substack{b\geq 1\\ (b,u_1u_2)=1}} \frac{\mu(b)}{b^3} \sum_{c\geq 1} \frac{\mathcal{G}(cb,\boldsymbol{a},\boldsymbol{b})\mathcal{F}(cb\delta)}{c^3} \sideset{}{^*}\sum_{\substack{x \pmod{c\delta b}\\ x\equiv u_2\bar{u}_1 \pmod{b\delta}}} 1,\end{aligned}$$

where the sum over x is 0 if $(u_1u_2, b\delta) \neq 1$ since $(u_1, u_2) = 1$. Applying Lemma 7.1 to the sum over x, we then obtain that

$$
\begin{aligned}
&\mathcal{A}(\boldsymbol{a}, \boldsymbol{b}, M, N)\\
&\quad = \sum_{\substack{\delta\geq 1\\(\delta,u_1u_2)=1}} \frac{1}{\delta} \sum_{\substack{b\geq 1\\(b,u_1u_2)=1}} \frac{\mu(b)}{b^3} \sum_{c\geq 1} \frac{1}{c^2} \prod_{\substack{p\mid c\\ p\nmid b\delta}} \Big(1-\frac{1}{p}\Big) \mathcal{F}(cb\delta)\mathcal{G}(cb, \boldsymbol{a}, \boldsymbol{b})\\
&\quad = \sum_{\substack{h\geq 1\\ h\mid u_1u_2}} \frac{\mu(h)}{h} \sum_{\delta\geq 1} \frac{1}{\delta} \sum_{\substack{b\geq 1\\(b,u_1u_2)=1}} \frac{\mu(b)}{b^3} \sum_{c\geq 1} \frac{1}{c^2} \prod_{\substack{p\mid c\\ p\nmid bh\delta}} \Big(1-\frac{1}{p}\Big) \mathcal{F}(cbh\delta)\mathcal{G}(cb, \boldsymbol{a}, \boldsymbol{b})\\
&\quad = \sum_{\substack{h\geq 1\\ h\mid u_1u_2}} \frac{\mu(h)}{h} \sum_{\substack{b\geq 1\\(b,u_1u_2)=1}} \frac{\mu(b)}{b^3} \sum_{c\geq 1} \frac{\mathcal{G}(cb, \boldsymbol{a}, \boldsymbol{b})}{c^2}\\
&\qquad\qquad \times \sum_{\gamma\mid c} \prod_{\substack{p\mid c\\ p\nmid bh\gamma}} \Big(1-\frac{1}{p}\Big) \sum_{\delta\geq 1} \frac{1}{\delta} \mathcal{F}(cbh\delta) \sum_{g\mid (c/\gamma, \delta/\gamma)} \mu(g)\\
&\quad = \sum\nolimits^{\#} \mathscr{G}_{\boldsymbol{a},\boldsymbol{b}}(1; h, b, c, \gamma, g) \sum_{\delta\geq 1} \frac{1}{\delta} \mathcal{F}(cbhg\gamma\delta),
\end{aligned}
$$

where

$$
\begin{aligned}
&\sum\nolimits^{\#} \mathscr{G}_{\boldsymbol{a},\boldsymbol{b}}(s; h, b, c, \gamma, g)\\
&= \sum_{\substack{h\geq 1\\ h\mid u_1u_2}} \frac{\mu(h)}{h^s} \sum_{\substack{b\geq 1\\(b,u_1u_2)=1}} \frac{\mu(b)}{b^{2+s}} \sum_{c\geq 1} \frac{\mathcal{G}(cb, \boldsymbol{a}, \boldsymbol{b})}{c^{1+s}} \sum_{\gamma\mid c} \frac{1}{\gamma^s} \sum_{g\mid (c/\gamma)} \frac{\mu(g)}{g^s} \prod_{\substack{p\mid c\\ p\nmid bh\gamma}} \Big(1-\frac{1}{p}\Big). \qquad (7\text{-}6)
\end{aligned}
$$

Next, applying Lemma 7.2 to the sum over δ and summing $\sum_M^d \sum_N^d$, we have that

$$
\begin{aligned}
&\sum_M\nolimits^{d} \sum_N\nolimits^{d} \mathcal{A}(\boldsymbol{a}, \boldsymbol{b}, M, N)\\
&\quad = 2\pi \frac{1}{2\pi i} \Big(\frac{q}{4\pi^2}\Big)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})} \int_{-\infty}^{\infty} \int_{(1)} \sum\nolimits^{\#} \mathscr{G}_{\boldsymbol{a},\boldsymbol{b}}(1; h, b, c, \gamma, g)\\
&\qquad\qquad \times \bigg\{\sum_{\ell=1}^{\infty} w\Big(\frac{\ell}{L}\Big) - \int_0^{\infty} w\Big(\frac{\ell}{L}\Big)\, d\ell\bigg\} G\Big(\frac{1}{2}+s; \boldsymbol{\alpha}+it, \boldsymbol{\beta}+it\Big) H(s; \boldsymbol{\alpha}, \boldsymbol{\beta})\\
&\qquad\qquad\qquad \times \Big(\frac{a_2^3b_2^2}{a_1^3b_1^2}\Big)^{it} (a_1^3b_1^2a_2^3b_2^2)^{-s} \frac{q^{3s}}{(4\pi^2)^{3s}}\\
&\qquad\qquad \times \int_0^{\infty}\int_0^{\infty} \frac{y^{-1/2-\alpha_1}}{y^{s+it}} \frac{z^{-1/2+\beta_1}}{z^{s-it}} J_{k-1}\bigg(4\pi\sqrt{\frac{a_1^2b_1y\ell}{a_2b_2cbhg\gamma q}}\bigg)\\
&\qquad\qquad\qquad \times J_{k-1}\bigg(4\pi\sqrt{\frac{a_2^2b_2z\ell}{a_1b_1cbhg\gamma q}}\bigg)\, dy\, dz\, \frac{ds}{s}\, dt.
\end{aligned}
$$

The integration over y and z can be evaluated by Equation 707.14 in [Gradshteyn and Ryzhik 2007], which is

$$\int_0^\infty v^\mu (vk)^{1/2} J_\nu(vk)\, dv = 2^{\mu+1/2} k^{-\mu-1} \frac{\Gamma(\mu/2+\nu/2+3/4)}{\Gamma(\nu/2-\mu/2+1/4)},$$

for $-\operatorname{Re}\nu - \frac{3}{2} < \operatorname{Re}\mu < 0$. Then we apply Lemma 7.3 to the sum and the integration over ℓ. Therefore, after summing over a_1, a_2, b_1, b_2, we obtain that the main term of R_{α_1,β_1} is

$$\frac{1}{2\pi i}\Big(\frac{q}{4\pi^2}\Big)^{\delta(\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))} \int_{-\infty}^{\infty}\int_{(\varepsilon)} G_{\alpha_1,\beta_1}\Big(\frac{1}{2}+s; \boldsymbol{\alpha}+it, \boldsymbol{\beta}+it\Big) H(s;\boldsymbol{\alpha},\boldsymbol{\beta}) \mathcal{M}_{\alpha_1,\beta_1}(s) \times \zeta(1-\alpha_1+\beta_1-2s)\frac{q^s}{(4\pi^2)^s}\frac{ds}{s}\,dt, \tag{7-7}$$

where $(\pi(\alpha), \pi(\beta)) = (\beta_1, \alpha_2, \alpha_3; \alpha_1, \beta_2, \beta_3)$,

$$G_{\alpha_i,\beta_j}\Big(\frac{1}{2}+s;\boldsymbol{\alpha},\boldsymbol{\beta}\Big) = \Gamma\Big(\frac{k}{2}-s-\alpha_i\Big)\Gamma\Big(\frac{k}{2}-s+\beta_j\Big)\prod_{\ell\neq i}\Gamma\Big(\frac{k}{2}+s+\alpha_\ell\Big)\prod_{\ell\neq j}\Gamma\Big(\frac{k}{2}+s-\beta_\ell\Big),$$

and

$$\mathcal{M}_{\alpha_1,\beta_1}(s) := \sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\geq 1\\ (a_1a_2b_1b_2,q)=1}} \frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{2-2\alpha_1-\beta_1+2s} b_1^{1-\alpha_1-\beta_1+2s}} \frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{2+\alpha_1+2\beta_1+2s} b_2^{1+\alpha_1+\beta_1+2s}} \times \sum{}^{\#} \mathscr{G}_{\boldsymbol{a},\boldsymbol{b}}(2s+\alpha_1-\beta_1; h, b, c, \gamma, g). \tag{7-8}$$

Now, in the ensuing discussion, we temporarily assume $\operatorname{Re}(\alpha_1) < \operatorname{Re}(\alpha_2), \operatorname{Re}(\alpha_3)$ and $\operatorname{Re}(\beta_1) > \operatorname{Re}(\beta_2), \operatorname{Re}(\beta_3)$. In this region,

$$\mathcal{R}_1\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big) = \sum\sum_{\substack{n_2,n_3\geq 1\\ \frac{u_2c}{\delta(a_1,u_2c/\delta)}\mid n_2n_3}} \frac{1}{n_2^{1+\alpha_2-\alpha_1} n_3^{1+\alpha_3-\alpha_1}}$$

and

$$\operatorname*{Res}_{s_1=1-\alpha_i} D_3\Big(s_1, \pm\frac{\lambda_1}{\eta_1}, \boldsymbol{\alpha}\Big) y^{s_1-1} = \mathcal{R}_1\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big) y^{-\alpha_1} = y^{-\alpha_1} \sum\sum_{\substack{n_2,n_3\geq 1\\ \frac{u_2c}{\delta(a_1,u_2c/\delta)}\mid n_2n_3}} \frac{1}{n_2^{1+\alpha_2-\alpha_1} n_3^{1+\alpha_3-\alpha_1}}.$$

Similarly,

$$\operatorname*{Res}_{s_2=1+\beta_1} D_3\Big(s_2, \pm\frac{\lambda_2}{\eta_2}, -\boldsymbol{\beta}\Big) z^{s_2-1} = \mathcal{R}_2\Big(\frac{c}{\delta},\boldsymbol{a},\boldsymbol{b}\Big) z^{\beta_1} = z^{\beta_1} \sum\sum_{\substack{m_2,m_3\geq 1\\ \frac{u_1c}{\delta(a_2,u_1c/\delta)}\mid m_2m_3}} \frac{1}{m_2^{1+\beta_1-\beta_2} m_3^{1+\beta_1-\beta_3}}.$$

Thus,

$$
\begin{aligned}
\frac{1}{c^2}&\mathcal{G}(c,\boldsymbol{a},\boldsymbol{b})\\
&= \sum_{\substack{n_2,n_3\ge 1\\ u_2c/(a_1,u_2c)\,|\,n_2n_3}}\sum \frac{1}{n_2^{1+\alpha_2-\alpha_1}n_3^{1+\alpha_3-\alpha_1}} \sum_{\substack{m_2,m_3\ge 1\\ u_1c/(a_2,u_1c)\,|\,m_2m_3}}\sum \frac{1}{m_2^{1+\beta_1-\beta_2}m_3^{1+\beta_1-\beta_3}}\\
&= \frac{(a_1,u_2c)(a_2,u_1c)}{u_1u_2c^2}\sum_{n=1}^{\infty}\frac{\sigma_2(u_2cn/(a_1,u_2c);\alpha_2-\alpha_1,\alpha_3-\alpha_1)}{n}\\
&\qquad\qquad\times\sum_{m=1}^{\infty}\frac{\sigma_2(u_1cm/(a_2,u_1c);\beta_1-\beta_2,\beta_1-\beta_3)}{m}. \qquad (7\text{-}9)
\end{aligned}
$$

From this, we may then check that

$$
\mathcal{M}_{\alpha_1,\beta_1}(s)=\prod_{j=2}^{3}\zeta(1+2s+\alpha_j-\beta_1)\zeta(1+2s+\alpha_1-\beta_j)\mathcal{J}_{\alpha_1,\beta_1}(s), \qquad (7\text{-}10)
$$

where $\mathcal{J}_{\alpha_1,\beta_1}$ is absolutely convergent in the region $\mathrm{Re}(s)=-\frac{1}{4}+\varepsilon$. Although we have a priori only verified (7-10) for the region $\mathrm{Re}(\alpha_1)<\mathrm{Re}(\alpha_2),\mathrm{Re}(\alpha_3)$ and $\mathrm{Re}(\beta_1)>\mathrm{Re}(\beta_2),\mathrm{Re}(\beta_3)$, we see that (7-10) must hold for all values of α_i,β_j by analytic continuation.

We note that the pole of $\zeta(1-\alpha_1+\beta_1-2s)$ at $s=\frac{1}{2}(\alpha_1-\beta_1)$ and the poles of $\zeta(1+2s+\alpha_i-\beta_j)$ at $s=\frac{1}{2}(\alpha_i-\beta_j)$ cancel with the zeros at the same point from $H(s;\boldsymbol{\alpha},\boldsymbol{\beta})$. Thus, the integrand in (7-7) has only a simple pole at $s=0$ and is analytic for all values of s with $\mathrm{Re}\,s>-\frac{1}{4}+\epsilon$. Moving the line of integration to $\mathrm{Re}(s)=-\frac{1}{4}+\varepsilon$, we then obtain the main term

$$
\zeta(1-\alpha_1+\beta_1)\mathcal{M}_{\alpha_1,\beta_1}(0)\biggl\{\Bigl(\frac{q}{4\pi^2}\Bigr)^{\delta(\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))}H(0;\boldsymbol{\alpha},\boldsymbol{\beta})
\times\int_{-\infty}^{\infty}G\Bigl(\frac{1}{2};\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it\Bigr)\,dt\biggr\},
$$

with negligible error term. To finish the proof of Proposition 6.1, we will show that the local factor at prime p of the Euler product of $\zeta(1-\alpha_1+\beta_1)\mathcal{M}_{\alpha_1,\beta_1}(0)$ is the same as the one in $\mathcal{AZ}\bigl(\frac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\bigr)$ defined in (2-16) and (2-17). The details of this are in Appendix A.

8. Proof of Proposition 6.2

To prove the proposition, it suffices to show that $\mathcal{E}_i^{+}(q;\boldsymbol{\alpha},\boldsymbol{\beta})\ll q^{-1/4+\varepsilon}$ for $i=1$ and $i=5$, since the proofs of upper bounds for other terms are similar. We start with a lemma that will be used in the proof.

Lemma 8.1. *Let λ, η be integers such that $(\lambda, \eta) = 1$ and $\lambda, \eta \ll q^A$, where A is a fixed constant. Moreover, for $i, j = 1, 2, 3$, let $\alpha_i \ll 1/\log q$ and $|\alpha_i - \alpha_j| \gg 1/q^{\varepsilon_1}$ when $i \neq j$. Then for $\varepsilon > \varepsilon_1$,*

$$\operatorname*{Res}_{s=1-\alpha_i} D_3\Big(s, \frac{\lambda}{\eta}, \boldsymbol{\alpha}\Big) \ll \frac{q^{\varepsilon}}{\eta},$$

where $D_3(s, \lambda/\eta, \boldsymbol{\alpha})$ is defined in (B-4).

Proof. By symmetry, it suffices to prove the statement for the residue at $1 - \alpha_1$. For $\mathrm{Re}(s) > 1 + \mathrm{Re}(\alpha_1 - \alpha_j)$, where $j = 2, 3$, let

$$\begin{aligned} D(s) &:= \sum_{\substack{m=1\\ \eta \mid m}}^{\infty} \frac{\sigma_2(m; \alpha_2 - \alpha_1, \alpha_3 - \alpha_1)}{m^s} \\ &= \frac{1}{\eta^s} \sum_{d \mid \eta} \frac{\mu(d)\sigma_2(\eta/d; \alpha_2 - \alpha_1, \alpha_3 - \alpha_1)}{d^{s+\alpha_2+\alpha_3-2\alpha_1}} \zeta(s + \alpha_2 - \alpha_1)\zeta(s + \alpha_3 - \alpha_1), \end{aligned}$$

where we have used Lemma 2.1 to derive the last line. Now, $D(s)$ can be continued analytically to the whole complex plane except for poles at $s = 1 + \alpha_1 - \alpha_j$ for $j = 2, 3$. Moreover, $D(1) \ll q^{\varepsilon}/\eta$.

For $i = 1, 2, 3$, and $\mathrm{Re}(s + \alpha_i) > 1$, the sum in the lemma can be rewritten as

$$\frac{1}{\eta^{3s+\alpha_1+\alpha_2+\alpha_3}} \sum\sum\sum_{r_1, r_2, r_3 \pmod{\eta}} \mathrm{e}\Big(\frac{\lambda r_1 r_2 r_3}{\eta}\Big) \zeta\Big(s + \alpha_1; \frac{r_1}{\eta}\Big) \zeta\Big(s + \alpha_2; \frac{r_2}{\eta}\Big) \zeta\Big(s + \alpha_3; \frac{r_3}{\eta}\Big).$$

This sum can be analytically continued to the whole complex plane except for poles at $s = 1 - \alpha_i$ for $i = 1, 2, 3$. After some arrangement, the contribution of the residue at $s = 1 - \alpha_1$ is

$$\frac{1}{\eta^{2+\alpha_2+\alpha_3-2\alpha_1}} \sum\sum_{\substack{r_2, r_3 \pmod{\eta}\\ \eta \mid r_2 r_3}} \zeta\Big(1 + \alpha_2 - \alpha_1; \frac{r_2}{\eta}\Big) \zeta\Big(1 + \alpha_3 - \alpha_1; \frac{r_3}{\eta}\Big) = D(1),$$

and the lemma follows. □

8A. *Bounding $\mathcal{E}_1^+(q; \boldsymbol{\alpha}, \boldsymbol{\beta})$.* With the same notation as in Section 7 and $\mathscr{B}$ defined as in (2-6), we recall that

$$\mathcal{E}^+_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(q; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{2\pi}{q} \sum\sum\sum\sum_{\substack{a_1, b_1, a_2, b_2 \geq 1\\ (a_1 a_2 b_1 b_2, q) = 1}} \frac{\mathscr{B}(a_1, b_1; \boldsymbol{\alpha})}{a_1^{3/2} b_1} \frac{\mathscr{B}(a_2, b_2; -\boldsymbol{\beta})}{a_2^{3/2} b_2} \times \sum_{M}^{d} \sum_{N}^{d} E^+_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a}, \boldsymbol{b}, M, N),$$

where

$$E^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a},\boldsymbol{b},M,N) := \sum_{c<C} \sideset{}{^*}\sum_{x \pmod{\delta c}} \frac{\mathcal{T}^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(c,x)}{c}$$

$$= \sum_{\delta<C} \frac{1}{\delta} \sum_{c<C/\delta} \sideset{}{^*}\sum_{\substack{x \pmod{\delta c} \\ (u_1x-u_2,c\delta)=\delta}} \frac{\mathcal{T}^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(c\delta,x)}{c},$$

$$\mathcal{T}^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(c\delta,x) := \frac{\pi^{3/2+\alpha_1+\alpha_2+\alpha_3}}{\eta_1^{3+\alpha_1+\alpha_2+\alpha_3}} \sum_{i=1}^{3} \operatorname*{Res}_{s=1+\beta_i} D_3\Big(s, \frac{\lambda_2}{\eta_2}, -\boldsymbol{\beta}\Big) \times \sum_{n=1}^{\infty} A_3\Big(n, \frac{\lambda_1}{\eta_1}, \boldsymbol{\alpha}\Big) \mathcal{F}_1^{+}(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

for

$$\lambda_1 = \frac{\bar{q}(u_2\bar{x}-u_1)a_1}{\delta(a_1,u_2c)}, \qquad \eta_1 = \frac{u_2c}{(a_1,u_2c)},$$
$$\lambda_2 = \frac{\bar{q}(u_1x-u_2)a_2}{\delta(a_2,u_1c)}, \qquad \eta_2 = \frac{u_1c}{(a_2,u_1c)}, \qquad u_i = \frac{a_ib_i}{(a_1b_1,a_2b_2)},$$

and

$$\mathcal{F}_1^{+}(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_0^\infty \int_0^\infty \frac{1}{y^{1/2}} \frac{z^{s-1}}{z^{1/2}} V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2y, a_2^3b_2^2z; q) J_{k-1}\bigg(\frac{4\pi\sqrt{a_2ya_1z}}{c\delta q}\bigg) f\Big(\frac{y}{N}\Big) f\Big(\frac{z}{M}\Big) \times i^{-k}\, \mathrm{e}\bigg(\frac{a_1^2b_1y}{c\delta q a_2b_2} + \frac{a_2^2b_2z}{c\delta q a_1b_1}\bigg) U_3\Big(\frac{\pi^3 ny}{\eta_1^3}; \boldsymbol{\alpha}\Big)\, dy\, dz.$$

We first note that the contribution from the terms $a_1^3b_1^2y \gg q^{3/2+\varepsilon}$ or $a_2^3b_2^2z \gg q^{3/2+\varepsilon}$ can be bounded by q^{-A} for any A due to the factor $V_{\alpha,\beta}(a_1^3b_1^2y, a_2^3b_2^2z; q)$. So from now on we assume $a_1^3b_1^2N \ll q^{3/2+\varepsilon}$ and $a_2^3b_2^2M \ll q^{3/2+\varepsilon}$.

Moreover, the dyadic sum over M and N contains only $\ll \log^2 q$ terms, so it suffices to prove that

$$E^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a},\boldsymbol{b},M,N) \ll a_1^{1/2} q^{3/4+\varepsilon} \tag{8-1}$$

for fixed $\boldsymbol{a}, \boldsymbol{b}, M, N$ satisfying $a_1^3b_1^2N \ll q^{3/2+\varepsilon}$ and $a_2^3b_2^2M \ll q^{3/2+\varepsilon}$. On a first reading, the reader may set $a_1 = a_2 = b_1 = b_2 = 1$ as this simplifies the notation without substantially changing the calculation.

We now write

$$E^{+}_{1,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a},\boldsymbol{b},M,N) = H_1 + H_2, \tag{8-2}$$

where H_1 is the contribution from the sum over $n \le \eta_1^3 q^{\varepsilon}/N$, and H_2 is the rest.

8A1. *Bounding H_1.* By (B-6), $U_3(\pi^3 ny/\eta_1^3) \ll q^\varepsilon$ when $n \ll \eta_1^3 q^\varepsilon/N$. This and (5-6) gives us that

$$\mathcal{F}_1^+(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta}) \ll M^{1/2}N^{1/2}q^\varepsilon \min\biggl\{\biggl(\frac{\sqrt{a_1a_2MN}}{c\delta q}\biggr)^{-1/2}, \biggl(\frac{\sqrt{a_1a_2MN}}{c\delta q}\biggr)^{k-1}\biggr\}.$$

Then, from Lemma 8.1, Lemma B.2, (B-1), and the fact that $(a_2, u_1c) \le a_2$ and $1/(a_1, u_2c) \le 1$, we obtain that H_1 is bounded by

$$\begin{aligned} &M^{1/2}N^{1/2}q^\varepsilon \sum_{\delta<C}\sum_{c<C/\delta}\frac{1}{\eta_1^3\eta_2} \\ &\qquad\times \sum_{n\ll\eta_1^3q^\varepsilon/N}\ \sum_{h|\eta_1,\, h|n^2} \eta_1^{3/2}\sqrt{h}\min\biggl\{\biggl(\frac{\sqrt{a_1a_2MN}}{c\delta q}\biggr)^{-1/2}, \biggl(\frac{\sqrt{a_1a_2MN}}{c\delta q}\biggr)^{k-1}\biggr\} \\ &\qquad\qquad \ll M^{5/4}N^{1/4}q^\varepsilon \frac{a_1^{3/4}a_2^{7/4}u_2^{3/2}}{u_1q^{3/2}} \ll q^{3/4+\varepsilon}, \end{aligned}$$

as desired. In the above, we have used $(a_1b_1, a_2b_2) \ge 1$, $a_1^3b_1^2N \ll q^{3/2+\varepsilon}$, and $a_2^3b_2^2M \ll q^{3/2+\varepsilon}$.

8A2. *Bounding H_2.* We start by rewriting $\mathcal{F}_1^+(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta})$ as

$$\int_{-\infty}^{\infty}\int_{(1/\log q)} V_1(s,t)\int_0^\infty z^{\beta_1}F_{s-it}\Bigl(\frac{z}{M}\Bigr)\,\mathrm{e}\Bigl(\frac{a_2^2b_2z}{c\delta qa_1b_1}\Bigr)\mathcal{I}(n,z)\,dz\,\frac{ds}{s}\,dt,$$

where $V_1(s,t) := V_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, s, t, M, N)$ is defined as

$$\begin{aligned} &\frac{1}{2\pi i}\Bigl(\frac{q}{4\pi^2}\Bigr)^{\delta(\boldsymbol{\alpha},\boldsymbol{\beta})}\Bigl(\frac{a_2^3b_2^2}{a_1^3b_1^2}\Bigr)^{it} G\bigl(\tfrac12+s; \boldsymbol{\alpha}+it, \boldsymbol{\beta}+it\bigr)H(s;\boldsymbol{\alpha},\boldsymbol{\beta}) \\ &\qquad\qquad \times \Bigl(\frac{q}{4\pi^2}\Bigr)^{3s} M^{-1/2-s+it}N^{-1/2-s-it}, \end{aligned} \tag{8-3}$$

and $\mathcal{I}(n,z) := \mathcal{I}_{\boldsymbol{\alpha}}(\boldsymbol{a}, \boldsymbol{b}, N, n, z, c, \delta)$ is defined as

$$\int_0^\infty F_{s+it}\Bigl(\frac{y}{N}\Bigr)\,\mathrm{e}\Bigl(\frac{a_1^2b_1y}{c\delta qa_2b_2}\Bigr)J_{k-1}\Bigl(\frac{4\pi\sqrt{a_2ya_1z}}{c\delta q}\Bigr)U_3\Bigl(\frac{\pi^3ny}{\eta_1^3};\boldsymbol{\alpha}\Bigr)\,dy, \tag{8-4}$$

and $F_v(x) := f(x)/x^{1/2+v}$. Note that the j-th derivative, $F_{s\pm it}^{(j)}(y/N) = O(|t|^jN^{-j})$.

Note that the trivial bound for $\mathcal{F}_1^+(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta})$ is

$$\mathcal{F}_1^+(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta}) \ll \sqrt{MN}(qn)^\varepsilon. \tag{8-5}$$

There are two cases to consider: $c > \dfrac{8\pi\sqrt{a_1a_2MN}}{\delta q}$ and $c \le \dfrac{8\pi\sqrt{a_1a_2MN}}{\delta q}$.

Case 1. $$c > \frac{8\pi\sqrt{a_1a_2MN}}{\delta q}.$$

By (5-5), (B-7), and $\pi^3 ny/\eta_1^3 \gg q^\varepsilon$, we can rewrite $\mathcal{I}(n,z)$ as

$$\sum_{j=1}^{K} \Big(\frac{\pi^3 ny}{\eta_1^3}\Big)^{(\beta_1+\beta_2+\beta_3)/3} \Big(\frac{\eta_1}{\pi n^{1/3}N^{1/3}}\Big)^j \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!(\ell+k-1)!} \\ \times \int_0^\infty \mathscr{F}_j(y,z,\ell)\, \mathrm{e}\Big(\frac{a_1^2 b_1 y}{c\delta q a_2 b_2}\Big)\Big\{c_j\, \mathrm{e}\Big(\frac{3n^{1/3}y^{1/3}}{\eta_1}\Big) + d_j\, \mathrm{e}\Big(-\frac{3n^{1/3}y^{1/3}}{\eta_1}\Big)\Big\}\, dy \\ + O(q^{-\varepsilon(K+1)}),$$

where c_j, d_j are some constants, and

$$\mathscr{F}_j(y,z,\ell) = F_{s+it}\Big(\frac{y}{N}\Big)\Big(\frac{2\pi\sqrt{a_2 y a_1 z}}{c\delta q}\Big)^{2\ell+k-1}\Big(\frac{y}{N}\Big)^{-j/3}$$

is supported on $y \in [N, 2N]$. Moreover,

$$\frac{\partial^i \mathscr{F}_j(y,z,\ell)}{\partial y^i} \ll \frac{1}{2^{2\ell}}\frac{|t|^i}{N^i} \quad \text{and} \quad \mathscr{F}_j(y,z,\ell) \ll 1.$$

Thus, picking K large enough so that $q^{-\varepsilon(K+1)}$ is negligible, it suffices to bound integrals of the form

$$\int_0^\infty \mathscr{F}_j(y,z,\ell)\, \mathrm{e}(\theta_z(y,n))\, dy,$$

where

$$\theta_z(y,n) = \pm\frac{3n^{1/3}y^{1/3}}{\eta_1} + By \quad \text{and} \quad B = \frac{a_1^2 b_1}{c\delta q a_2 b_2}.$$

Taking the derivative of $\theta_z(y,n)$ with respect to y, we have that

$$\theta_z'(y,n) = B \pm \frac{n^{1/3}}{y^{2/3}\eta_1}.$$

When $n \ge 64(B\eta_1)^3 N^2$ or $n \le \frac{1}{4}(B\eta_1)^3 N^2$, since $n \gg \eta_1^3 q^\varepsilon/N$ we have

$$|\theta_z'(y,n)| \gg \frac{n^{1/3}}{y^{2/3}\eta_1} \gg \frac{q^\varepsilon}{N}.$$

Thus, integrating by parts many times shows that the contribution from these terms is negligible. Therefore we only consider the contribution from when $\frac{1}{4}(B\eta_1)^3N^2 \le n \le 64(B\eta_1)^3N^2$. Note however that

$$(B\eta_1)^3 N^2 \ll \frac{(a_1^2 b_1)^3 N^2}{\delta^3 q^3} \ll \frac{q^\varepsilon}{\delta^3},$$

and that there are no terms of this form unless $N \gg q^{3/2}/(a_1^2 b_1)^{3/2}$ and $\delta \ll q^\varepsilon$. From (8-5), trivially $\mathcal{F}_1^+(c\delta, n; \boldsymbol{\alpha}, \boldsymbol{\beta}) = O(M^{1/2}N^{1/2}(nq)^\varepsilon)$. Hence the contribution to H_2 from these terms is bounded by

$$M^{1/2}N^{1/2}q^{\varepsilon}\sum_{\delta<q^{\varepsilon}}\sum_{c\gg\sqrt{a_1a_2MN}/\delta q}\frac{1}{\eta_1^{3/2}\eta_2}\sum_{h|\eta_1}\sqrt{h}\sum_{\substack{n\ll q^{\varepsilon}\\ h^2|n}}\frac{\eta_1}{n^{1/3}N^{1/3}} \ll a_1^{5/4}b_1^{1/2}a_2^{3/4}M^{1/4}N^{1/4}q^{\varepsilon}\ll a_1^{1/2}q^{3/4+\varepsilon},$$

similar to before.

Case 2. $$c\le\frac{8\pi\sqrt{a_1a_2MN}}{\delta q}.$$

By (5-4), we write $\mathcal{I}(n,z)$ as

$$\int_0^{\infty}\Big\{R^+(y,z)+R^-(y,z)\Big\}\frac{\sqrt{c\delta q}}{\pi(a_1a_2yz)^{1/4}}U_3\Big(\frac{\pi^3ny}{\eta_1^3};\boldsymbol{\alpha}\Big)\,\mathrm{e}\Big(\frac{a_1^2b_1y}{c\delta qa_2b_2}\Big)\,dy,$$

where

$$R^{\pm}(y,z):=F_{s+it}\Big(\frac{y}{N}\Big)W^{\pm}\Big(\frac{4\pi\sqrt{a_1a_2yz}}{c\delta q}\Big)\,\mathrm{e}\Big(\pm\Big(\frac{2\sqrt{a_1a_2yz}}{c\delta q}-\frac{k}{4}+\frac{1}{8}\Big)\Big),$$

and $W^+=W$, $W^-=\overline{W}$.

Similar to Case 1, we explicitly write $U_3(\pi^3ny/\eta_1^3;\boldsymbol{\alpha})$ as in (B-7) so it suffices to bound

$$\sum_{j=1}^{K}\Big(\frac{\pi^3ny}{\eta_1^3}\Big)^{(\beta_1+\beta_2+\beta_3)/3}\Big(\frac{\eta_1}{\pi n^{1/3}N^{1/3}}\Big)^j\frac{1}{N^{1/4}}\frac{\sqrt{c\delta q}}{\pi(a_1a_2z)^{1/4}} \times\int_0^{\infty}\mathscr{H}_j^{\pm}(y,z)\Big\{c_j\,\mathrm{e}\Big(\frac{3n^{1/3}y^{1/3}}{\eta_1}\Big)+d_j\,\mathrm{e}\Big(-\frac{3n^{1/3}y^{1/3}}{\eta_1}\Big)\Big\}\,\mathrm{e}\Big(\frac{a_1^2b_1y}{c\delta qa_2b_2}\Big) \times\mathrm{e}\Big(\pm\Big(\frac{2\sqrt{a_1a_2yz}}{c\delta q}-\frac{k}{4}+\frac{1}{8}\Big)\Big)\,dy+O\big(q^{-\varepsilon(K+1)}\big),$$

where

$$\mathscr{H}_j^{\pm}(y,z)=\frac{F_{s+it}(y/N)W^{\pm}(4\pi\sqrt{a_1a_2yz}/c\delta q)}{(y/N)^{j/3+1/4}}$$

is supported on $y\in[N,2N]$. Note that

$$\frac{\partial^{(i)}\mathscr{H}_j^{\pm}(y,z)}{\partial y^i}\ll_{j,i}\frac{|t|^i}{N^i}\quad\text{and}\quad\mathscr{H}_j^{\pm}(y,z)\ll 1.$$

Thus, the integration over y is of the form

$$\int_0^{\infty}\mathscr{H}_j^{\pm}(y,z)\,\mathrm{e}\big(g_z(y,n)\big)\,dy,$$

where

$$g_z(y,n)=\pm\frac{3n^{1/3}y^{1/3}}{\eta_1}\pm\Bigl(2A\sqrt{y}+\frac{k}{4}-\frac{1}{8}\Bigr)+By,\quad A=\frac{\sqrt{a_1a_2z}}{c\delta q},\quad B=\frac{a_1^2b_1}{c\delta q a_2b_2}.$$

Differentiating $g_z(y,n)$ with respect to y, we have

$$g_z'(y,n)=\pm\frac{n^{1/3}}{y^{2/3}\eta_1}\pm\frac{A}{y^{1/2}}+B.$$

When $a_2^{3/2}b_2M^{1/2}\ge 4a_1^{3/2}b_1N^{1/2}$, it follows that

$$\frac{A}{y^{1/2}}\ge\frac{A}{y^{1/2}}-B\ge\frac{1}{2}\frac{A}{y^{1/2}}\quad\text{and}\quad\frac{3}{2}\frac{A}{y^{1/2}}\ge\frac{A}{y^{1/2}}+B\ge\frac{A}{y^{1/2}}.$$

Therefore

$$\frac{1}{2}\frac{A}{y^{1/2}}\le\left|\pm\frac{A}{y^{1/2}}+B\right|\le\frac{3}{2}\frac{A}{y^{1/2}}.$$

When $n\ge 54(A\eta_1)^3N^{1/2}$ or $n\le\frac{1}{64}(A\eta_1)^3N^{1/2}$, since $n\gg\eta_1^3q^\varepsilon/N$ we have that

$$|g_z'(y,n)|\gg\frac{n^{1/3}}{y^{2/3}\eta_1}\gg\frac{q^\varepsilon}{N}.$$

Integrating by parts many times shows that these terms are negligible. We then consider only the terms where $\frac{1}{64}(A\eta_1)^3N^{1/2}\le n\le 54(A\eta_1)^3N^{1/2}$. Note however that

$$(A\eta_1)^3N^{1/2}\ll\left(\frac{\sqrt{a_1a_2M}u_2c}{c\delta q}\right)^3N^{1/2}\ll\left(\frac{\sqrt{a_1}}{\delta q^{1/4}}\right)^3N^{1/2}\ll\frac{q^\varepsilon}{\delta^3},$$

and that the left side is only $\gg 1$ if $N\gg q^{3/2}/a_1^3$ and $\delta\ll q^\varepsilon$.

By (8-5), the contribution of $\mathcal{F}_1^+(c\delta,n;\boldsymbol{\alpha},\boldsymbol{\beta})$ to the terms in this range is $O(M^{1/2}N^{1/2}(nq)^\varepsilon)$. So the contribution to H_2 from these terms is bounded by

$$M^{1/2}N^{1/2}q^\varepsilon\sum_{\delta<q^\varepsilon}\sum_{c\ll\sqrt{a_1a_2MN}/(\delta q)}\frac{1}{\eta_1^{3/2}\eta_2}\sum_{h\mid\eta_1}\sqrt{h}\sum_{\substack{n\ll q^\varepsilon\\ h^2\mid n}}\Bigl(\frac{\eta_1}{n^{1/3}N^{1/3}}\Bigr)\frac{\sqrt{c\delta q}}{(a_1a_2MN)^{1/4}}$$
$$\ll a_1^{5/4}b_1^{1/2}a_2^{3/4}M^{1/4}N^{1/4}q^\varepsilon\ll a_1^{1/2}q^{3/4+\varepsilon},$$

which suffices.

When $4a_2^{3/2}b_2M^{1/2}\le a_1^{3/2}b_1N^{1/2}$, we have that $\frac{1}{2}B<B-A/y^{1/2}<B$ and $\frac{3}{2}B>Ay^{1/2}+B>B$. By the same arguments as in Case 1, the range of n that should be considered is of the size $(B\eta_1)^3N^2$ and give a contribution to H_2 bounded by $a_1^{1/2}q^{3/4+\varepsilon}$.

When $\frac{1}{4}a_1^{3/2}b_1N^{1/2}<a_2^{3/2}b_2M^{1/2}<4a_1^{3/2}b_1N^{1/2}$, we have that $A/y^{1/2}\asymp B$, and so the range of n that should be considered is of the size $(A\eta_1)^3N^{1/2}$ by the same arguments as above. Hence the contribution from these terms to H_2 is $O\big(a_1^{1/4}q^{3/4+\varepsilon}\big)$.

This completes the proof of Proposition 6.2 for $\mathcal{E}_1^+(q;\boldsymbol{\alpha},\boldsymbol{\beta})$. The same proof applies to bound $\mathcal{E}_i^\pm(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ for $i=2,3,4$.

8B. ***Bounding*** $\boldsymbol{\mathcal{E}_5^+(q;\alpha,\beta)}$**.** We first recall that

$$\mathcal{E}_5^+(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{2\pi}{q}\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2\ge 1\\(a_1a_2b_1b_2,q)=1}}\frac{\mathscr{B}(a_1,b_1;\boldsymbol{\alpha})}{a_1^{3/2}b_1}\frac{\mathscr{B}(a_2,b_2;-\boldsymbol{\beta})}{a_2^{3/2}b_2}\times\sum_M^d\sum_N^d E_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}^+(\boldsymbol{a},\boldsymbol{b},M,N),$$

where

$$\begin{aligned}E_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}^+(\boldsymbol{a},\boldsymbol{b},M,N):=&\sum_{c<C}\ \sideset{}{^*}\sum_{x\ (\mathrm{mod}\ \delta c)}\frac{\mathcal{T}_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}^+(c,x)}{c}\\=&\sum_{\delta<C}\frac{1}{\delta}\sum_{c<C/\delta}\ \sideset{}{^*}\sum_{\substack{x\ (\mathrm{mod}\ \delta c)\\(u_1x-u_2,c\delta)=\delta}}\frac{\mathcal{T}_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}^+(c\delta,x)}{c},\end{aligned}$$

$$\mathcal{T}_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}^+(c\delta,x):=\frac{\pi^{3+\alpha_1-\beta_1+\alpha_2-\beta_2+\alpha_3-\beta_3}}{\eta_1^{3+\alpha_1+\alpha_2+\alpha_3}\eta_2^{3-\beta_1-\beta_2-\beta_3}}\sum\sum_{n,m\ge 1}A_3\Big(n,\frac{\lambda_1}{\eta_1},\boldsymbol{\alpha}\Big)\times A_3\Big(m,\frac{\lambda_2}{\eta_2},-\boldsymbol{\beta}\Big)\mathcal{F}_5^+(c\delta,n,m;\boldsymbol{\alpha},\boldsymbol{\beta});$$

for

$$\lambda_1=\frac{\bar q(u_2\bar x-u_1)a_1}{\delta(a_1,u_2c)},\quad \eta_1=\frac{u_2c}{(a_1,u_2c)},\qquad u_i=\frac{a_ib_i}{(a_1b_1,a_2b_2)},$$
$$\lambda_2=\frac{\bar q(u_1x-u_2)a_2}{\delta(a_2,u_1c)},\quad \eta_2=\frac{u_1c}{(a_2,u_1c)},$$

and

$$\mathcal{F}_5^+(c\delta,n,m,\boldsymbol{\alpha},\boldsymbol{\beta})=\int_0^\infty\int_0^\infty\frac{1}{y^{1/2}}\frac{1}{z^{1/2}}V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3b_1^2y,a_2^3b_2^2z;q)\times J_{k-1}\Big(\frac{4\pi\sqrt{a_2ya_1z}}{c\delta q}\Big)f\Big(\frac{y}{N}\Big)f\Big(\frac{z}{M}\Big)i^{-k}\,\mathrm{e}\Big(\frac{a_1^2b_1y}{c\delta qa_2b_2}+\frac{a_2^2b_2z}{c\delta qa_1b_1}\Big)\times U_3\Big(\frac{\pi^3ny}{\eta_1^3};\boldsymbol{\alpha}\Big)U_3\Big(\frac{\pi^3mz}{\eta_2^3};-\boldsymbol{\beta}\Big)\,dy\,dz.$$

The proofs in this section are very similar to the ones in the previous section. Previously, we had one sum over n and now we have a double sum over m and n which can be treated in a similar manner. To be precise, we begin by dividing

$E^{+}_{5,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a},\boldsymbol{b},M,N)$ into $\sum_{i=1}^{4} E^{+}_{5,i}(\boldsymbol{a},\boldsymbol{b},M,N)$, where

$$E^{+}_{5,i}(\boldsymbol{a},\boldsymbol{b},M,N) := E^{+}_{5,i,\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{a},\boldsymbol{b},M,N)$$

is the contribution from case (i) below:

(1) $n \ll \dfrac{\eta_1^3 q^{\varepsilon}}{N}$ and $m \ll \dfrac{\eta_2^3 q^{\varepsilon}}{M}$;

(2) $n \gg \dfrac{\eta_1^3 q^{\varepsilon}}{N}$ and $m \ll \dfrac{\eta_2^3 q^{\varepsilon}}{M}$;

(3) $n \ll \dfrac{\eta_1^3 q^{\varepsilon}}{N}$ and $m \gg \dfrac{\eta_2^3 q^{\varepsilon}}{M}$;

(4) $n \gg \dfrac{\eta_1^3 q^{\varepsilon}}{N}$ and $m \gg \dfrac{\eta_2^3 q^{\varepsilon}}{M}$.

By symmetry, the treatment for cases (2) and (3) is the same, so we show only the second case.

Similar to Section 8A, the contribution from the terms $a_1^3 b_1^2 y \gg q^{3/2+\varepsilon}$ or $a_2^3 b_2^2 z \gg q^{3/2+\varepsilon}$ can be bounded by q^{-A} due to the factor $V_{\boldsymbol{\alpha},\boldsymbol{\beta}}(a_1^3 b_1^2 y, a_2^3 b_2^2 z; q)$. Thus it suffices to prove that

$$E^{+}_{5,i}(\boldsymbol{a},\boldsymbol{b},M,N) \ll a_1^{1/2} a_2^{1/2} q^{3/4+\varepsilon}, \tag{8-6}$$

for fixed $\boldsymbol{a},\boldsymbol{b},M,N$ satisfying

$$a_1^3 b_1^2 N \ll q^{3/2+\varepsilon} \quad \text{and} \quad a_2^3 b_2^2 M \ll q^{3/2+\varepsilon}.$$

In fact, we prove the stronger bound $E^{+}_{5,i}(\boldsymbol{a},\boldsymbol{b},M,N) \ll a_1^{1/2} a_2^{1/2} q^{1/2+\varepsilon}$.

8B1. *Bounding $E^{+}_{5,1}(\boldsymbol{a},\boldsymbol{b},M,N)$.* For this case, we have $U_3(\pi^3 ny/\eta_1^3;\boldsymbol{\alpha}) \ll q^{\varepsilon}$ and $U_3(\pi^3 mz/\eta_2^3;-\boldsymbol{\beta}) \ll q^{\varepsilon}$ by (B-6). Similar to the arguments in Section 8A1, from Lemma 8.1, Lemma B.2, and (5-6), we have that for $k \geq 5$,

$$\begin{aligned}
&E^{+}_{5,1}(\boldsymbol{a},\boldsymbol{b},M,N)\\
&\quad \ll M^{1/2}N^{1/2}q^{\varepsilon} \sum_{\delta<C} \sum_{c<C/\delta} \frac{1}{\eta_1^3\eta_2^3} \sum_{n \ll \eta_1^3 q^{\varepsilon}/N} \sum_{h_1|\eta_1,\, h_1|n^2} \eta_1^{3/2}\sqrt{h_1}\\
&\qquad \times \sum_{m \ll \eta_2^3 q^{\varepsilon}/M} \sum_{h_2|\eta_2,\, h_2|m^2} \eta_2^{3/2}\sqrt{h_2} \min\left\{ \left(\frac{\sqrt{a_1a_2MN}}{c\delta q}\right)^{-1/2}, \left(\frac{\sqrt{a_1a_2MN}}{c\delta q}\right)^{k-1} \right\}
\end{aligned}$$

$$\ll M^{-1/2}N^{-1/2}q^{\varepsilon}\sum_{\delta<C}\Bigg\{\sum_{\sqrt{a_1a_2MN}/(q\delta)\ll c<C/\delta}\eta_1^{3/2}\eta_2^{3/2}\bigg(\frac{\sqrt{a_1a_2MN}}{c\delta q}\bigg)^{k-1} + \sum_{c\ll\sqrt{a_1a_2MN}/(q\delta)}\eta_1^{3/2}\eta_2^{3/2}\bigg(\frac{\sqrt{a_1a_2MN}}{c\delta q}\bigg)^{-1/2}\Bigg\}$$

$$\ll M^{3/2}N^{3/2}q^{\varepsilon}\frac{a_1^2a_2^2u_1^{3/2}u_2^{3/2}}{q^4}\ll q^{1/2+\varepsilon}.$$

8B2. *Bounding* $E^{+}_{5,2}(\boldsymbol{a},\boldsymbol{b},M,N)$. We can write $\mathcal{F}^{+}_{5}(c\delta,n,m;\boldsymbol{\alpha},\boldsymbol{\beta})$ as

$$\int_{-\infty}^{\infty}\int_{(1/\log q)}V_1(s,t)\int_0^{\infty}F_{s-it}\Big(\frac{z}{M}\Big)\,\mathrm{e}\Big(\frac{a_2^2b_2z}{c\delta qa_1b_1}\Big)U_3\Big(\frac{\pi^3mz}{\eta_2^3};-\boldsymbol{\beta}\Big)\mathcal{I}(n,z)\,dz\,\frac{ds}{s}\,dt,$$

where $V_1(s,t)$ and $\mathcal{I}(n,z)$ are defined as in (8-3) and (8-4), respectively, and $F_v(x)=1/x^{1/2+v}f(x)$. Note that $F^{(j)}_{s\pm it}(y/N)\ll|t|^jN^{-j}$.

The integration over z can be bounded trivially, and the sum over m, h_2 can be treated in the same way as in Section 8B1. For the integration over y, we argue as in Cases 1 and 2 of Section 8A2 and obtain that $E^{+}_{5,2}(\boldsymbol{a},\boldsymbol{b},M,N)\ll a_1^{1/2}q^{1/2+\varepsilon}$.

8B3. *Bounding* $E^{+}_{5,4}(\boldsymbol{a},\boldsymbol{b},M,N)$. We split into two cases as follows.

Case 1.
$$c>\frac{8\pi\sqrt{a_1a_2MN}}{\delta q}.$$

We use (5-5) and (B-7), and the integral that we consider is of the form

$$\int_0^{\infty}\int_0^{\infty}G(y,z)\,\mathrm{e}\bigg(\frac{a_1^2b_1y}{c\delta qa_2b_2}+\frac{a_2^2b_2z}{c\delta qa_1b_1}\pm\frac{3n^{1/3}y^{1/3}}{\eta_1}\pm\frac{3m^{1/3}z^{1/3}}{\eta_2}\bigg)\,dy\,dz,$$

where $\partial^j\partial^iG(y,z)/\partial y^j\partial z^i\ll 1/(N^jM^i)$, $G(x,y)\ll 1$, and it is supported in $[N,2N]\times[M,2M]$. Therefore, the integration over y, z above is $O(MN)$.

By the same arguments as Case 1 of Section 8A2, it is sufficient to consider when $c_1(B_1\eta_1)^3N^2\ll n\ll c_2(B_1\eta_1)^3N^2$ and $c_1(B_2\eta_2)^3M^2\ll m\ll c_2(B_2\eta_2)^3M^2$, where c_1, c_2 are some constants, $B_1=a_1^2b_1/(c\delta qa_2b_2)$, and $B_2=a_2^2b_2/(c\delta qa_1b_1)$, since the terms outside these ranges give negligible contribution from integration by parts many times. By the same arguments as in Section 8A,

$$(B\eta_1)^3N^2\ll\frac{(a_1^2b_1)^3N^2}{\delta^3q^3}\ll\frac{q^{\varepsilon}}{\delta^3}\quad\text{and}\quad(B\eta_2)^3M^2\ll\frac{(a_2^2b_2)^3M^2}{\delta^3q^3}\ll\frac{q^{\varepsilon}}{\delta^3}.$$

So there are no terms of this form unless $N\gg q^{3/2}/(a_1^2b_1)^{3/2}$, $M\gg q^{3/2}/(a_2^2b_2)^{3/2}$, and $\delta\ll q^{\varepsilon}$. We then obtain that the contribution from these terms to $E^{+}_{5,4}(\boldsymbol{a},\boldsymbol{b},M,N)$

is bounded by

$$M^{1/2}N^{1/2}q^{\varepsilon}\sum_{\delta<q^{\varepsilon}}\sum_{c\gg\sqrt{a_1a_2MN}/(\delta q)}\frac{1}{\eta_1^{3/2}\eta_2^{3/2}}\sum_{h_1\mid\eta_1}\sqrt{h_1}\sum_{\substack{n\ll q^{\varepsilon}\\ h_1^2\mid n}}\sum_{h_2\mid\eta_1}\sqrt{h_2}\sum_{\substack{m\ll q^{\varepsilon}\\ h_2^2\mid m}}1 \ll a_1^{1/2}a_2^{1/2}q^{1/2+\varepsilon}.$$

Case 2. $$c\le\frac{8\pi\sqrt{a_1a_2MN}}{\delta q}.$$

For this case, we use (5-4), and the integral that we consider is of the form

$$\frac{\eta_1\eta_2}{m^{1/3}n^{1/3}M^{1/3}N^{1/3}}\frac{\sqrt{c\delta q}}{M^{1/4}N^{1/4}(a_1a_2)^{1/4}}\int_0^\infty\int_0^\infty\mathcal{H}(y,z)\,\mathrm{e}(g(y,z,n,m))\,dy\,dz,$$

where
$$\frac{\partial^j\partial^i\mathcal{H}(y,z)}{\partial y^j\partial z^i}\ll 1/(N^jM^i),$$

$\mathcal{H}(x,y)$ is supported in $[N,2N]\times[M,2M]$, and

$$g(y,z,n,m)=\frac{a_1^2b_1y}{c\delta q a_2b_2}+\frac{a_2^2b_2z}{c\delta q a_1b_1}\pm\frac{3n^{1/3}y^{1/3}}{\eta_1}\pm\frac{3m^{1/3}z^{1/3}}{\eta_2}\pm\frac{2\sqrt{a_1a_2yz}}{c\delta q}.$$

We note that the integration over y, z above is $O(MN)$. Hence we obtain that

$$\frac{\partial g(y,z,n,m)}{\partial y}=B_1\pm\frac{n^{1/3}}{y^{2/3}\eta_1}\pm\frac{A_1}{y^{1/2}},$$
$$\frac{\partial g(y,z,n,m)}{\partial z}=B_2\pm\frac{m^{1/3}}{z^{2/3}\eta_2}\pm\frac{A_2}{z^{1/2}},$$

where $A_1=\sqrt{a_1a_2z}/(c\delta q)$ and $A_2=\sqrt{a_1a_2y}/(c\delta q)$. We divide into three cases to consider.

Case 2.1: $a_2^{3/2}b_2M^{1/2}\ge 4a_1^{3/2}b_1N^{1/2}$. For this case, we have that

$$\left|\frac{A_1}{y^{1/2}}\pm B_1\right|\asymp\frac{A_1}{y^{1/2}}\quad\text{and}\quad\left|\frac{A_2}{z^{1/2}}\pm B_2\right|\asymp B_2.$$

Similar to Case 2 of Section 8A2, we consider the ranges $n\asymp(A_1\eta_1)^3N^{1/2}$ and $m\asymp(B_2\eta_2)^3M^2$. By the same arguments as in Section 8A, we note that

$$(A\eta_1^3)N^{1/2}\ll\left(\frac{\sqrt{a_1}}{\delta q^{1/2}}\right)^3N^{1/2}\ll\frac{q^{\varepsilon}}{\delta^3}\quad\text{and}\quad(B\eta_2)^3M^2\ll\frac{(a_2^2b_2)^3M^2}{\delta^3q^3}\ll\frac{q^{\varepsilon}}{\delta^3},$$

and there are no terms of this form unless $N\gg q^{3/2}/a_1^3$, $M\gg q^{3/2}/(a_2^2b_2)^{3/2}$, and $\delta\ll q^{\varepsilon}$. Hence the contribution from these terms to $E_{5,4}^{+}(\boldsymbol{a},\boldsymbol{b},M,N)$ is

$$M^{1/6}N^{1/6}\sum_{\delta<q^\varepsilon}\sum_{c\ll\sqrt{a_1a_2MN}/(\delta q)}\frac{1}{\eta_1^{1/2}\eta_2^{1/2}}\sum_{h_1|\eta_1}\sqrt{h_1}\sum_{\substack{n\ll q^\varepsilon\\ h_1^2|n}}\frac{1}{n^{1/3}}\sum_{h_2|\eta_1}\sqrt{h_2}\sum_{m\ll q^\varepsilon h_2^2|m}\frac{1}{m^{1/3}}$$
$$\ll a_1^{1/2}a_2^{1/2}q^{1/2+\varepsilon}.$$

Case 2.2: $a_1^{3/2}b_1N^{1/2}\ge 4a_2^{3/2}b_2M^{1/2}$. For this case, we do the same calculation as in Case 2.1 and obtain that the contribution is also $O(a_1^{1/2}a_2^{1/2}q^{1/2+\varepsilon})$.

Case 2.3: $\frac{1}{4}a_1^{3/2}b_1N^{1/2}<a_2^{3/2}b_2M^{1/2}<4a_1^{3/2}b_1N^{1/2}$. For this case, we have that

$$\frac{A_1}{y^{1/2}}\asymp B_1 \quad\text{and}\quad \frac{A_2}{z^{1/2}}\asymp B_2.$$

By similar arguments to Case 2 of Section 8A2, we can focus on the ranges $n\asymp(A_1\eta_1)^3N^{1/2}$ and $m\asymp(A_1\eta_1)^3N^{1/2}$. The contribution from these terms to $E^+_{5,4}(\boldsymbol{a},\boldsymbol{b},M,N)$ is then $\ll a_1^{1/2}a_2^{1/2}q^{1/2+\varepsilon}$.

9. Conclusion of the proof of Theorem 2.5

Recall that from Section 4 and (4-1), we want to evaluate

$$H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{M}_6(q)=\mathscr{M}_1(q;\boldsymbol{\alpha},\boldsymbol{\beta})+\mathscr{M}_1(q;\boldsymbol{\beta},\boldsymbol{\alpha}).$$

By (4-2), we see that $\mathscr{M}_1(q;\boldsymbol{\alpha},\boldsymbol{\beta})=\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})+\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$, and in Lemma 4.1, we showed that

$$\mathscr{D}(q;\boldsymbol{\alpha},\boldsymbol{\beta})=H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\int_{-\infty}^{\infty}\mathcal{M}(q;\boldsymbol{\alpha}+it,\boldsymbol{\beta}+it)\,dt+O(q^{-3/4+\varepsilon}),$$

which is one of the twenty main terms of the asymptotic formula. Then we decomposed $\mathscr{K}(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ as $\mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta})+\mathscr{K}_E(q;\boldsymbol{\alpha},\boldsymbol{\beta})$. We proved in Section 5A that $\mathscr{K}_E(q;\boldsymbol{\alpha},\boldsymbol{\beta})\ll q^{-1/2+\varepsilon}$, and then using the Voronoi summation formula, we extracted another nine main terms of the asymptotic formula from $\mathscr{K}_M(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ with an error term $O(q^{-1/4+\varepsilon})$ (see Propositions 6.1 and 6.2, Section 7, Section 8, and Appendix A). As briefly discussed in Section 4, those terms correspond to $\mathcal{M}(q;\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it)$, where π is the transposition (α_i,β_j) for $i=1,2,3$ in $S_6/S_3\times S_3$. Hence $\mathscr{M}_1(q;\boldsymbol{\alpha},\boldsymbol{\beta})$ gives ten main terms the desired asymptotic formula, and similarly the remaining ten terms come from $\mathscr{M}_1(q;\boldsymbol{\beta},\boldsymbol{\alpha})$.

Therefore, combining everything together, we have that

$$H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\mathcal{M}_6(q)$$
$$=H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\int_{-\infty}^{\infty}\sum_{\pi\in S_6/S_3\times S_3}\mathcal{M}(q;\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it)\,dt+O(q^{-1/4+\varepsilon}).$$

If $|\alpha_i-\beta_j|\gg q^{-\varepsilon}$ for all $1\le i,\ j\le 3$, then $H(0;\boldsymbol{\alpha},\boldsymbol{\beta})\gg q^{-\varepsilon}$ and we immediately get

$$\mathcal{M}_6(q)=\int_{-\infty}^{\infty}\sum_{\pi\in S_6/S_3\times S_3}\mathcal{M}(q;\pi(\boldsymbol{\alpha})+it,\pi(\boldsymbol{\beta})+it)\,dt+O(q^{-1/4+\varepsilon}).$$

However, since all expressions above, including the term bounded by $O(q^{-1/2+\varepsilon})$, are analytic in the α_i and β_j, we see that this in fact holds in general.

Appendix A: Comparing the main term of R_{α_1,β_1} and $\mathcal{M}(q;\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))$

To finish the proof of Proposition 6.1, we show that the local factor at prime p of the Euler product of $\zeta(1-\alpha_1+\beta_1)\mathcal{M}_{\alpha_1,\beta_1}(0)$ is the same as the one in $\mathcal{AZ}\big(\tfrac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\big)$, where

$$(\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))=(\beta_1,\alpha_1,\alpha_2;\alpha_1,\beta_2,\beta_3),$$

and $\mathcal{M}_{\alpha_1,\beta_1}(s)$ is defined as in (7-10). To simplify the presentation, we work within the ring of formal Dirichlet series, so that we need not worry about convergence issues in this section. Indeed, if we show that $\zeta(1-\alpha_1+\beta_1)\mathcal{M}_{\alpha_1,\beta_1}(0)$ is the same as $\mathcal{AZ}\big(\tfrac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\big)$ as formal series, then they must have the same region of absolute convergence. Thus, as analytic functions, they agree on the region of absolute convergence, and so must be the same by analytic continuation. Note that we have already verified that there is a nonempty open region of absolute convergence at the end of Section 7.

For notational convenience, $\boldsymbol{\alpha}_{2,3}=(\alpha_2,\alpha_3)$, and $-\boldsymbol{\beta}_{2,3}=(-\beta_2,-\beta_3)$ in this section.

A1. ***Euler product at prime p of $\mathcal{AZ}\big(\tfrac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\big)$.*** We start by rearranging the sums in $\mathcal{AZ}(s;\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))$ by the same method as in (2-9). When

$$\operatorname{Re}(s+\beta_1+\alpha_2+\alpha_3),\ \operatorname{Re}(s-\alpha_1-\beta_2-\beta_3)>1,$$

we recall that from (2-13) and (2-17), $\mathcal{AZ}(s;\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta}))$ is

$$\sum\sum\sum\sum\sum\sum_{\substack{a_1,b_1,a_2,b_2,m,n\ge 1\\ a_1n=a_2m\\ a_1b_1=a_2b_2\\ (a_i,q)=(b_j,q)=1}}\frac{\mathscr{B}(a_1,b_1;\pi(\boldsymbol{\alpha}))}{(a_1b_1)^{2s}}\frac{\mathscr{B}(a_2,b_2;-\pi(\boldsymbol{\beta}))}{(a_2b_2)^{2s}}\times\frac{\sigma_3(n;\pi(\boldsymbol{\alpha}))\sigma_3(m;-\pi(\boldsymbol{\beta}))}{(a_1n)^s(a_2m)^s}.$$

Using Lemma 2.1, the proof of (2-9), and the fact that

$$\sigma_3(a;\alpha_1,\alpha_2,\alpha_3)=\sum_{df=a}d^{-\alpha_1}\sigma_2(f;\alpha_2,\alpha_3),\tag{A-1}$$

we see after a change of variables that

$$\mathcal{AZ}\big(\tfrac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\big) = \zeta(1-\alpha_1+\beta_1)\sum\sum\sum\sum_{\substack{d_1,d_2,e_1,e_2\ge 1\\ d_1e_1=d_2e_2\\ (d_ie_i,q)=1}} \frac{1}{d_1^{1+\alpha_2+\alpha_3}d_2^{1-\beta_2-\beta_3}}\frac{1}{e_1^{1+\beta_1}e_2^{1-\alpha_1}}\mathcal{J}(e_1,e_2), \tag{A-2}$$

where

$$\mathcal{J}(e_1,e_2)=\sum\sum_{j_1,j_2\ge 1}\frac{\sigma_2(j_1e_1;\boldsymbol{\alpha}_{2,3})\sigma_2(j_2e_2;-\boldsymbol{\beta}_{2,3})(j_1,j_2)^{1-\alpha_1+\beta_1}}{j_1^{1-\alpha_1}j_2^{1+\beta_1}}. \tag{A-3}$$

Because both $\zeta(1-\alpha_1+\beta_1)\mathcal{M}_{\alpha_1,\beta_1}(0)$ and $\mathcal{AZ}\big(\frac{1}{2};\pi(\boldsymbol{\alpha}),\pi(\boldsymbol{\beta})\big)$ have the factor $\zeta(1-\alpha_1+\beta_1)$, it suffices to consider only the local factor at prime p of the sum over d_i, e_i in (A-2). For $p\ne q$, this is

$$\sum\sum\sum\sum_{\substack{\delta_1,\delta_2,\epsilon_1,\epsilon_2\ge 0\\ \delta_1+\epsilon_1=\delta_2+\epsilon_2}}\frac{1}{p^{D+\epsilon_1+\epsilon_2+\epsilon_1\beta_1-\epsilon_2\alpha_1}}\sum_{k\ge 0}\frac{\mathcal{J}_p(\epsilon_1,\epsilon_2,k)}{p^k}, \tag{A-4}$$

where $p^{\delta_i}\parallel d_i$, $p^{\epsilon_i}\parallel e_i$, $p^{\iota_i}\parallel j_i$,

$$D:=\delta_1+\delta_2+\delta_1(\alpha_2+\alpha_3)-\delta_2(\beta_2-\beta_3), \tag{A-5}$$

and

$$\begin{aligned}\mathcal{J}_p(\epsilon_1,\epsilon_2,k) &:= \sigma_2(p^{k+\epsilon_1};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{k+\epsilon_2};-\boldsymbol{\beta}_{2,3})\\ &\quad+\sum_{0\le\iota_1<k}\frac{\sigma_2(p^{\iota_1+\epsilon_1};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{k+\epsilon_2};-\boldsymbol{\beta}_{2,3})}{p^{\beta_1(k-\iota_1)}}\\ &\quad+\sum_{0\le\iota_2<k}\frac{\sigma_2(p^{k+\epsilon_1};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{\iota_2+\epsilon_2};-\boldsymbol{\beta}_{2,3})}{p^{-\alpha_1(k-\iota_2)}}.\end{aligned} \tag{A-6}$$

For $p=q$, we have that $\delta_i=\epsilon_i=0$, and the local factor at p is

$$\sum_{k\ge 0}\frac{\mathcal{J}_p(0,0,k)}{p^k}. \tag{A-7}$$

We also comment here that when $\alpha_i=\beta_i=0$ for $i=1,2,3$, using $\sigma_2(p^k)=k+1$ in (A-4), (A-6), (A-7) and some straightforward calculation, we derive that the local factor at $p\ne q$ of $\mathcal{AZ}\big(\frac{1}{2};0,0\big)$ is

$$\Big(1-\frac{1}{p}\Big)^{-9}C_p,$$

where C_p is defined in (1-4), and the local factor at q is

$$\Big(1-\frac{1}{q}\Big)^{-5}\Big(1+\frac{4}{q}+\frac{1}{q^2}\Big).$$

This explains the presence of the arithmetic factor in our Conjecture 1.1, as in the work [Conrey et al. 2005].

A2. ***The Euler product at p of $\mathcal{M}_{\alpha_1,\beta_1}(0)$.*** By the definition of $\mathscr{B}(a,b;\alpha_1,\alpha_2,\alpha_3)$ in (2-6), $\mathcal{G}(c,\boldsymbol{a},\boldsymbol{b})$ in (7-9), $\sum^{\#}\mathscr{G}_{\boldsymbol{a},\boldsymbol{b}}(s;h,b,c,\gamma,g)$ in (7-6), (A-1), and a change of variables, we obtain that $\mathcal{M}_{\alpha_1,\beta_1}(0)$ can be rewritten as

$$\sum\sum\sum\sum_{\substack{d_1,f_1,d_2,f_2\ge 1\\ (d_if_i,q)=1}} \frac{1}{d_1^{1-\alpha_1-\beta_1+\alpha_2+\alpha_3}d_2^{1+\alpha_1+\beta_1-\beta_2-\beta_3}}\frac{1}{f_1^{1-\beta_1}f_2^{1+\alpha_1}}$$
$$\times\sum_{\substack{h\ge 1\\ h\mid u_1u_2}}\frac{\mu(h)}{h^{\alpha_1-\beta_1}}\sum_{\substack{b\ge 1\\ (b,u_1u_2)=1}}\frac{\mu(b)}{b^{\alpha_1-\beta_1}}\sum_{c\ge 1}\frac{c}{c^{\alpha_1-\beta_1}}\sum_{\gamma\mid c}\frac{1}{\gamma^{\alpha_1-\beta_1}}\sum_{g\mid(c/\gamma)}\frac{\mu(g)}{g^{\alpha_1-\beta_1}}\prod_{\substack{p\mid c\\ p\nmid bh\gamma}}\Big(1-\frac{1}{p}\Big)$$
$$\times\sum_{n\ge 1}\sum_{a_1\mid f_1}\frac{\mu(a_1)}{a_1^{\alpha_2+\alpha_3}}\bigg(\frac{(a_1,u_2cb)}{a_1u_2cbn}\bigg)^{1-\alpha_1}\sigma_2\bigg(\frac{u_2cbn}{(a_1,u_2cb)};\boldsymbol{\alpha}_{2,3}\bigg)\sigma_2\bigg(\frac{f_1}{a_1};\boldsymbol{\alpha}_{2,3}\bigg)$$
$$\times\sum_{m\ge 1}\sum_{a_2\mid f_2}\frac{\mu(a_2)}{a_2^{-\beta_2-\beta_3}}\bigg(\frac{(a_2,u_1cb)}{a_2u_1cbm}\bigg)^{1+\beta_1}\sigma_2\bigg(\frac{u_1cbm}{(a_2,u_1cb)};-\boldsymbol{\beta}_{2,3}\bigg)\sigma_2\bigg(\frac{f_2}{a_2};-\boldsymbol{\beta}_{2,3}\bigg).$$

In Section 7B, $u_i=a_ib_i/(a_1b_1,a_2b_2)$, but after changing variables, we write that $u_i=f_id_i/(f_1d_1,f_2d_2)$. By comparing Euler products, we can show that

$$\sum_{n\ge 1}\sum_{a_1\mid f_1}\frac{\mu(a_1)}{a_1^{\alpha_2+\alpha_3}}\bigg(\frac{(a_1,u_2cb)}{a_1u_2cbn}\bigg)^{1-\alpha_1}\sigma_2\bigg(\frac{u_2cbn}{(a_1,u_2cb)};\boldsymbol{\alpha}_{2,3}\bigg)\sigma_2\bigg(\frac{f_1}{a_1};\boldsymbol{\alpha}_{2,3}\bigg)$$
$$=\sum_{n'\ge 1}\frac{\sigma_2(f_1u_2cbn';\boldsymbol{\alpha}_{2,3})}{(u_2cbn')^{1-\alpha_1}}.$$

We also have a similar expression for the sum over m and a_2. Hence we can write $\mathcal{M}_{\alpha_1,\beta_1}(0)$ as

$$\sum\sum\sum\sum_{\substack{d_1,f_1,d_2,f_2\ge 1\\ (d_if_i,q)=1}} \frac{1}{d_1^{1-\alpha_1-\beta_1+\alpha_2+\alpha_3}d_2^{1+\alpha_1+\beta_1-\beta_2-\beta_3}}\frac{1}{f_1^{1-\beta_1}u_2^{1-\alpha_1}f_2^{1+\alpha_1}u_1^{1+\beta_1}}$$
$$\times\sum_{\substack{h\ge 1\\ h\mid u_1u_2}}\frac{\mu(h)}{h^{\alpha_1-\beta_1}}\sum_{\substack{b\\ (b,u_1u_2)=1}}\frac{\mu(b)}{b^2}\sum_{c}\frac{1}{c}\sum_{\gamma\mid c}\frac{1}{\gamma^{\alpha_1-\beta_1}}\sum_{g\mid(c/\gamma)}\frac{\mu(g)}{g^{\alpha_1-\beta_1}}\prod_{\substack{p\mid c\\ p\nmid bh\gamma}}\Big(1-\frac{1}{p}\Big)$$
$$\times\sum_{n,m\ge 1}\frac{\sigma_2(f_1u_2cbn;\boldsymbol{\alpha}_{2,3})}{n^{1-\alpha_1}}\frac{\sigma_2(f_2u_1cbm;-\boldsymbol{\beta}_{2,3})}{m^{1+\beta_1}}. \tag{A-8}$$

We note here that $d_1f_1u_2=d_2f_2u_1$ by the definition of u_1,u_2. Next, we consider the local factor at $p \neq q$ of (A-8), which is of the form

$$\sum\sum\sum\sum_{\substack{\delta_1,\delta_2,\xi_1,\xi_2\ge 0\\ \delta_1+\ell_1=\delta_2+\ell_2}} \frac{1}{p^{D'(\ell_1,\ell_2)+\ell_2\beta_1-\ell_1\alpha_1-(\xi_1+\xi_2)(\beta_1-\alpha_1)}}\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2), \tag{A-9}$$

where $p^{\delta_i} \parallel d_i$, $p^{\xi_i} \parallel f_i$, $p^{\upsilon_i} \parallel u_i$, $\ell_1 = \xi_1+\upsilon_2$, $\ell_2 = \xi_2+\upsilon_1$, $\min\{\upsilon_1,\upsilon_2\} = 0$, $D'(\ell_1,\ell_2) = D+(\delta_2-\delta_1)(\alpha_1+\beta_1)+\ell_1+\ell_2$, and D is defined in (A-5). We examine $\mathscr{L}_p(d_1,d_2,f_1,f_2)$ below, but before that analysis, we need the following two lemmas.

Lemma A.1. *The contribution to the local factor at p from*

$$\sum_{\gamma\mid c}\frac{1}{\gamma^{\alpha_1-\beta_1}}\sum_{g\mid(c/\gamma)}\frac{\mu(g)}{g^{\alpha_1-\beta_1}}\prod_{\substack{p\mid c\\ p\nmid bh\gamma}}\Big(1-\frac{1}{p}\Big)$$

is 1 *if $p\nmid c$ or $p\mid bh$. Otherwise, it is*

$$1-\frac{1}{p}+\frac{1}{p^{1+\alpha_1-\beta_1}}.$$

Proof. For $p\nmid c$ or $p\mid bh$, the contribution to the local factor is 1 because

$$\sum_{\gamma\mid c}\frac{1}{\gamma^{\alpha_1-\beta_1}}\sum_{g\mid(c/\gamma)}\frac{\mu(g)}{g^{\alpha_1-\beta_1}} = \sum_{a\mid c}\frac{1}{a^{\alpha_1-\beta_1}}\sum_{g\mid a}\mu(g) = 1.$$

Now suppose $p\mid c$ and $p\nmid bh$. Below we write $p^{c_p}\parallel c$ and $p^{\gamma_p}\parallel \gamma$. Then the contribution to the local factor at p is

$$\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\Big(1-\frac{1}{p}\Big)+\sum_{1\le\gamma_p<c_p}\frac{1}{p^{\gamma_p(\alpha_1-\beta_1)}}\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)+\frac{1}{p^{c_p(\alpha_1-\beta_1)}} = 1-\frac{1}{p}+\frac{1}{p^{1+\alpha_1-\beta_1}}. \qquad \square$$

Lemma A.2. *Let*

$$\mathscr{P}(\ell_1,\ell_2) := \sum\sum\sum_{c_p\ge 0,\ n_p,m_p\ge 0}\frac{1}{p^{c_p}}\frac{\sigma_2(p^{\ell_1+n_p+c_p};\boldsymbol{\alpha}_{2,3})}{p^{n_p(1-\alpha_1)}}\frac{\sigma_2(p^{\ell_2+m_p+c};-\boldsymbol{\beta}_{2,3})}{p^{m_p(1+\beta_1)}}. \tag{A-10}$$

Then

$$\mathscr{P}(\ell_1,\ell_2) = \sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k}+\frac{p^{\alpha_1-\beta_1}}{p^2}\mathscr{P}(\ell_1+1,\ell_2+1),$$

where $\mathcal{J}_p(\ell_1,\ell_2,k)$ is defined as in (A-6).

Proof. We have

$$\begin{aligned}
\mathscr{P}(\ell_1,\ell_2) &= \sum_{c_p\ge 0} \frac{\sigma_2(p^{\ell_1+c_p};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{\ell_2+c_p};-\boldsymbol{\beta}_{2,3})}{p^{c_p}} \\
&\qquad + \sum_{c_p\ge 0,}\sum_{n_p\ge 1} \frac{\sigma_2(p^{\ell_1+n_p+c_p};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{\ell_2+c_p};-\boldsymbol{\beta}_{2,3})}{p^{c_p+n_p-n_p\alpha_1}} \\
&\qquad + \sum_{c_p\ge 0,}\sum_{m_p\ge 1} \frac{\sigma_2(p^{\ell_1+c_p};\boldsymbol{\alpha}_{2,3})\sigma_2(p^{\ell_2+m_p+c_p};-\boldsymbol{\beta}_{2,3})}{p^{c_p+m_p+m_p\beta_1}} \\
&\qquad + \frac{p^{\alpha_1-\beta_1}}{p^2}\mathscr{P}(\ell_1+1,\ell_2+1) \\
&= \sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k} + \frac{p^{\alpha_1-\beta_1}}{p^2}\mathscr{P}(\ell_1+1,\ell_2+1),
\end{aligned}$$

after some arrangement. □

Now we examine $\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2)$, which we separate into two cases below.

Case 1: $\boldsymbol{\delta_1+\xi_1=\delta_2+\xi_2}$. For this case, we have $\upsilon_1=\upsilon_2=0$, so $\xi_i=\ell_i$. Hence $u_1u_2=1$ and $p\nmid h$. From Lemma A.1, we then obtain that

$$\begin{aligned}
&\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2) \\
&\quad = \mathscr{P}(\ell_1,\ell_2) + \frac{1}{p^2}(p^{\beta_1-\alpha_1}-1)\mathscr{P}(\ell_1+1,\ell_2+1) - \frac{1}{p^2}\mathscr{P}(\ell_1+1,\ell_2+1).
\end{aligned}$$

From Lemma A.2, $\mathscr{L}_p(d_1,d_2,f_1,f_2)$ can be written as

$$\begin{aligned}
&\sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k} + \frac{1}{p^2}(p^{\alpha_1-\beta_1}-2+p^{\beta_1-\alpha_1})\mathscr{P}(\ell_1+1,\ell_2+1) \\
&\quad = \sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k} + \frac{1}{p^2}(p^{\alpha_1-\beta_1}-2+p^{\beta_1-\alpha_1})\sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1+1,\ell_2+1,k)}{p^k} \\
&\qquad + \frac{p^{\alpha_1-\beta_1}}{p^4}(p^{\alpha_1-\beta_1}-2+p^{\beta_1-\alpha_1})\mathscr{P}(\ell_1+2,\ell_2+2) \\
&\quad = \sum_{k\ge 0}\frac{1}{p^k}\bigg\{\mathcal{J}_p(\ell_1,\ell_2,k) + (p^{\alpha_1-\beta_1}-2+p^{\beta_1-\alpha_1}) \\
&\qquad \times \sum_{1\le m_p\le \lfloor k/2\rfloor} p^{(m_p-1)(\alpha_1-\beta_1)}\mathcal{J}_p(\ell_1+m_p,\ell_2+m_p,k-2m_p)\bigg\}.
\end{aligned}$$

Case 2: $\boldsymbol{\delta_1+\xi_1\ne\delta_2+\xi_2}$. For this case, $\upsilon_1+\upsilon_2\ge 1$. So $p\mid u_1u_2$, and $b_p=0$, where $p^{b_p}\parallel b$. By Lemma A.1 and A.2, we have

$$
\begin{aligned}
&\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2)\\
&\quad=\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\mathscr{P}(\ell_1,\ell_2)-\frac{1}{p^2}\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\mathscr{P}(\ell_1+1,\ell_2+1)\\
&\quad=\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\bigg\{\sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k}+\frac{1}{p^2}(p^{\alpha_1-\beta_1}-1)\mathscr{P}(\ell_1+1,\ell_2+1)\bigg\}\\
&\quad=\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\bigg\{\sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1,\ell_2,k)}{p^k}+\frac{1}{p^2}(p^{\alpha_1-\beta_1}-1)\sum_{k\ge 0}\frac{\mathcal{J}_p(\ell_1+1,\ell_2+1,k)}{p^k}\\
&\qquad\qquad+\frac{p^{\alpha_1-\beta_1}}{p^4}(p^{\alpha_1-\beta_1}-1)\mathscr{P}(\ell_1+2,\ell_2+2)\bigg\}\\
&\quad=\Big(1-\frac{1}{p^{\alpha_1-\beta_1}}\Big)\sum_{k\ge 0}\frac{1}{p^k}\bigg\{\mathcal{J}_p(\ell_1,\ell_2,k)\\
&\qquad+(p^{\alpha_1-\beta_1}-1)\sum_{1\le m_p\le\lfloor k/2\rfloor}p^{(m_p-1)(\alpha_1-\beta_1)}\mathcal{J}_p(\ell_1+m_p,\ell_2+m_p,k-2m_p)\bigg\}.
\end{aligned}
$$

From both cases, we obtain that (A-9) is, say,

$$
\begin{aligned}
&\sum\sum\sum\sum_{\substack{\delta_1,\delta_2,\ell_1,\ell_2\ge 0\\ \delta_1+\ell_1=\delta_2+\ell_2}}\frac{1}{p^{D'(\ell_1,\ell_2)-\ell_1\beta_1+\ell_2\alpha_1}}\mathscr{L}_p(\delta_1,\delta_2,\ell_1,\ell_2)\\
&\qquad+\sum\sum\sum\sum_{\substack{\delta_1,\delta_2,\ell_1,\ell_2\ge 0\\ \delta_1+\ell_1=\delta_2+\ell_2}}\bigg(\sum_{\substack{0\le\xi_2<\ell_2\\ \xi_1=\ell_1}}\frac{\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2)}{p^{(\ell_2-\ell_1)\beta_1-\xi_2(\beta_1-\alpha_1)}}\\
&\qquad+\sum_{\substack{0\le\xi_1<\ell_1\\ \xi_2=\ell_2}}\frac{\mathscr{L}_p(\delta_1,\delta_2,\xi_1,\xi_2)}{p^{(\ell_2-\ell_1)\alpha_1-\xi_1(\beta_1-\alpha_1)}}\bigg)\frac{1}{p^{D'(\ell_1,\ell_2)}}=:\sum\sum_{\delta_1,\delta_2\ge 0}\mathcal{S}_p(\delta_1,\delta_2).
\end{aligned}
\tag{A-11}
$$

For fixed δ_1,δ_2, where $\delta_2\ge\delta_1$, we rearrange the term $\mathcal{S}_p(\delta_1,\delta_2)$ and obtain that

$$
\begin{aligned}
&S_p(\delta_1,\delta_2)\\
&\ =\sum\sum_{\substack{\epsilon_1,\epsilon_2\ge 0\\ \delta_1+\epsilon_1=\delta_2+\epsilon_2}}\sum_{k\ge 0}\frac{\mathcal{J}_p(\epsilon_1,\epsilon_2,k)}{p^{D'(\epsilon_1,\epsilon_2)+k}}\bigg\{\frac{1}{p^{(\epsilon_2-\epsilon_1)\alpha_1}}+\frac{1}{p^{(\epsilon_2-\epsilon_1)\beta_1}}-\frac{1}{p^{-\epsilon_1\beta_1+\epsilon_2\alpha_1}}\\
&\quad+(p^{\alpha_1-\beta_1}-2+p^{\beta_1-\alpha_1})\sum_{1\le\ell_2\le\epsilon_2}\frac{p^{(\ell_2-1)(\alpha_1-\beta_1)}}{p^{-(\epsilon_1-\ell_2)\beta_1+(\epsilon_2-\ell_2)\alpha_1}}\\
&\quad+(p^{\alpha_1-\beta_1}-1)\sum_{1\le\ell_2\le\epsilon_2}p^{(\ell_2-1)(\alpha_1-\beta_1)}\bigg(\frac{1}{p^{(\epsilon_2-\epsilon_1)\alpha_1}}+\frac{1}{p^{(\epsilon_2-\epsilon_1)\beta_1}}-\frac{2p^{\ell_2(\alpha_1-\beta_1)}}{p^{-\epsilon_1\beta_1+\epsilon_2\alpha_1}}\bigg)\bigg\}\\
&\ =\sum\sum_{\substack{\epsilon_1,\epsilon_2\ge 0\\ \delta_1+\epsilon_1=\delta_2+\epsilon_2}}\sum_{k\ge 0}\frac{\mathcal{J}_p(\epsilon_1,\epsilon_2,k)}{p^{D+\epsilon_1+\epsilon_2+k+\epsilon_1\beta_1-\epsilon_2\alpha_1}}.
\end{aligned}
\tag{A-12}
$$

By a similar calculation, $\mathcal{S}_p(\delta_1, \delta_2)$ yields the same value for $\delta_1 > \delta_2$. In summary, from (A-9), (A-11), and (A-12), the local factor at p of $\mathcal{M}_{\alpha_1,\beta_1}$ is

$$\sum_{\substack{\epsilon_1,\epsilon_2\geq 0\\ \delta_1+\epsilon_1=\delta_2+\epsilon_2}} \sum_{k\geq 0} \frac{\mathcal{J}_p(\epsilon_1,\epsilon_2,k)}{p^{D+\epsilon_1+\epsilon_2+k+\epsilon_1\beta_1-\epsilon_2\alpha_1}},$$

which is the same as the local factor at p of $\mathcal{AZ}\big(\frac{1}{2}; \pi(\boldsymbol{\alpha}), \pi(\boldsymbol{\beta})\big)$ in (A-4), as desired.

For $p = q$, we use similar arguments, with $\delta_i = \epsilon_i = 0$, so that the Euler factor is

$$\sum_{k\geq 0} \frac{\mathcal{J}_p(0,0,k)}{p^k},$$

which is the same as the local factor at q of $\mathcal{AZ}\big(\frac{1}{2}; \pi(\boldsymbol{\alpha}), \pi(\boldsymbol{\beta})\big)$ in (A-7). This completes the proof of Proposition 6.1.

Appendix B. Voronoi summation

In this section, we state the Voronoi summation formula for the shifted k-divisor function defined in (2-5). The proof of this formula is essentially the same as the proof by Ivić of the Voronoi summation formula for the k-divisor function, so we state the results and refer the reader to [Ivić 1997] for detailed proofs.

Let ω be a smooth compactly supported function. For $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$, let $\sigma_k(n; \boldsymbol{\alpha}) = \sigma_k(n; \alpha_1, \dots, \alpha_k)$. Define

$$S\Big(\frac{a}{c}, \boldsymbol{\alpha}\Big) = \sum_{n=1}^{\infty} \sigma_k(n; \boldsymbol{\alpha})\, \mathrm{e}\Big(\frac{an}{c}\Big)\omega(n),$$

where $(a, c) = 1$, and the Mellin transform of ω,

$$\tilde{\omega}(s) = \int_0^\infty \omega(x) x^{s-1}\, dx.$$

Since ω was chosen to be from the Schwarz class, $\tilde{\omega}$ is entire and decays rapidly on vertical lines. We have the Mellin inversion formula

$$\omega(n) = \frac{1}{2\pi i}\int_{(c)} \tilde{\omega}(s)\, \frac{ds}{n^s},$$

where c is any vertical line. Let

$$A_k\Big(m, \frac{a}{c}, \boldsymbol{\alpha}\Big) = \frac{1}{2} \sum_{\substack{m_1,\dots,m_k\geq 1\\ m_1\cdots m_k=m}} m_1^{\alpha_1}\cdots m_k^{\alpha_k} \\ \times \sum_{a_1 \ (\mathrm{mod}\ c)} \cdots \sum_{a_k \ (\mathrm{mod}\ c)} \Big\{\mathrm{e}\Big(\frac{aa_1\cdots a_k + \boldsymbol{a}\cdot\boldsymbol{m}}{c}\Big) + \mathrm{e}\Big(\frac{-aa_1\cdots a_k + \boldsymbol{a}\cdot\boldsymbol{m}}{c}\Big)\Big\}, \tag{B-1}$$

and

$$B_k\Big(m, \frac{a}{c}, \boldsymbol{\alpha}\Big) = \frac{1}{2} \sum_{\substack{m_1,\dots,m_k \geq 1 \\ m_1\cdots m_k = m}} m_1^{\alpha_1} \cdots m_k^{\alpha_k} \times \sum_{a_1 \pmod c} \cdots \sum_{a_k \pmod c} \Big\{ \mathrm{e}\Big(\frac{aa_1\cdots a_k + \boldsymbol{a}\cdot\boldsymbol{m}}{c}\Big) - \mathrm{e}\Big(\frac{-aa_1\cdots a_k + \boldsymbol{a}\cdot\boldsymbol{m}}{c}\Big)\Big\}. \tag{B-2}$$

Moreover, we define

$$G_k(s, n, \boldsymbol{\alpha}) = \frac{c^{ks}}{\pi^{ks} n^s} \Bigg(\prod_{\ell=1}^{k} \frac{\Gamma\big(\frac{1}{2}(s - \alpha_\ell)\big)}{\Gamma\big(\frac{1}{2}(1 - s + \alpha_\ell)\big)} \Bigg),$$

$$H_k(s, n, \boldsymbol{\alpha}) = \frac{c^{ks}}{\pi^{ks} n^s} \Bigg(\prod_{\ell=1}^{k} \frac{\Gamma\big(\frac{1}{2}(1 + s - \alpha_\ell)\big)}{\Gamma\big(\frac{1}{2}(2 - s + \alpha_\ell)\big)} \Bigg),$$

and for $0 < \sigma < \frac{1}{2} - \frac{1}{k} + \frac{1}{k}\sum_{\ell=1}^{k} \mathrm{Re}(\alpha_\ell)$, we let

$$\begin{aligned} U_k(x; \boldsymbol{\alpha}) &= \frac{1}{2\pi i} \int_{(\sigma)} \prod_{\ell=1}^{k} \frac{\Gamma\big(\frac{1}{2}(s - \alpha_\ell)\big)}{\Gamma\big(\frac{1}{2}(1 - s + \alpha_\ell)\big)} \frac{ds}{x^s}, \\ V_k(x; \boldsymbol{\alpha}) &= \frac{1}{2\pi i} \int_{(\sigma)} \prod_{\ell=1}^{k} \frac{\Gamma\big(\frac{1}{2}(1 + s - \alpha_\ell)\big)}{\Gamma\big(\frac{1}{2}(2 - s + \alpha_\ell)\big)} \frac{ds}{x^s}. \end{aligned} \tag{B-3}$$

We note that by Stirling's formula, both integrals for U_k and V_k are absolutely convergent.

Finally, we define the Dirichlet series to be

$$D_k\Big(s, \frac{a}{c}, \boldsymbol{\alpha}\Big) = \sum_{n} \frac{\sigma_k(n; \boldsymbol{\alpha}) e(an/c)}{n^s},$$

which converges absolutely for $\mathrm{Re}\, s > 1$. We have that

$$\begin{aligned} D_k\Big(s, \frac{a}{c}, \boldsymbol{\alpha}\Big) &= \sum_{n_1,\dots,n_k} \frac{e(an_1 n_2 \cdots n_k / c)}{n_1^{s+\alpha_1} \cdots n_k^{s+\alpha_k}} \\ &= \frac{1}{c^{ks+\alpha_1+\cdots+\alpha_k}} \sum_{a_1,\dots,a_k \pmod c} e\Big(\frac{aa_1\cdots a_k}{c}\Big) \prod_{j=1}^{k} \zeta\Big(s + \alpha_j, \frac{a_j}{c}\Big), \end{aligned} \tag{B-4}$$

where $\zeta(s, a/c)$ is the Hurwitz zeta function defined for $\mathrm{Re}\, s > 1$ as

$$\zeta\Big(s, \frac{a}{c}\Big) = \sum_{n=0}^{\infty} \frac{1}{(n + a/c)^s}. \tag{B-5}$$

The Hurwitz zeta function may be analytically continued to all of $\mathbb{C}$ except for a simple pole at $s = 1$. Therefore, $D_k(s, a/c)$ can be analytically continued to all of $\mathbb{C}$ except for a simple pole at $s = 1 - \alpha_j$ for $j = 1, \dots, k$.

Theorem B.1. *With notation as above and* $(a, c) = 1$ *and*

$$\alpha_i \ll \frac{1}{1000k \log(|c| + 100)},$$

we have

$$\begin{aligned} S\Big(\frac{a}{c}, \boldsymbol{\alpha}\Big) = \sum_{\ell=1}^{k} \operatorname*{Res}_{s=1-\alpha_\ell} \tilde{\omega}(s) D_k\Big(s, \frac{a}{c}, \boldsymbol{\alpha}\Big) \\ + \frac{\pi^{k/2+\alpha_1+\cdots+\alpha_k}}{c^{k+\alpha_1+\cdots+\alpha_k}} \sum_{n=1}^{\infty} A_k\Big(n, \frac{a}{c}, \boldsymbol{\alpha}\Big) \int_0^\infty \omega(x) U_k\Big(\frac{\pi^k n x}{c^k}; \boldsymbol{\alpha}\Big)\, dx \\ + i^{3k} \frac{\pi^{k/2+\alpha_1+\cdots+\alpha_k}}{c^{k+\alpha_1+\cdots+\alpha_k}} \sum_{n=1}^{\infty} B_k\Big(n, \frac{a}{c}, \boldsymbol{\alpha}\Big) \int_0^\infty \omega(x) V_k\Big(\frac{\pi^k n x}{c^k}; \boldsymbol{\alpha}\Big)\, dx. \end{aligned}$$

We refer the reader to the proof of Theorem 2 in [Ivić 1997] for details. Next, we collect properties of $A_3(n, a/c, \boldsymbol{\alpha})$, $B_3(n, a/c, \boldsymbol{\alpha})$, $U_3(x; \boldsymbol{\alpha})$, and $V_3(x; \boldsymbol{\alpha})$. These are useful for bounding error terms of $\mathcal{M}_6(q)$.

Lemma B.2. *Let* a, n, γ *be integers such that* $(a, \gamma) = 1$. *Moreover,* $A_3(n, a/c, \boldsymbol{\alpha})$ *and* $B_3(n, a/c, \boldsymbol{\alpha})$ *are defined as in* (B-1) *and* (B-2). *Then*

$$\begin{aligned} \sum_{n_1 n_2 n_3 = n} n_1^{\alpha_1} n_2^{\alpha_2} n_3^{\alpha_3} \sum\sum\sum_{r_1, r_2, r_3 \pmod{\gamma}} \mathrm{e}\Big(\frac{a r_1 r_2 r_3 + r_1 n_1 + r_2 n_2 + r_3 n_3}{\gamma}\Big) \\ = \gamma \sum_{h \mid \gamma,\, h^2 \mid n} h \Delta(n, h, \gamma) S\Big(\frac{n}{h^2}, -\bar{a}, \frac{\gamma}{h}\Big), \end{aligned}$$

where $\Delta(n, h, \gamma)$ *is a divisor function satisfying* $\Delta(n, h, \gamma) \ll (\gamma n)^\varepsilon$. *Moreover,*

$$A_3\Big(n, \frac{a}{\gamma}, \boldsymbol{\alpha}\Big) \ll (\gamma n)^\varepsilon \gamma^{3/2} \sum_{h \mid \gamma,\, h^2 \mid n} \sqrt{h},$$

$$B_3\Big(n, \frac{a}{\gamma}, \boldsymbol{\alpha}\Big) \ll (\gamma n)^\varepsilon \gamma^{3/2} \sum_{h \mid \gamma,\, h^2 \mid n} \sqrt{h}.$$

The proof of this lemma can be found in equations (8.7)–(8.9) in [Ivić 1997].

Lemma B.3. *If* $U(x; \boldsymbol{\alpha}) := U_3(x; \boldsymbol{\alpha})$ *and* $V(x; \boldsymbol{\alpha}) := V_3(x; \boldsymbol{\alpha})$, *as defined in* (B-3), *and*

$$\alpha_i \ll \frac{1}{1000k \log(|c| + 100)},$$

then for any $0<\varepsilon<\frac{1}{6}$ *and* $x>0$, *we have*

$$U(x;\boldsymbol{\alpha})\ll x^{\varepsilon},\qquad V(x;\boldsymbol{\alpha})\ll x^{\varepsilon}. \tag{B-6}$$

Moreover for any fixed integer $K\geq 1$ *and* $x\geq x_0>0$,

$$U(x;\boldsymbol{\alpha})=\sum_{j=1}^{K}\frac{c_j\cos(6x^{1/3})+d_j\sin(6x^{1/3})}{x^{j/3+(\alpha_1+\alpha_2+\alpha_3)/3}}+O\Big(\frac{1}{x^{(K+1)/3+(\alpha_1+\alpha_2+\alpha_3)/3}}\Big), \tag{B-7}$$

$$V(x;\boldsymbol{\alpha})=\sum_{j=1}^{K}\frac{e_j\cos(6x^{1/3})+f_j\sin(6x^{1/3})}{x^{j/3+(\alpha_1+\alpha_2+\alpha_3)/3}}+O\Big(\frac{1}{x^{(K+1)/3+(\alpha_1+\alpha_2+\alpha_3)/3}}\Big), \tag{B-8}$$

with suitable constants $c_j,\dots,f_j$, *and* $c_1=0$, $d_1=-2/\sqrt{3\pi}$, $e_1=-2/\sqrt{3\pi}$, $f_1=0$.

The proof of this lemma is a minor modification of the proof of Lemma 3 in [Ivić 1997].

Acknowledgment

Vorrapan Chandee acknowledges support from the Coordinating Center for Thai Government Science and Technology Scholarship Students (CSTS) and the National Science and Technology Development Agency (NSTDA) grant of 2014. Part of this work was done while Chandee was visiting the Mathematical Institute, University of Oxford; she is grateful for their kind hospitality.

References

[Chandee and Li 2014] V. Chandee and X. Li, "The eighth moment of Dirichlet L-functions", *Adv. Math.* **259** (2014), 339–375. MR Zbl

[Conrey and Ghosh 1998] J. B. Conrey and A. Ghosh, "A conjecture for the sixth power moment of the Riemann zeta-function", *Internat. Math. Res. Notices* 15 (1998), 775–780. MR Zbl

[Conrey and Gonek 2001] J. B. Conrey and S. M. Gonek, "High moments of the Riemann zeta-function", *Duke Math. J.* **107**:3 (2001), 577–604. MR Zbl

[Conrey et al. 2005] J. B. Conrey, D. W. Farmer, J. P. Keating, M. O. Rubinstein, and N. C. Snaith, "Integral moments of L-functions", *Proc. London Math. Soc.* (3) **91**:1 (2005), 33–104. MR Zbl

[Conrey et al. 2012] J. B. Conrey, H. Iwaniec, and K. Soundararajan, "The sixth power moment of Dirichlet L-functions", *Geom. Funct. Anal.* **22**:5 (2012), 1257–1288. MR Zbl

[Diaconu et al. 2003] A. Diaconu, D. Goldfeld, and J. Hoffstein, "Multiple Dirichlet series and moments of zeta and L-functions", *Compositio Math.* **139**:3 (2003), 297–360. MR Zbl

[Djanković 2011] G. Djanković, "The sixth moment of the family of $\Gamma_1(q)$-automorphic L-functions", *Arch. Math.* (*Basel*) **97**:6 (2011), 535–547. MR Zbl

[Gradshteyn and Ryzhik 2007] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., Elsevier, Amsterdam, 2007. MR Zbl

[Harper 2013] A. J. Harper, "Sharp conditional bounds for moments of the Riemann zeta function", preprint, 2013. arXiv

[Ivić 1997] A. Ivić, "On the ternary additive divisor problem and the sixth moment of the zeta-function", pp. 205–243 in *Sieve methods, exponential sums, and their applications in number theory* (Cardiff, 1995), edited by G. R. H. Greaves et al., London Math. Soc. Lecture Note Ser. **237**, Cambridge University Press, 1997. MR Zbl

[Iwaniec and Li 2007] H. Iwaniec and X. Li, "The orthogonality of Hecke eigenvalues", *Compos. Math.* **143**:3 (2007), 541–565. MR Zbl

[Katz and Sarnak 1999] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications **45**, American Mathematical Society, Providence, RI, 1999. MR Zbl

[Keating and Snaith 2000] J. P. Keating and N. C. Snaith, "Random matrix theory and $\zeta(1/2+it)$", *Comm. Math. Phys.* **214**:1 (2000), 57–89. MR Zbl

[Oberhettinger 1972] F. Oberhettinger, *Tables of Bessel transforms*, Springer, Berlin, 1972. MR Zbl

[Rudnick and Soundararajan 2005] Z. Rudnick and K. Soundararajan, "Lower bounds for moments of L-functions", *Proc. Natl. Acad. Sci. USA* **102**:19 (2005), 6837–6838. MR Zbl

[Soundararajan 2009] K. Soundararajan, "Moments of the Riemann zeta function", *Ann. of Math.* (2) **170**:2 (2009), 981–993. MR Zbl

[Titchmarsh 1986] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd ed., Oxford University Press, 1986. MR Zbl

[Watson 1944] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1944. MR Zbl

Communicated by Philippe Michel
Received 2016-02-16 Revised 2016-07-25 Accepted 2016-12-16

vorrapan@buu.ac.th — *Department of Mathematics, Burapha University, 169 Long-Hard Bangsaen rd Saensuk, Mueng 20131, Thailand*

xiannan@math.ksu.edu — *Department of Mathematics, Kansas State University, 138 Cardwell Hall, Manhattan, KS 66506, United States*

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.635

A duality in Buchsbaum rings and triangulated manifolds

Satoshi Murai, Isabella Novik and Ken-ichi Yoshida

Let Δ be a triangulated homology ball whose boundary complex is $\partial\Delta$. A result of Hochster asserts that the canonical module of the Stanley–Reisner ring $\mathbb{F}[\Delta]$ of Δ is isomorphic to the Stanley–Reisner module $\mathbb{F}[\Delta, \partial\Delta]$ of the pair $(\Delta, \partial\Delta)$. This result implies that an Artinian reduction of $\mathbb{F}[\Delta, \partial\Delta]$ is (up to a shift in grading) isomorphic to the Matlis dual of the corresponding Artinian reduction of $\mathbb{F}[\Delta]$. We establish a generalization of this duality to all triangulations of connected orientable homology manifolds with boundary. We also provide an explicit algebraic interpretation of the h''-numbers of Buchsbaum complexes and use it to prove the monotonicity of h''-numbers for pairs of Buchsbaum complexes as well as the unimodality of h''-vectors of barycentric subdivisions of Buchsbaum polyhedral complexes. We close with applications to the algebraic manifold g-conjecture.

1. Introduction

In this paper, we study an algebraic duality of Stanley–Reisner rings of triangulated homology manifolds with nonempty boundary. Our starting point is the following (unpublished) result of Hochster — see [Stanley 1996, Chapter II, §7]. (We defer most definitions until later sections.) Let Δ be a triangulated $(d-1)$-dimensional homology ball whose boundary complex is $\partial\Delta$, let $\mathbb{F}[\Delta]$ be the Stanley–Reisner ring of Δ, and let $\mathbb{F}[\Delta, \partial\Delta]$ be the Stanley–Reisner module of $(\Delta, \partial\Delta)$. (Throughout the paper $\mathbb{F}$ denotes an infinite field.) Hochster's result asserts that the canonical module $\omega_{\mathbb{F}[\Delta]}$ of $\mathbb{F}[\Delta]$ is isomorphic to $\mathbb{F}[\Delta, \partial\Delta]$. In the last decade or so, this result had a lot of impact on the study of face numbers of simplicial complexes, especially in connection with the g-conjecture for spheres; see, for instance, the proof of Theorem 3.1 in the recent survey paper [Swartz 2014].

Murai's research is partially supported by Grant-in-Aid for Scientific Research (C) 25400043.
Novik's research is partially supported by NSF grant DMS-1361423.
Yoshida's research is partially supported by Grant-in-Aid for Scientific Research (C) 25400050.
MSC2010: primary 13F55; secondary 13H10, 05E40, 52B05, 57Q15.
Keywords: triangulated manifolds, Buchsbaum rings, h-vectors, Stanley–Reisner rings.

One numerical consequence of Hochster's result is the following symmetry of h-numbers of homology balls: $h_i(\Delta, \partial\Delta) = h_{d-i}(\Delta)$. The h-numbers are certain linear combinations of the face numbers; they are usually arranged in a vector called the h-vector. In fact, Hochster's result implies a stronger statement: it implies that there is an isomorphism

$$\mathbb{F}[\Delta, \partial\Delta]/\Theta\mathbb{F}[\Delta, \partial\Delta] \cong (\mathbb{F}[\Delta]/\Theta\mathbb{F}[\Delta])^{\vee}(-d), \tag{1}$$

where Θ is a linear system of parameters for $\mathbb{F}[\Delta]$ and $N^{\vee}$ is the (graded) Matlis dual of N (see, e.g., [Murai and Yanagawa 2014, Lemma 3.6]). As the $\mathbb{F}$-dimensions of the i-th graded components of modules in (1) are equal to $h_i(\Delta, \partial\Delta)$ and $h_{d-i}(\Delta)$, respectively, the above-mentioned symmetry, $h_i(\Delta, \partial\Delta) = h_{d-i}(\Delta)$, follows.

Hochster's result was generalized to homology manifolds with boundary by Gräbe [1984]. To state Gräbe's result, we recall the definition of homology manifolds. We denote by $\mathbb{S}^d$ and $\mathbb{B}^d$ the d-dimensional sphere and ball, respectively. A pure d-dimensional simplicial complex Δ is an $\mathbb{F}$-*homology d-manifold without boundary* if the link of each nonempty face τ of Δ has the homology of $\mathbb{S}^{d-|\tau|}$ (over $\mathbb{F}$). An $\mathbb{F}$-*homology d-manifold with boundary* is a pure d-dimensional simplicial complex Δ such that

(i) the link of each nonempty face τ of Δ has the homology of either $\mathbb{S}^{d-|\tau|}$ or $\mathbb{B}^{d-|\tau|}$, and

(ii) the set of all *boundary faces*, that is,

$$\partial\Delta := \left\{\tau \in \Delta : \text{ the link of } \tau \text{ has the same homology as } \mathbb{B}^{d-|\tau|}\right\} \cup \{\varnothing\}$$

is a $(d-1)$-dimensional $\mathbb{F}$-homology manifold without boundary.

A connected $\mathbb{F}$-homology d-manifold with boundary is said to be *orientable* if the top homology $\widetilde{H}_d(\Delta, \partial\Delta)$ is isomorphic to $\mathbb{F}$.

Gräbe proved [1984] that if Δ is an orientable homology manifold with boundary, then $\mathbb{F}[\Delta, \partial\Delta]$ is the canonical module of $\mathbb{F}[\Delta]$. Gräbe also established [1987] a symmetry of h-numbers for such a Δ. While Gräbe's original statement of symmetry is somewhat complicated, it was recently observed by the first two authors [Murai and Novik 2016] that it takes the following simple form when expressed in the language of h''-numbers.

Theorem 1.1. *Let Δ be a connected orientable $\mathbb{F}$-homology $(d-1)$-manifold with nonempty boundary $\partial\Delta$. Then $h''_i(\Delta, \partial\Delta) = h''_{d-i}(\Delta)$ for all $i = 0, 1, \dots, d$.*

The h''-numbers are certain modifications of h-numbers (see Section 3 for their definition). Similarly to the h-numbers, the h''-numbers are usually arranged in a vector, called the h''-vector. For homology manifolds, this vector appears to be a "correct" analog of the h-vector. Indeed, many properties of h-vectors of homology balls and spheres are now known to hold for the h''-vectors of homology manifolds (with and without boundary); see recent survey articles [Klee and Novik 2016; Swartz 2014]. In light of Gräbe's result [1984] and Theorem 1.1, it is natural to ask if Theorem 1.1 can be explained by Matlis duality. The first goal of this paper is to provide such an explanation.

To this end, the key object is the submodule $\Sigma(\Theta; M)$ defined by Goto [1983]. Several definitions are in order. Let $S = \mathbb{F}[x_1, \ldots, x_n]$ be a graded polynomial ring over a field $\mathbb{F}$ with $\deg x_i = 1$ for $i = 1, 2, \ldots, n$. Let M be a finitely generated graded S-module of Krull dimension d and let $\Theta = \theta_1, \ldots, \theta_d$ be a homogeneous system of parameters for M. The module $\Sigma(\Theta; M)$ is defined as follows:

$$\Sigma(\Theta; M) := \Theta M + \sum_{i=1}^{d} ((\theta_1, \ldots, \hat{\theta}_i, \ldots, \theta_d)M :_M \theta_i) \subseteq M.$$

This module was introduced by Goto [1983] and has been used in the study of Buchsbaum local rings. Note that if M is a Cohen–Macaulay module, then $\Sigma(\Theta; M) = \Theta M$. We first show that this submodule is closely related to the h''-vectors. Specifically, we establish the following explicit algebraic interpretation of h''-numbers.

Theorem 1.2. *Let (Δ, Γ) be a Buchsbaum relative simplicial complex of dimension $d-1$ and let Θ be a linear system of parameters for $\mathbb{F}[\Delta, \Gamma]$. Then*

$$\dim_{\mathbb{F}}(\mathbb{F}[\Delta, \Gamma]/\Sigma(\Theta; \mathbb{F}[\Delta, \Gamma]))_j = h''_j(\Delta, \Gamma) \quad \textit{for all } j = 0, 1, \ldots, d.$$

Theorems 1.1 and 1.2 suggest that when Δ is a homology manifold with boundary there might be a duality between the quotients of $\mathbb{F}[\Delta]$ and $\mathbb{F}[\Delta, \partial\Delta]$ by $\Sigma(\Theta; \mathbb{F}[\Delta])$ and $\Sigma(\Theta; \mathbb{F}[\Delta, \partial\Delta])$, respectively. We prove that this is indeed the case. In fact, we prove a more general algebraic result on canonical modules of Buchsbaum graded algebras. Let $\mathfrak{m} = (x_1, \ldots, x_n)$ be the graded maximal ideal of S. If M is a finitely generated graded S-module of Krull dimension d, then the *canonical module* of M is the module

$$\omega_M := (H^d_{\mathfrak{m}}(M))^\vee,$$

where $H^i_{\mathfrak{m}}(M)$ denotes the i-th local cohomology module of M. We prove that the following isomorphism holds for *all* Buchsbaum graded algebras.

Theorem 1.3. *Let $R = S/I$ be a Buchsbaum graded $\mathbb{F}$-algebra of Krull dimension $d \geq 2$, let $\Theta = \theta_1, \ldots, \theta_d \in S$ be a homogeneous system of parameters for R, and let $\delta = \sum_{i=1}^{d} \deg \theta_i$. If* depth $R \geq 2$, *then*

$$\omega_R/\Sigma(\Theta; \omega_R) \cong (R/\Sigma(\Theta; R))^\vee(-\delta).$$

As we mentioned above, if Δ is a connected orientable homology manifold, then (by Gräbe's result) the module $\mathbb{F}[\Delta, \partial\Delta]$ is the canonical module of $\mathbb{F}[\Delta]$; furthermore it is not hard to see that $\mathbb{F}[\Delta]$ satisfies the assumptions of Theorem 1.3. (Indeed, the connectivity of Δ implies that depth $\mathbb{F}[\Delta] \geq 2$, and the fact that $\mathbb{F}[\Delta]$ is Buchsbaum follows from Schenzel's theorem — see Theorem 3.1 below.) Hence we obtain the following corollary that generalizes (1).

Corollary 1.4. *Let Δ be a connected orientable $\mathbb{F}$-homology $(d-1)$-manifold with nonempty boundary $\partial\Delta$, and let Θ be a linear system of parameters for $\mathbb{F}[\Delta]$. Then*

$$\mathbb{F}[\Delta, \partial\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta, \partial\Delta]) \cong (\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]))^{\vee}(-d).$$

We also consider combinatorial and algebraic applications of Theorems 1.2 and 1.3. Specifically, we prove the monotonicity of h''-vectors for pairs of Buchsbaum simplicial complexes, establish the unimodality of h''-vectors of barycentric subdivisions of Buchsbaum polyhedral complexes, provide a combinatorial formula for the a-invariant of Buchsbaum Stanley–Reisner rings, and extend [Swartz 2014, Theorem 3.1] as well as [Böhm and Papadakis 2015, Corollary 4.5], which is related to the sphere g-conjecture, to the generality of the manifold g-conjecture. More precisely, Swartz's result asserts that most bistellar flips when applied to a homology sphere preserve the weak Lefschetz property, while Böhm and Papadakis' result asserts that stellar subdivisions at large-dimensional faces of homology spheres preserve the weak Lefschetz property; we extend both of these results to bistellar flips and stellar subdivisions performed on connected orientable homology manifolds.

The structure of the paper is as follows. In Section 2 we prove Theorem 1.3 (although we defer part of the proof to the Appendix). In Section 3, we study Stanley–Reisner rings and modules of Buchsbaum simplicial complexes. There, after reviewing basics of simplicial complexes and Stanley–Reisner rings and modules, we verify Theorem 1.2 and derive several combinatorial consequences. Section 4 is devoted to applications of our results to the manifold g-conjecture. Finally, in the Appendix, we prove a graded version of [Goto 1983, Proposition 3.6] — a result on which our proof of Theorem 1.3 is based.

2. Duality in Buchsbaum rings

In this section, we prove Theorem 1.3. We start by recalling some definitions and results pertaining to Buchsbaum rings and modules.

Let $S = \mathbb{F}[x_1, \ldots, x_n]$ be a graded polynomial ring with $\deg x_i = 1$, $i = 1, 2, \ldots, n$, and let $\mathfrak{m} = (x_1, \ldots, x_n)$ be the graded maximal ideal of S. Given a graded S-module N, we denote by $N(a)$ the module N with grading shifted by $a \in \mathbb{Z}$, that is, $N(a)_j = N_{a+j}$. If M is a finitely generated graded S-module of Krull dimension d, then a *homogeneous system of parameters* (or h.s.o.p.) for M is a sequence

$\Theta = \theta_1, \dots, \theta_d \in \mathfrak{m}$ of homogeneous elements such that $\dim_{\mathbb{F}} M/\Theta M < \infty$. A sequence $\theta_1, \dots, \theta_r \in \mathfrak{m}$ of homogeneous elements is said to be a *weak M-sequence* if

$$(\theta_1, \dots, \theta_{i-1})M :_M \theta_i = (\theta_1, \dots, \theta_{i-1})M :_M \mathfrak{m}$$

for all $i = 1, 2, \dots, r$. We say that M is *Buchsbaum* if every h.s.o.p. for M is a weak M-sequence.

Let M be a finitely generated graded S-module and let Θ be its h.s.o.p. The function $P_{\Theta,M} : \mathbb{Z} \to \mathbb{Z}$ defined by

$$P_{\Theta,M}(n) := \dim_{\mathbb{F}}(M/(\Theta)^{n+1}M)$$

is called the *Hilbert–Samuel function of M with respect to the ideal* (Θ). It is known that there is a polynomial in n of degree d, denoted by $p_{\Theta,M}(n)$, such that $P_{\Theta,M}(n) = p_{\Theta,M}(n)$ for $n \gg 0$ (see [Bruns and Herzog 1993, Proposition 4.6.2]). The leading coefficient of the polynomial $p_{\Theta,M}(n)$ multiplied by $d!$ is called the *multiplicity of M with respect to the ideal* (Θ), and is denoted by $e_{\Theta}(M)$. We will make use of the following known characterization of the Buchsbaum property; see [Stückrad and Vogel 1986, Theorem I.1.12 and Proposition I.2.6].

Lemma 2.1. *A finitely generated graded S-module M of Krull dimension d is Buchsbaum if and only if, for every h.s.o.p. Θ of M,*

$$\dim_{\mathbb{F}}(M/\Theta M) - e_{\Theta}(M) = \sum_{i=0}^{d-1} \binom{d-1}{i} \dim_{\mathbb{F}} H^i_{\mathfrak{m}}(M).$$

We also recall some known results on canonical modules of Buchsbaum modules. For a finitely generated graded S-module M, the *depth* of M is defined by $\operatorname{depth}(M) := \min\{i : H^i_{\mathfrak{m}}(M) \neq 0\}$.

Lemma 2.2. *Let M be a finitely generated graded S-module of Krull dimension d. If M is Buchsbaum, then the following properties hold:*

(i) *ω_M is Buchsbaum.*

(ii) *$H^i_{\mathfrak{m}}(\omega_M) \cong (H^{d-i+1}_{\mathfrak{m}}(M))^{\vee}$ (as graded modules) for all $i = 2, 3, \dots, d-1$.*

(iii) *If* $\operatorname{depth}(M) \geq 2$, *then $(H^d_{\mathfrak{m}}(\omega_M))^{\vee} \cong M$ (as graded modules).*

(iv) *If Θ is an h.s.o.p. for M, then Θ is also an h.s.o.p. for ω_M and $e_{\Theta}(\omega_M) = e_{\Theta}(M)$.*

See [Stückrad and Vogel 1986, Theorem II.4.9] for (i) and (ii), [Schenzel 1982, Korollar 3.13] or [Aoyama and Goto 1986, (1.16)] for (iii), and [Suzuki 1981, Lemma 2.2] for (iv).

The following theorem is a graded version of [Goto 1983, Proposition 3.6]. The original result by Goto is a statement about Buchsbaum *local rings*. It may be possible to prove Theorem 2.3 in the same way as in [Goto 1983] by replacing rings with modules and by carefully keeping track of grading. However, since we could not find any literature allowing us to easily check this statement, we will provide its proof in the Appendix.

Theorem 2.3. *Let M be a finitely generated graded S-module of Krull dimension $d > 0$. Assume further that M is Buchsbaum and that $\Theta = \theta_1, \dots, \theta_d$ is an h.s.o.p. for M with $\deg \theta_i = \delta_i$. Let $\delta_C := \sum_{i \in C} \delta_i$ for $C \subseteq [d] = \{1, 2, \dots, d\}$. Then*

(i) $\Sigma(\Theta; M)/\Theta M \cong \bigoplus_{C \subsetneq [d]} H_{\mathfrak{m}}^{|C|}(M)(-\delta_C)$, *and*

(ii) *there is an injection* $M/\Sigma(\Theta; M) \to H_{\mathfrak{m}}^d(M)(-\delta_{[d]})$.

We are now in a position to prove Theorem 1.3.

Proof of Theorem 1.3. We first prove that $R/\Sigma(\Theta; R)$ and $\omega_R/\Sigma(\Theta; \omega_R)$ have the same $\mathbb{F}$-dimension. Indeed,

$$\begin{aligned}\dim_{\mathbb{F}}(R/\Sigma(\Theta; R)) &= \dim_{\mathbb{F}}(R/\Theta R) - \dim_{\mathbb{F}}(\Sigma(\Theta; R)/\Theta R)\\ &= e_\Theta(R) + \sum_{i=0}^{d-1} \binom{d-1}{i} \dim_{\mathbb{F}} H_{\mathfrak{m}}^i(R) - \sum_{i=0}^{d-1} \binom{d}{i} \dim_{\mathbb{F}} H_{\mathfrak{m}}^i(R)\\ &= e_\Theta(R) - \sum_{i=1}^{d-1} \binom{d-1}{i-1} \dim_{\mathbb{F}} H_{\mathfrak{m}}^i(R),\end{aligned}$$

where we use Lemma 2.1 and Theorem 2.3(i) for the second equality. Similarly, since ω_R is Buchsbaum, the same computation yields

$$\dim_{\mathbb{F}}(\omega_R/\Sigma(\Theta; \omega_R)) = e_\Theta(\omega_R) - \sum_{i=1}^{d-1} \binom{d-1}{i-1} \dim_{\mathbb{F}} H_{\mathfrak{m}}^i(\omega_R).$$

Now, since $\operatorname{depth}(R) \geq 2$ by the assumptions of the theorem and since $\operatorname{depth}(\omega_R) \geq 2$ always holds (see [Aoyama 1980, Lemma 1]), parts (ii) and (iv) of Lemma 2.2 guarantee that

$$\dim_{\mathbb{F}}(R/\Sigma(\Theta; R)) = \dim_{\mathbb{F}}(\omega_R/\Sigma(\Theta; \omega_R)).$$

Thus, to complete the proof of the statement, it suffices to show that there is a surjection from $R/\Sigma(\Theta; R)(+\delta)$ to $(\omega_R/\Sigma(\Theta; \omega_R))^\vee$. By Theorem 2.3(ii), there is an injection

$$\omega_R/\Sigma(\Theta; \omega_R) \to H_{\mathfrak{m}}^d(\omega_R)(-\delta).$$

Dualizing and using Lemma 2.2(iii), we obtain a surjection

$$R(+\delta) \cong (H_{\mathfrak{m}}^d(\omega_R)(-\delta))^\vee \to (\omega_R/\Sigma(\Theta; \omega_R))^\vee. \tag{2}$$

Since ω_R is an R-module, it follows from the definition of $\Sigma(\Theta; \omega_R)$ that

$$\Sigma(\Theta; R) \cdot \omega_R \subseteq \Sigma(\Theta; \omega_R).$$

Hence

$$\Sigma(\Theta;\, R)\cdot(\omega_R/\Sigma(\Theta;\,\omega_R))=0,$$

which in turn implies

$$\Sigma(\Theta;\, R)\cdot(\omega_R/\Sigma(\Theta;\,\omega_R))^{\vee}=0. \tag{3}$$

Now (2) and (3) put together guarantee the existence of a surjection

$$(R/\Sigma(\Theta;\, R))(+\delta)\to(\omega_R/\Sigma(\Theta;\,\omega_R))^{\vee},$$

as desired. □

Remark 2.4. The above proof also works in the local setting: it shows that if R is a Buchsbaum Noetherian local ring with depth $R\geq 2$ (and if the canonical module of R exists), then $\omega_R/\Sigma(\Theta;\,\omega_R)$ is isomorphic to the Matlis dual of $R/\Sigma(\Theta;\, R)$.

3. Duality in Stanley–Reisner rings of manifolds

In this section, we study Buchsbaum Stanley–Reisner rings and modules. Some objects in this section such as homology groups, Betti numbers, and h'- and h''-numbers depend on the characteristic of $\mathbb{F}$; however, we fix a field $\mathbb{F}$ throughout this section and omit $\mathbb{F}$ from our notation.

We start by reviewing basics of simplicial complexes and Stanley–Reisner rings. A *simplicial complex* Δ on $[n]$ is a collection of subsets of $[n]$ that is closed under inclusion. A *relative simplicial complex* Ψ on $[n]$ is a collection of subsets of $[n]$ with the property that there are simplicial complexes $\Delta\supseteq\Gamma$ such that $\Psi=\Delta\setminus\Gamma$. We identify such a pair of simplicial complexes (Δ,Γ) with the relative simplicial complex $\Delta\setminus\Gamma$. Also, a simplicial complex Δ will be identified with $(\Delta,\varnothing)$. A *face* of (Δ,Γ) is an element of $\Delta\setminus\Gamma$. The *dimension* of a face τ is its cardinality minus one, and the dimension of (Δ,Γ) is the maximal dimension of its faces. A relative simplicial complex is said to be *pure* if all its maximal faces have the same dimension.

We denote by $\widetilde{H}_i(\Delta,\Gamma)$ the i-th reduced homology group of the pair (Δ,Γ) computed with coefficients in $\mathbb{F}$: when $\Gamma\neq\varnothing$, $\widetilde{H}_*(\Delta,\Gamma)$ is the usual relative homology of a pair, and when $\Gamma=\varnothing$, $\widetilde{H}_*(\Delta,\Gamma)=\widetilde{H}_*(\Delta)$ is the reduced homology of Δ. The Betti numbers of (Δ,Γ) are defined by $\widetilde{\beta}_i(\Delta,\Gamma):=\dim_{\mathbb{F}}\widetilde{H}_i(\Delta,\Gamma)$. If Δ is a simplicial complex on $[n]$ and $\tau\in\Delta$ is a face of Δ, then the *link* of τ in Δ is

$$\mathrm{lk}_\Delta(\tau):=\{\sigma\in\Delta:\tau\cup\sigma\in\Delta,\ \tau\cap\sigma=\varnothing\}.$$

For convenience, we also define $\mathrm{lk}_\Delta(\tau)=\varnothing$ if $\tau\notin\Delta$.

Let Δ be a simplicial complex on $[n]$. The Stanley–Reisner ideal of Δ (in S) is the ideal

$$I_\Delta=(x^\tau:\tau\subseteq[n],\ \tau\notin\Delta)\subseteq S,$$

where $x^\tau = \prod_{i\in\tau} x_i$. If (Δ, Γ) is a relative simplicial complex, then the *Stanley–Reisner module* of (Δ, Γ) is the S-module

$$\mathbb{F}[\Delta, \Gamma] = I_\Gamma / I_\Delta.$$

When $\Gamma = \varnothing$, the ring $\mathbb{F}[\Delta] = \mathbb{F}[\Delta, \varnothing] = S/I_\Delta$ is called the *Stanley–Reisner ring* of Δ.

A relative simplicial complex (Δ, Γ) is said to be *Buchsbaum* if $\mathbb{F}[\Delta, \Gamma]$ is a Buchsbaum module. The following characterization of the Buchsbaum property was given in [Schenzel 1981, Theorem 3.2]; a proof for relative simplicial complexes appears in [Adiprasito and Sanyal 2016, Theorem 1.11].

Theorem 3.1. *A pure relative simplicial complex* (Δ, Γ) *of dimension d is Buchsbaum if and only if*

$$\widetilde{H}_i(\mathrm{lk}_\Delta(\tau), \mathrm{lk}_\Gamma(\tau)) = 0$$

for every nonempty face $\tau \in \Delta \setminus \Gamma$ and all $i \neq d - |\tau|$.

In particular, if Δ is a homology manifold with boundary, then Δ and $(\Delta, \partial\Delta)$ are Buchsbaum (relative) simplicial complexes. (However, most Buchsbaum complexes are not homology manifolds.)

Next, we discuss face numbers of Buchsbaum simplicial complexes. For a relative simplicial complex (Δ, Γ) of dimension $d-1$, let $f_i(\Delta, \Gamma)$ be the number of i-dimensional faces of (Δ, Γ) and let

$$h_j(\Delta, \Gamma) = \sum_{i=0}^{j} (-1)^{j-i} \binom{d-i}{d-j} f_{i-1}(\Delta, \Gamma) \quad \text{for } j = 0, 1, \ldots, d.$$

For convenience, we also define $h_j(\Delta, \Gamma) = 0$ for $j > \dim(\Delta, \Gamma) + 1$. The h-numbers play a central role in the study of face numbers of Cohen–Macaulay simplicial complexes. On the other hand, for Buchsbaum simplicial complexes, the following modifications of h-numbers, called h'-numbers and h''-numbers, behave better than the usual h-numbers.

Recall that a *linear system of parameters* (or l.s.o.p.) is an h.s.o.p. $\Theta = \theta_1, \ldots, \theta_d$ consisting of linear forms. Note that when $\mathbb{F}$ is infinite, any finitely generated graded S-module has an l.s.o.p. The following result, established in [Schenzel 1981, §4] (a proof for Stanley–Reisner modules appears in [Adiprasito and Sanyal 2016, Theorem 2.5]), is known as Schenzel's formula.

Theorem 3.2. *Let* (Δ, Γ) *be a Buchsbaum relative simplicial complex of dimension* $d - 1$ *and let* Θ *be an l.s.o.p. for* $\mathbb{F}[\Delta, \Gamma]$. *Then, for* $j = 0, 1, \ldots, d$,

$$\dim_{\mathbb{F}}(\mathbb{F}[\Delta, \Gamma]/\Theta\mathbb{F}[\Delta, \Gamma])_j = h_j(\Delta, \Gamma) - \binom{d}{j} \sum_{i=1}^{j-1} (-1)^{j-i} \tilde{\beta}_{i-1}(\Delta, \Gamma).$$

In view of Schenzel's formula, we define the *h'-numbers* of a $(d-1)$-dimensional relative simplicial complex (Δ, Γ) by

$$h'_j(\Delta, \Gamma) = h_j(\Delta, \Gamma) - \binom{d}{j} \sum_{i=1}^{j-1} (-1)^{j-i} \tilde{\beta}_{i-1}(\Delta, \Gamma).$$

Furthermore, we define the *h''-numbers* of (Δ, Γ) by

$$h''_j(\Delta, \Gamma) = \begin{cases} h'_j(\Delta, \Gamma) - \binom{d}{j} \tilde{\beta}_{j-1}(\Delta, \Gamma) & \text{if } 0 \le j < d, \\ h'_d(\Delta, \Gamma) & \text{if } j = d. \end{cases}$$

We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. Let $M = \mathbb{F}[\Delta, \Gamma]$. Observe that

$$\dim_{\mathbb{F}}(M/\Sigma(\Theta; M))_j = \dim_{\mathbb{F}}(M/\Theta M)_j - \dim_{\mathbb{F}}(\Sigma(\Theta; M)/\Theta M)_j.$$

Since $H^i_{\mathfrak{m}}(M) = (H^i_{\mathfrak{m}}(M))_0 \cong \widetilde{H}_{i-1}(\Delta, \Gamma)$ for $i < d$ (see [Adiprasito and Sanyal 2016, Theorem 1.8]), Theorem 2.3(i) implies

$$\dim_{\mathbb{F}}(\Sigma(\Theta; M)/\Theta M)_j = \begin{cases} \binom{d}{j} \tilde{\beta}_{j-1}(\Delta, \Gamma) & \text{if } 0 \le j \le d-1, \\ 0 & \text{if } j \ge d. \end{cases}$$

The statement then follows from Theorem 3.2, asserting that $\dim_{\mathbb{F}}(M/\Theta M)_j = h'_j(\Delta, \Gamma)$ for all j. $\square$

It was proved in [Novik and Swartz 2009b, Theorem 3.4] that the j-th graded component of the socle of $\mathbb{F}[\Delta]/\Theta\mathbb{F}[\Delta]$ has dimension at least $\binom{d}{j}\tilde{\beta}_{j-1}(\Delta)$. This implies that the h''-numbers of a Buchsbaum simplicial complex form the Hilbert function of some quotient of its Stanley–Reisner ring. A new contribution and the significance of Theorem 1.2 is that it provides an explicit algebraic interpretation of h''-numbers via a submodule $\Sigma(\Theta; -)$.

In the rest of this section we discuss a few algebraic and combinatorial applications of our results. As was proved by Kalai and, independently, Stanley [1993; 1996, Chapter III, §9], if $\Delta \supseteq \Gamma$ are Cohen–Macaulay simplicial complexes of the same dimension, then $h_i(\Delta) \ge h_i(\Gamma)$ for all i. The interpretation of the h''-numbers given in Theorem 1.2 allows us to prove the following generalization of this fact.

Theorem 3.3. *Let $\Delta \supseteq \Gamma$ be Buchsbaum simplicial complexes of the same dimension. Then $h''_i(\Delta) \ge h''_i(\Gamma)$ for all i.*

Proof. We may assume that Δ and Γ are simplicial complexes on $[n]$ and that $\mathbb{F}$ is infinite. Let $d = \dim \Delta + 1$. Then there is a common linear system of parameters

$\Theta = \theta_1, \ldots, \theta_d$ for $\mathbb{F}[\Delta]$ and $\mathbb{F}[\Gamma]$. By the definition of $\Sigma(\Theta; -)$,

$$\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]) = S\Big/\bigg((\Theta) + \sum_{k=1}^{d}\big(((\theta_1, \ldots, \hat{\theta}_k, \ldots, \theta_d) + I_\Delta) :_S \theta_k\big)\bigg),$$

and an analogous formula holds for $\mathbb{F}[\Gamma]/\Sigma(\Theta; \mathbb{F}[\Gamma])$. Since $I_\Gamma \supseteq I_\Delta$, the above formula implies that $\mathbb{F}[\Gamma]/\Sigma(\Theta; \mathbb{F}[\Gamma])$ is a quotient ring of $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta])$. The desired statement then follows from Theorem 1.2. $\square$

Corollary 3.4. *Let Δ be a Buchsbaum simplicial complex and let $\tau \in \Delta$ be any nonempty face. Then $h_i''(\Delta) \geq h_i(\mathrm{lk}_\Delta(\tau))$ for all i.*

Proof. Consider the star of τ, $\mathrm{st}_\Delta(\tau) = \{\sigma \in \Delta : \sigma \cup \tau \in \Delta\}$. Then $\mathrm{st}_\Delta(\tau)$ is Cohen–Macaulay (by Theorem 3.1 and Reisner's criterion) and has the same dimension as Δ, and so by Theorem 3.3 we have $h_i''(\Delta) \geq h_i''(\mathrm{st}_\Delta(\tau))$ for all i. The result follows since $\mathrm{lk}_\Delta(\tau)$ and $\mathrm{st}_\Delta(\tau)$ have the same h-numbers (this is because $\mathrm{st}_\Delta(\tau)$ is just a cone over $\mathrm{lk}_\Delta(\tau)$; cf. [Stanley 1996, Corollary III.9.2]) and since for Cohen–Macaulay simplicial complexes the h''-numbers coincide with the h-numbers. $\square$

Let $R = S/I$ be a graded $\mathbb{F}$-algebra. The *a-invariant* of R is the number

$$a(R) = -\min\{k : (\omega_R)_k \neq 0\}.$$

This number is an important invariant in commutative algebra. When R is Cohen–Macaulay, it is well-known that $a(R) = \max\{i : h_i(R) \neq 0\} - \dim R$, where $h_i(R)$ is the i-th h-number of R. (See [Bruns and Herzog 1993, §4.1] for the definition of h-numbers for modules.) The following result provides a generalization of this fact.

Theorem 3.5. *Let $R = S/I$ be a Buchsbaum graded $\mathbb{F}$-algebra of Krull dimension d with* $\mathrm{depth}(R) \geq 2$, *and let $\Theta = \theta_1, \ldots, \theta_d$ be an l.s.o.p. for R. Then*

$$a(R) = \max\{k : (R/\Sigma(\Theta; R))_k \neq 0\} - d.$$

In particular, for any connected Buchsbaum simplicial complex Δ of dimension $d-1$, we have $a(\mathbb{F}[\Delta]) = \max\{k : h_k''(\Delta) \neq 0\} - d$.

Proof. Let $m = \max\{k : (R/\Sigma(\Theta; R))_k \neq 0\}$. Then by Theorem 1.3,

$$\min\{k : (\omega_R/\Sigma(\Theta; \omega_R))_k \neq 0\} = d - m.$$

As $\min\{k : (\omega_R)_k \neq 0\} = \min\{k : (\omega_R/\Theta\omega_R)_k \neq 0\}$, the theorem would follow if we prove that $\min\{k : (\omega_R/\Theta\omega_R)_k \neq 0\} = d - m$. To this end, it is enough to show that

$$(\Sigma(\Theta; \omega_R)/\Theta\omega_R)_k = 0 \quad \text{for } k \leq d - m - 1. \tag{4}$$

Since $\mathfrak{m}H_\mathfrak{m}^i(R) = 0$ for $i < d$ (see [Stückrad and Vogel 1986, Proposition I.2.1]), we conclude from Theorem 2.3(i) that $\mathfrak{m}(\Sigma(\Theta; R)/\Theta R) = 0$. Furthermore, since

$(R/\Theta R)_k = (\Sigma(\Theta; R)/\Theta R)_k$ for $k \geq m+1$, it follows that, for $k \geq m+2$,

$$(\Sigma(\Theta; R)/\Theta R)_k = (R/\Theta R)_k = (\mathfrak{m}(R/\Theta R))_k = (\mathfrak{m}(\Sigma(\Theta; R)/\Theta R))_k = 0.$$

The isomorphism

$$\Sigma(\Theta; R)/\Theta R \cong \bigoplus_{C \subsetneq [d]} H_{\mathfrak{m}}^{|C|}(R)(-|C|)$$

established in Theorem 2.3(i) then implies that

$$H_{\mathfrak{m}}^i(R)_j = 0 \quad \text{for all } i \leq d-1,\ j \geq m+2-i.$$

Therefore,

$$(H_{\mathfrak{m}}^{d-i+1}(R)^{\vee}(-i))_k = 0 \quad \text{for all } 2 \leq i \leq d-1,\ -(k-i) \geq m+2-(d-i+1),$$

and so

$$(H_{\mathfrak{m}}^{d-i+1}(R)^{\vee}(-i))_k = 0 \quad \text{for all } 2 \leq i \leq d-1,\ k \leq d-m-1. \tag{5}$$

Finally, since $\operatorname{depth}(\omega_R) \geq 2$, we infer from Theorem 2.3(i) and Lemma 2.2(ii) that

$$\Sigma(\Theta; \omega_R)/\Theta\omega_R \cong \bigoplus_{C \subsetneq [d], |C| \geq 2} H_{\mathfrak{m}}^{|C|}(\omega_R)(-|C|) \cong \bigoplus_{C \subsetneq [d], |C| \geq 2} H_{\mathfrak{m}}^{d-|C|+1}(R)^{\vee}(-|C|).$$

The above isomorphisms and (5) then yield the desired property (4). □

Observe that, by Hochster's formula on local cohomology [Bruns and Herzog 1993, Theorem 5.3.8], if Δ is a $(d-1)$-dimensional Buchsbaum simplicial complex, then $-a(\mathbb{F}[\Delta])$ equals the minimum cardinality of a face whose link has a nonvanishing top homology. Thus the "in particular" part of Theorem 3.5 can be equivalently restated as follows: if, for every face $\tau \in \Delta$ of dimension $< d-k$, the link of τ has vanishing top homology, then $h''_k(\Delta) = 0$. For the case of homology manifolds with boundary, faces with vanishing top homology are precisely the boundary faces; in this case, the above statement reduces to [Murai and Nevo 2014, Theorem 3.1].

A sequence $h_0, h_1, \ldots, h_m$ of numbers is said to be *unimodal* if there is an index p such that $h_0 \leq h_1 \leq \cdots \leq h_p \geq \cdots \geq h_m$. It was proved in [Murai 2010] that the h''-numbers of the barycentric subdivision of any connected Buchsbaum simplicial complex form a unimodal sequence. Here we use the Matlis duality established in Theorem 1.3 to generalize this result to Buchsbaum polyhedral complexes. (We refer our readers to [Murai 2010] for the definition of barycentric subdivisions.)

Theorem 3.6. *Let Δ be the barycentric subdivision of a connected polyhedral complex Γ of dimension $d-1$. Suppose that the characteristic of $\mathbb{F}$ is zero and*

Δ is Buchsbaum. Then, for a generic choice of linear forms $\Theta = \theta_1, \dots, \theta_d$, and $\theta_{d+1} \in \mathbb{F}[\Delta]$, the multiplication

$$\times\theta_{d+1} : (\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]))_i \to (\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]))_{i+1}$$

is injective for $i \le \frac{d}{2} - 1$ and is surjective for $i \ge \frac{d}{2}$. In particular, the sequence $h''_0(\Delta), h''_1(\Delta), \dots, h''_d(\Delta)$ is unimodal.

Proof. By genericity of linear forms, $\Theta = \theta_1, \dots, \theta_d$ is a common l.s.o.p. for $\mathbb{F}[\Delta]$ and $\omega_{\mathbb{F}[\Delta]}$. Let P be the face poset of Γ. Then the Stanley–Reisner ring $\mathbb{F}[\Delta]$ is a squarefree P-module (a notion introduced in [Murai and Yanagawa 2014, Definition 2.1]); furthermore, by [Murai and Yanagawa 2014, Theorem 3.1] $\omega_{\mathbb{F}[\Delta]}$ is also a squarefree P-module. It then follows from [Murai and Yanagawa 2014, Theorem 6.2(ii)] that the multiplication maps

$$\times\theta_{d+1} : (\mathbb{F}[\Delta]/\Theta\mathbb{F}[\Delta])_i \to (\mathbb{F}[\Delta]/\Theta\mathbb{F}[\Delta])_{i+1} \tag{6}$$

and

$$\times\theta_{d+1} : (\omega_{\mathbb{F}[\Delta]}/(\Theta\omega_{\mathbb{F}[\Delta]}))_i \to (\omega_{\mathbb{F}[\Delta]}/(\Theta\omega_{\Theta\mathbb{F}[\Delta]}))_{i+1}$$

are surjective for $i \ge \frac{d}{2}$. (While Cohen–Macaulayness was assumed in [Murai and Yanagawa 2014, Theorem 6.2], one can see from their proof of the theorem that this assumption was used only in the proof of part (i) and is unnecessary to derive surjectivity.) Since $\Theta\mathbb{F}[\Delta]$ is contained in $\Sigma(\Theta; \mathbb{F}[\Delta])$, the map in (6) remains surjective if we replace $\Theta\mathbb{F}[\Delta]$ with $\Sigma(\Theta; \mathbb{F}[\Delta])$, and a similar statement holds for $\omega_{\mathbb{F}[\Delta]}$. These surjectivities and the Matlis duality $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]) \cong (\omega_{\mathbb{F}[\Delta]}/\Sigma(\Theta; \omega_{\mathbb{F}[\Delta]}))^\vee(d)$ of Theorem 1.3 yield the desired statement. $\square$

In fact, in view of results from [Juhnke-Kubitzke and Murai 2015], it is tempting to conjecture that if Δ is a barycentric subdivision of a Buchsbaum regular CW-complex of dimension $d-1$, then even the sequence

$$h''_0(\Delta)\Big/\binom{d}{0},\ h''_1(\Delta)\Big/\binom{d}{1},\ h''_2(\Delta)\Big/\binom{d}{2},\ \dots,\ h''_d(\Delta)\Big/\binom{d}{d}$$

is unimodal.

We close this section with a couple of remarks.

Remark 3.7. Our results on h''-vectors can be generalized to the following setting. Consider a finitely generated graded S-module M of Krull dimension d such that

$$H^i_{\mathfrak{m}}(M) = (H^i_{\mathfrak{m}}(M))_0 \quad \text{for all } 0 \le i < d. \tag{7}$$

Define the h'- and h''-numbers of M in the same way as for the Stanley–Reisner modules but with $\dim_{\mathbb{F}}(H^j_{\mathfrak{m}}(M))$ used as a replacement for $\tilde{\beta}_{j-1}(\Delta, \Gamma)$. Then suitably modified statements of Theorems 1.2 and 3.2 continue to hold for such an M. Furthermore, if M satisfies (7), then M must be Buchsbaum (see [Stückrad and Vogel

1986, Proposition I.3.10]), in which case ω_M also satisfies (7) by Lemma 2.2(ii). In particular, if Δ is an arbitrary connected Buchsbaum simplicial complex of dimension $d-1$, then

$$h''_i(\omega_{\mathbb{F}[\Delta]}) = h''_{d-i}(\mathbb{F}[\Delta]) \quad \text{for all } i = 0, 1, \ldots, d.$$

Remark 3.8. A statement analogous to Corollary 1.4 also holds for homology manifolds without boundary. Indeed, if Δ is an orientable homology manifold without boundary, then by Gräbe's result [1984], $\mathbb{F}[\Delta]$ is isomorphic to its own canonical module, and hence, by Theorem 1.3, $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta])$ is an Artinian Gorenstein algebra. This fact was essentially proved in [Novik and Swartz 2009a, Theorem 1.4].

Let us also point out that if Δ is a connected orientable homology manifold with or without boundary, then for $M = \mathbb{F}[\Delta]$ the statement of part (ii) of Lemma 2.2 is a simple consequence of the Poincaré–Alexander–Lefschetz duality along with Gräbe's result [1984] that $\omega_M \cong \mathbb{F}[\Delta, \partial\Delta]$.

4. Applications to the manifold g-conjecture

In this section we discuss connected orientable homology manifolds without boundary. One of the most important open problems in algebraic combinatorics is the algebraic g-conjecture; it asserts that every homology sphere has the weak Lefschetz property (the WLP, for short). Kalai proposed [Novik 1998, Conjecture 7.5] a far-reaching generalization of this conjecture to homology manifolds. Using the $\Sigma(\Theta, -)$ module allows us to restate Kalai's conjecture as follows.

Conjecture 4.1. *Let Δ be a connected orientable $\mathbb{F}$-homology $(d-1)$-manifold without boundary. Then, for a generic choice of linear forms $\Theta = \theta_1, \ldots, \theta_d$, the ring $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta])$ has the WLP; that is, for a generic linear form ω, the multiplication*

$$\times\omega : (\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]))_{\lfloor d/2 \rfloor} \to (\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta]))_{\lfloor d/2 \rfloor + 1}$$

is surjective.

It is worth mentioning that while Kalai's original statement of the conjecture did not involve $\Sigma(\Theta, -)$, the two statements are equivalent (see [Novik and Swartz 2009a] and Remark 3.8 above). Somewhat informally, we say that Δ (or $\mathbb{F}[\Delta]$) has the WLP if Δ satisfies the conclusions of the above conjecture. Enumerative consequences of this conjecture are discussed in [Novik and Swartz 2009a, §1].

Given a finite set A, we denote by $\bar{A}$ the simplex on A, i.e., the simplicial complex whose set of faces consists of all subsets of A. When $A = \{a\}$ consists of a single vertex, we write $\bar{a}$ to denote the vertex a, viewed as a 0-dimensional simplex. If

Δ and Γ are two simplicial complexes on disjoint vertex sets, then the *join* of Δ and Γ, $\Delta * \Gamma$, is the simplicial complex defined by

$$\Delta * \Gamma := \{\sigma \cup \tau : \sigma \in \Delta,\ \tau \in \Gamma\}.$$

Finally, if Δ is a simplicial complex and W is a set, then $\Delta_W = \{\sigma \in \Delta : \sigma \subseteq W\}$ is the subcomplex of Δ induced by W.

Let Δ be a $(d-1)$-dimensional homology manifold, and let A and B be disjoint subsets such that $|A|+|B|=d+1$. If $\Delta_{A\cup B}=\overline{A}*\partial\overline{B}$, then the operation of removing $\overline{A}*\partial\overline{B}$ from Δ and replacing it with $\partial\overline{A}*\overline{B}$ is called a $(|B|-1)$-*bistellar flip*. (For instance, a 0-flip is simply a stellar subdivision at a facet.) The resulting complex is a homology manifold homeomorphic to the original complex.

The following surprising property was proved by Pachner [1987; 1991]: if Δ_1 and Δ_2 are two PL homeomorphic combinatorial manifolds without boundary, then they can be connected by a sequence of bistellar flips. Since the boundary complex of a simplex has the WLP, Pachner's result suggests the following inductive approach to the algebraic g-conjecture for PL spheres: prove that bistellar flips applied to PL spheres preserve the WLP. Swartz showed [2014, §3] that most bistellar flips applied to homology *spheres* preserve the WLP. Here we extend Swartz's results to the generality of orientable homology manifolds without boundary:

Theorem 4.2. *Let Δ be a $(d-1)$-dimensional, connected, orientable homology manifold without boundary. Suppose Δ' is obtained from Δ via a $(p-1)$-bistellar flip with $p \neq \frac{1}{2}(d+1)$ if d is odd and with $p \notin \left\{\frac{d}{2}, \frac{1}{2}(d+2)\right\}$ if d is even, and let Θ be a common l.s.o.p. for $\mathbb{F}[\Delta]$ and $\mathbb{F}[\Delta']$. Then $\mathbb{F}[\Delta']/\Sigma(\Theta; \mathbb{F}[\Delta'])$ has the WLP if and only if $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta])$ has the WLP.*

If Δ is a simplicial complex and σ is a face of Δ, then the *stellar subdivision* of Δ at σ consists of (i) removing σ and all faces containing it from Δ, (ii) introducing a new vertex a, and (iii) adding new faces in $\bar{a} * \partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ to Δ:

$$\mathrm{sd}_\sigma(\Delta) := (\Delta \setminus \mathrm{st}_\Delta(\sigma)) \cup (\bar{a} * \partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)).$$

A classical result due to Alexander [1930] asserts that two simplicial complexes are PL homeomorphic if and only if they are stellar equivalent; that is, one of them can be obtained from another by a sequence of stellar subdivisions and their inverses. Thus, a different approach to the algebraic g-conjecture (at least for PL manifolds) is to show that the WLP is preserved by stellar subdivisions and their inverses. Böhm and Papadakis proved [2015, Corollary 4.5] that this is the case if one applies stellar subdivisions at faces of sufficiently large dimension (or their inverses) to homology spheres. We extend their result to the generality of orientable homology manifolds without boundary:

Theorem 4.3. *Let Δ be a $(d-1)$-dimensional, connected, orientable $\mathbb{F}$-homology manifold without boundary and let σ be a face of Δ with $\dim\sigma > \frac{d}{2}$. Then $\mathbb{F}[\Delta]$ has the WLP if and only if $\mathbb{F}[\mathrm{sd}_\sigma(\Delta)]$ has the WLP.*

The structure of the proofs of both theorems is similar to the proof of [Swartz 2014, Theorem 3.1]. The new key ingredient is given by the following lemma. Recall that Δ is an $\mathbb{F}$-*homology* $(d-1)$-*sphere* if Δ is an $\mathbb{F}$-homology manifold whose homology over $\mathbb{F}$ coincides with that of $\mathbb{S}^{d-1}$. Similarly, Δ is an $\mathbb{F}$-*homology ball* if (i) Δ is an $\mathbb{F}$-homology manifold with boundary, (ii) the homology of Δ (over $\mathbb{F}$) vanishes, and (iii) the boundary of Δ is an $\mathbb{F}$-homology sphere.

Lemma 4.4. *Let Δ be a $(d-1)$-dimensional $\mathbb{F}$-homology manifold without boundary, Γ a full-dimensional subcomplex of Δ, and Θ an l.s.o.p. for $\mathbb{F}[\Delta]$. Assume further that Γ is an $\mathbb{F}$-homology ball, and let D be the simplicial complex obtained from Δ by removing the interior faces of Γ. Then D is an $\mathbb{F}$-homology manifold with boundary and the natural surjection $\mathbb{F}[\Delta]\to\mathbb{F}[\Gamma]$ induces the short exact sequence*

$$0\to\mathbb{F}[D,\partial D]/\Sigma(\Theta;\mathbb{F}[D,\partial D])\to\mathbb{F}[\Delta]/\Sigma(\Theta;\mathbb{F}[\Delta])\to\mathbb{F}[\Gamma]/(\Theta\mathbb{F}[\Gamma])\to 0.$$

Proof. The fact that D is a homology manifold with boundary follows by a standard Mayer–Vietoris argument (note that $\partial D=\partial\Gamma$ is a homology $(d-2)$-sphere). Furthermore, by excision and since Γ has vanishing homology,

$$\tilde\beta_i(D,\partial D)=\tilde\beta_i(\Delta,\Gamma)=\tilde\beta_i(\Delta)\quad\text{for all } i. \tag{8}$$

Now, using that the homology ball Γ is a full-dimensional subcomplex of Δ, we conclude as in the proof of Theorem 3.3 that there is a natural surjection

$$\phi:\mathbb{F}[\Delta]/\Sigma(\Theta;\mathbb{F}[\Delta])\to\mathbb{F}[\Gamma]/(\Theta\mathbb{F}[\Gamma])=\mathbb{F}[\Gamma]/\Sigma(\Theta;\mathbb{F}[\Gamma]). \tag{9}$$

Thus, to finish the proof of the lemma, it suffices to show that the kernel of ϕ is $\mathbb{F}[D,\partial D]/\Sigma(\Theta;\mathbb{F}[D,\partial D])$. This, in turn, would follow if we verify that

(i) $\mathrm{Ker}(\phi)$ is a quotient of $\mathbb{F}[D,\partial D]/\Sigma(\Theta;\mathbb{F}[D,\partial D])$, and

(ii) $\dim_{\mathbb{F}}(\mathrm{Ker}(\phi))_j=\dim_{\mathbb{F}}(\mathbb{F}[D,\partial D]/\Sigma(\Theta;\mathbb{F}[D,\partial D]))_j$ for all j.

For brevity, let $R=\mathbb{F}[\Delta]$ and let $I=\mathbb{F}[D,\partial D]$. Since $(D,\partial D)=(\Delta,\Gamma)$ as relative simplicial complexes, it follows that I is an ideal of R, and R/I is $\mathbb{F}[\Gamma]$. Thus our surjection ϕ is the projection $\phi: R/\Sigma(\Theta;R)\to R/(I+\Theta R)$. Hence

$$\mathrm{Ker}(\phi)=(I+\Theta R)/\Sigma(\Theta;R)=(I+\Sigma(\Theta;R))/\Sigma(\Theta;R)=I/(I\cap\Sigma(\Theta;R)),$$

which together with an observation that $I\cap\Sigma(\Theta;R)$ contains $\Sigma(\Theta;I)$ yields assertion (i). To see that $I\cap\Sigma(\Theta;R)\supseteq\Sigma(\Theta;I)$, note that, since I is a subset of R, ΘI is contained in ΘR, and $(\theta_1,\dots,\hat\theta_i,\dots,\theta_d)I:_I\theta_i$ is contained in $(\theta_1,\dots,\hat\theta_i,\dots,\theta_d)R:_R\theta_i$ for all $i\in[d]$; therefore $\Sigma(\Theta;R)\supseteq\Sigma(\Theta;I)$.

As for assertion (ii), the dimension of the kernel of ϕ can be computed as follows: by (9) and Theorem 1.2,

$$\dim_{\mathbb{F}}(\mathrm{Ker}(\phi))_j = h''_j(\Delta) - h_j(\Gamma). \tag{10}$$

Now, since each face of Δ is either a face of $(D, \partial D)$ or a face of Γ (but not of both), $f_i(\Delta) = f_i(D, \partial D) + f_i(\Gamma)$ for all $i \geq -1$, and so

$$h_j(D, \partial D) = h_j(\Delta) - h_j(\Gamma) \quad \text{for all } j \geq 0. \tag{11}$$

Equations (8), (10) and (11) yield

$$\dim_{\mathbb{F}}(\mathrm{Ker}(\phi))_j = h''_j(D, \partial D) = \dim_{\mathbb{F}}(\mathbb{F}[D, \partial D]/\Sigma(\Theta; \mathbb{F}[D, \partial D]))_j \quad \text{for all } j \geq 0,$$

where the last equality is another application of Theorem 1.2. The lemma follows. $\square$

Another ingredient needed for both proofs is the following immediate consequence of the snake lemma.

Lemma 4.5. *Let $0 \to L \to N \to M \to 0$ be an exact sequence of graded S-modules, let $\omega \in S$ be a linear form, and let k be a fixed integer. Assume also that the map $\times\omega : M_k \to M_{k+1}$ is bijective. Then the map $\times\omega : N_k \to N_{k+1}$ is surjective if and only if the map $\times\omega : L_k \to L_{k+1}$ is surjective.*

We are now ready to prove both of the theorems.

Proof of Theorem 4.2. We are given that $\Delta' = (\Delta \setminus (\bar{A} * \partial \bar{B})) \cup (\partial \bar{A} * \bar{B})$, where $|B| = p$. Let D be the simplicial complex obtained from Δ by removing the interior faces of the ball $\Gamma_1 = \bar{A} * \partial \bar{B}$; equivalently, D is obtained from Δ' by removing the interior faces of $\Gamma_2 = \partial \bar{A} * \bar{B}$. Since Γ_1 and Γ_2 are full-dimensional subcomplexes of Δ and Δ', and Θ is an l.s.o.p. for both $\mathbb{F}[\Delta]$ and $\mathbb{F}[\Delta']$, it is also an l.s.o.p. for both $\mathbb{F}[\Gamma_1]$ and $\mathbb{F}[\Gamma_2]$. Hence

$$\dim_{\mathbb{F}}(\mathbb{F}[\Gamma_1]/\Theta\mathbb{F}[\Gamma_1])_j = h_j(\bar{A} * \partial \bar{B}) = h_j(\partial \bar{B}) = \begin{cases} 1 & \text{if } j \leq p-1, \\ 0 & \text{if } j > p-1, \end{cases}$$

which implies that $\mathbb{F}[\Gamma_1]/\Theta\mathbb{F}[\Gamma_1] \cong \mathbb{F}[x]/(x^p)$. Similarly, we have $\mathbb{F}[\Gamma_2]/\Theta\mathbb{F}[\Gamma_2] \cong \mathbb{F}[x]/(x^{d-p+1})$. These isomorphisms together with the assumption $p \notin \{\frac{d}{2}, \frac{d+1}{2}, \frac{d+2}{2}\}$ yield that, for a generic choice of a linear form w, the multiplication map

$$\times w : (\mathbb{F}[\Gamma_i]/\Theta\mathbb{F}[\Gamma_i])_{\lfloor d/2 \rfloor} \to (\mathbb{F}[\Gamma_i]/\Theta\mathbb{F}[\Gamma_i])_{\lfloor d/2 \rfloor + 1}$$

is bijective for $i \in \{1, 2\}$.

Applying Lemma 4.4 to Δ and Γ_1, we can conclude from Lemma 4.5 that $\mathbb{F}[\Delta]/\Sigma(\Theta; \mathbb{F}[\Delta])$ has the WLP if and only if, for a generic linear form w, the multiplication

$$\times w : (\mathbb{F}[D, \partial D]/\Sigma(\Theta; \mathbb{F}[D, \partial D]))_{\lfloor d/2 \rfloor} \to (\mathbb{F}[D, \partial D]/\Sigma(\Theta; \mathbb{F}[D, \partial D]))_{\lfloor d/2 \rfloor + 1}$$

is surjective. On the other hand, by Lemma 4.4 applied to Δ' and Γ_2, the latter condition is equivalent to the WLP of $\mathbb{F}[\Delta']/\Sigma(\Theta;\mathbb{F}[\Delta'])$. $\square$

Proof of Theorem 4.3. Let D be the homology manifold obtained from Δ by removing the interior faces of the homology ball $\Gamma_1 = \mathrm{st}_\Delta(\sigma)$; equivalently, D is obtained from $\mathrm{sd}_\sigma(\Delta)$ by removing the interior faces of the homology ball $\Gamma_2 = \bar{a} * \partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$. As the proof of Theorem 4.2 shows, to verify Theorem 4.3, it suffices to check that, for a generic choice of linear forms $\Theta = \theta_1, \dots, \theta_d$ and another generic linear form w, the multiplication

$$\times w : (\mathbb{F}[\Gamma_i]/\Theta\mathbb{F}[\Gamma_i])_{\lfloor d/2\rfloor} \to (\mathbb{F}[\Gamma_i]/\Theta\mathbb{F}[\Gamma_i])_{\lfloor d/2\rfloor+1} \tag{12}$$

is bijective for each $i \in \{1, 2\}$.

In the case of $i = 1$, the desired bijection is an immediate consequence of the fact that

$$\dim_{\mathbb{F}}(\mathbb{F}[\Gamma_1]/\Theta\mathbb{F}[\Gamma_1])_j = h_j(\mathrm{st}_\Delta(\sigma)) = h_j(\mathrm{lk}_\Delta(\sigma)) = 0 \quad \text{for } j \geq d - |\sigma| + 1$$

and the assumption $|\sigma| > \frac{d}{2} + 1$. We now prove that the map in (12) is also bijective for $i = 2$. To this end, observe that the homology $(d-2)$-sphere $\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ is the boundary of the homology $(d-1)$-ball $\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$, and hence $\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ is $(d - |\sigma|)$-stacked. (A homology $(d-2)$-sphere is called $(d-k)$-stacked if it is the boundary of a homology ball that has no interior faces of size $\leq k - 1$.) It then follows from [Swartz 2014, Corollary 6.3] that $\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ has the WLP. This, in turn, implies that the map in (12) is surjective, as $\mathbb{F}[\Gamma_2]$ is a polynomial ring over $\mathbb{F}[\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)]$. Also, since $|\sigma| > \frac{d}{2} + 1$, we infer from [McMullen and Walkup 1971, Theorem 2] and the $(d - |\sigma|)$-stackedness of $\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ that $h_{\lfloor d/2\rfloor}(\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)) = h_{\lfloor d/2\rfloor+1}(\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma))$. As Γ_2 and $\partial\bar{\sigma} * \mathrm{lk}_\Delta(\sigma)$ have the same h-vector, we conclude that $h_{\lfloor d/2\rfloor}(\Gamma_2) = h_{\lfloor d/2\rfloor+1}(\Gamma_2)$, and therefore that the map in (12) is bijective. $\square$

Let Δ be a triangulation of a closed surface and assume that one of the vertices of Δ is connected to all other vertices of Δ. Then by [Novik and Swartz 2009a, Theorem 1.6], Δ has the WLP. However, the problem of whether *every* triangulation of a closed surface other than the sphere possesses the WLP is at present wide open.

Remark 4.6. Note that if Δ is an arbitrary *odd-dimensional* Buchsbaum complex (e.g., a homology manifold), then there exists a simplicial complex Γ that is PL homeomorphic to Δ and has the WLP in characteristic 0. Simply take Γ to be the barycentric subdivision of Δ: the WLP of Γ is guaranteed by Theorem 3.6.

Appendix: Proof of Theorem 2.3

The goal of this appendix is to verify Theorem 2.3. Our proof is based on the proof of [Novik and Swartz 2009b, Theorem 2.2], and so some details are omitted.

Let M be a finitely generated Buchsbaum graded S-module of Krull dimension d, and let $\Theta = \theta_1, \dots, \theta_d$ be an h.s.o.p. for M with $\deg \theta_i = \delta_i$. We write $\delta_C = \sum_{i \in C} \delta_i$ for $C \subseteq [d]$. We need the following result (see [Stückrad and Vogel 1986, Lemma II.4.14′]).

Lemma A.1. *There is an isomorphism*

$$H^i_{\mathfrak{m}}(M/((\theta_1, \dots, \theta_j)M)) \cong \bigoplus_{C \subseteq [j]} H^{|C|+i}_{\mathfrak{m}}(M)(-\delta_C) \quad \textit{for all } i, j \textit{ with } i + j < d.$$

We use the following notation: if $C \subseteq [d]$, then $M\langle C\rangle$ is defined by

$$M\langle C\rangle = M/((\theta_i : i \notin C)M).$$

By [Stückrad and Vogel 1986, Proposition I.2.1], if $C \subsetneq [d]$ and $s \in [d] \setminus C$, then $H^0_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle) = 0 :_{M\langle C\cup\{s\}\rangle} \theta_s$. This leads to the short exact sequence

$$0 \longrightarrow M\langle C \cup \{s\}\rangle / H^0_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle)(-\delta_s) \xrightarrow{\times\theta_s} M\langle C \cup \{s\}\rangle \xrightarrow{\pi_s} M\langle C\rangle \longrightarrow 0, \quad (13)$$

where π_s is a natural projection. This short exact sequence gives rise to exact sequences

$$0 \longrightarrow H^k_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle) \xrightarrow{\pi_s^*} H^k_{\mathfrak{m}}(M\langle C\rangle) \xrightarrow{\varphi_s^*} H^{k+1}_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle)(-\delta_s) \longrightarrow 0 \quad (14)$$

for $k < |C|$, and

$$0 \longrightarrow H^{|C|}_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle) \xrightarrow{\pi_s^*} H^{|C|}_{\mathfrak{m}}(M\langle C\rangle) \xrightarrow{\varphi_s^*} H^{|C|+1}_{\mathfrak{m}}(M\langle C \cup \{s\}\rangle)(-\delta_s), \quad (15)$$

where we denote by φ_s^* the connecting homomorphism.

Similarly, if $C \subsetneq [d]$, $|C| \le d-2$, and $s, t \in [d] \setminus C$, we obtain the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & M\langle C\cup\{s\}\rangle/H^0_{\mathfrak{m}}(M\langle C\cup\{s\}\rangle)(-\delta_s) \arrow[r, "\times\theta_s"] & M\langle C\cup\{s\}\rangle \arrow[r, "\pi_s"] & M\langle C\rangle \arrow[r] & 0 \\
0 \arrow[r] & M\langle C\cup\{s,t\}\rangle/H^0_{\mathfrak{m}}(M\langle C\cup\{s,t\}\rangle)(-\delta_s) \arrow[r, "\times\theta_s"] \arrow[u, "\pi_t"] & M\langle C\cup\{s,t\}\rangle \arrow[r, "\pi_s"] \arrow[u, "\pi_t"] & M\langle C\cup\{t\}\rangle \arrow[r] \arrow[u, "\pi_t"] & 0
\end{tikzcd}$$

This diagram, in turn, induces the commutative diagram

$$\begin{tikzcd}
H^i_{\mathfrak{m}}(M\langle C\rangle) \arrow[r, "\varphi_s^*"] & H^{i+1}_{\mathfrak{m}}(M\langle C\cup\{s\}\rangle)(-\delta_s) \\
H^i_{\mathfrak{m}}(M\langle C\cup\{t\}\rangle) \arrow[r, "\varphi_s^*"] \arrow[u, "\pi_t^*"] & H^{i+1}_{\mathfrak{m}}(M\langle C\cup\{s,t\}\rangle)(-\delta_s) \arrow[u, "\pi_t^*"]
\end{tikzcd} \quad (16)$$

Now, for $k = 2, 3, \dots, d$, define the maps ϕ_k and ψ_k as compositions:

$$\phi_k : H^0_{\mathfrak{m}}(M\langle\varnothing\rangle) \xrightarrow{\varphi_1^*} H^1_{\mathfrak{m}}(M\langle[1]\rangle)(-\delta_{[1]}) \xrightarrow{\varphi_2^*} \cdots \xrightarrow{\varphi_{k-1}^*} H^{k-1}_{\mathfrak{m}}(M\langle[k-1]\rangle)(-\delta_{[k-1]}),$$

$$\psi_k : H^0_{\mathfrak{m}}(M\langle\{k\}\rangle) \xrightarrow{\varphi_1^*} H^1_{\mathfrak{m}}(M\langle[1]\cup\{k\}\rangle)(-\delta_{[1]}) \xrightarrow{\varphi_2^*} \cdots \xrightarrow{\varphi_{k-1}^*} H^{k-1}_{\mathfrak{m}}(M\langle[k]\rangle)(-\delta_{[k-1]}).$$

Then the commutativity of (16) implies the commutativity of

$$\begin{tikzcd}
H^0_{\mathfrak{m}}(M\langle\varnothing\rangle) \arrow[r, "\phi_k"] & H^{k-1}_{\mathfrak{m}}(M\langle[k-1]\rangle)(-\delta_{[k-1]}) \\
H^0_{\mathfrak{m}}(M\langle\{k\}\rangle) \arrow[r, "\psi_k"] \arrow[u, "\pi_k^*"] & H^{k-1}_{\mathfrak{m}}(M\langle[k]\rangle)(-\delta_{[k-1]}) \arrow[u, "\pi_k^*"]
\end{tikzcd} \tag{17}$$

To prove part (ii) of Theorem 2.3, we consider the diagram

$$\begin{tikzcd}
 & & H^0_{\mathfrak{m}}(M\langle\{1\}\rangle) \arrow[r, "\pi_1^*"] & H^0_{\mathfrak{m}}(M\langle\varnothing\rangle) \arrow[d, "\varphi_1^*"] \\
H^0_{\mathfrak{m}}(M\langle\{2\}\rangle) \arrow[r, "\psi_2"] & H^1_{\mathfrak{m}}(M\langle[2]\rangle)(-\delta_{[1]}) \arrow[rr, "\pi_2^*"] & & H^1_{\mathfrak{m}}(M\langle[1]\rangle)(-\delta_{[1]}) \arrow[d, "\varphi_2^*"] \\
 & & & \vdots \arrow[d, "\varphi_{d-1}^*"] \\
H^0_{\mathfrak{m}}(M\langle\{d\}\rangle) \arrow[r, "\psi_d"] & H^{d-1}_{\mathfrak{m}}(M\langle[d]\rangle)(-\delta_{[d-1]}) \arrow[rr, "\pi_d^*"] & & H^{d-1}_{\mathfrak{m}}(M\langle[d-1]\rangle)(-\delta_{[d-1]}) \arrow[d, "\varphi_d^*"] \\
 & & & H^d_{\mathfrak{m}}(M\langle[d]\rangle)(-\delta_{[d]})
\end{tikzcd}$$

The surjectivity of φ_s^* in (14) implies that ψ_k is surjective. Hence, in each horizontal line of the diagram,

$$\operatorname{Im}(\pi_k^* \circ \psi_k) = \operatorname{Im}(\pi_k^*). \tag{18}$$

Also, since the sequence in (15) is exact, it follows that in the diagram we have

$$\operatorname{Ker}(\varphi_k^*) = \operatorname{Im}(\pi_k^*). \tag{19}$$

Define $\phi_{d+1} = \varphi_d^* \circ \cdots \circ \varphi_1^*$ to be the composition of the vertical maps in the diagram. By (15), each π_k^* (in the diagram) is injective, and we conclude that

$$\operatorname{Ker}(\phi_{d+1}) \cong \bigoplus_{k=1}^{d} \operatorname{Im}(\pi_k^*) \cong \bigoplus_{k=1}^{d} H^{k-1}_{\mathfrak{m}}(M\langle[k]\rangle)(-\delta_{[k-1]}) \tag{20}$$

as $\mathbb{F}$-vector spaces. In addition, using (18) and the commutativity of (17), we obtain that $\mathrm{Ker}(\phi_{d+1})$ is the sum of the images of

$$\pi_k^* : H^0_{\mathfrak{m}}(M\langle\{k\}\rangle) \to H^0_{\mathfrak{m}}(M\langle\varnothing\rangle) = M/\Theta M.$$

Finally, since

$$H^0_{\mathfrak{m}}(M\langle\{k\}\rangle) = ((\theta_1, \ldots, \hat{\theta}_k, \ldots, \theta_d)M :_M \theta_k)/(\theta_1, \ldots, \hat{\theta}_k, \ldots, \theta_d)M$$

and since π_k^* is a natural projection, it follows that

$$\mathrm{Ker}(\phi_{d+1}) = \Sigma(\Theta; M)/\Theta M.$$

This proves Theorem 2.3(ii) since ϕ_{d+1} is a map from $M/\Theta M$ to $H^d_{\mathfrak{m}}(M)(-\delta_{[d]})$.

It remains to verify Theorem 2.3(i). Recall that $\mathrm{Ker}(\phi_{d+1})$ is the sum of images of $H^0_{\mathfrak{m}}(M\langle\{k\}\rangle)$ and that, by [Stückrad and Vogel 1986, Proposition I.2.1], $\mathfrak{m}\cdot H^0_{\mathfrak{m}}(M\langle\{k\}\rangle) = 0$. Hence the modules in (20) are in fact isomorphic as S-modules (since they are direct sums of copies of $\mathbb{F}$). Thus we infer from (20) and Lemma A.1 that

$$\begin{aligned}
\Sigma(\Theta; M)/\Theta M = \mathrm{Ker}(\phi_{d+1}) &\cong \bigoplus_{k=1}^{d} H^{k-1}_{\mathfrak{m}}(M\langle[k]\rangle)(-\delta_{[k-1]}) \\
&\cong \bigoplus_{k=1}^{d} \Bigl[\bigoplus_{C\subseteq[d]\setminus[k]} H^{|C|+k-1}_{\mathfrak{m}}(M)(-\delta_{C\cup[k-1]}) \Bigr] \\
&\cong \bigoplus_{C\subsetneq[d]} H^{|C|}_{\mathfrak{m}}(M)(-\delta_C),
\end{aligned}$$

as desired. □

Remark. As the proof in this appendix is based on the proof of [Novik and Swartz 2009b, Theorem 2.2], it is worth pointing out that their paper contains a minor mistake. Indeed, the short "exact" sequence that appears three lines after their statement of Theorem 2.4 is not necessarily exact. However, this mistake can be easily corrected by replacing this sequence with the short exact sequence of (13).

Acknowledgments

We thank Ed Swartz for raising the question of whether there is an algebraic version of Theorem 1.1 and for bringing the result of Böhm and Papadakis to our attention. We are also grateful to Shiro Goto for helpful conversations.

References

[Adiprasito and Sanyal 2016] K. A. Adiprasito and R. Sanyal, "Relative Stanley–Reisner theory and upper bound theorems for Minkowski sums", *Publ. Math. Inst. Hautes Études Sci.* **124** (2016), 99–163. MR

[Alexander 1930] J. W. Alexander, "The combinatorial theory of complexes", *Ann. of Math.* (2) **31**:2 (1930), 292–320. MR Zbl

[Aoyama 1980] Y. Aoyama, "On the depth and the projective dimension of the canonical module", *Japan. J. Math.* (*N.S.*) **6**:1 (1980), 61–66. MR Zbl

[Aoyama and Goto 1986] Y. Aoyama and S. Goto, "Some special cases of a conjecture of Sharp", *J. Math. Kyoto Univ.* **26**:4 (1986), 613–634. MR Zbl

[Böhm and Papadakis 2015] J. Böhm and S. A. Papadakis, "Weak Lefschetz property and stellar subdivisions of Gorenstein complexes", preprint, 2015. arXiv

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR Zbl

[Goto 1983] S. Goto, "On the associated graded rings of parameter ideals in Buchsbaum rings", *J. Algebra* **85**:2 (1983), 490–534. MR Zbl

[Gräbe 1984] H.-G. Gräbe, "Über den Stanley–Reisner–Ring von Quasimannigfaltigkeiten", *Math. Nachr.* **117** (1984), 161–174. MR Zbl

[Gräbe 1987] H.-G. Gräbe, "Generalized Dehn–Sommerville equations and an upper bound theorem", *Beiträge Algebra Geom.* **25** (1987), 47–60. MR Zbl

[Juhnke-Kubitzke and Murai 2015] M. Juhnke-Kubitzke and S. Murai, "Balanced generalized lower bound inequality for simplicial polytopes", preprint, 2015. arXiv

[Klee and Novik 2016] S. Klee and I. Novik, "Face enumeration on simplicial complexes", pp. 653–686 in *Recent Trends in Combinatorics*, edited by A. Beveridge et al., The IMA Volumes in Mathematics and its Applications **159**, Springer, 2016. Zbl

[McMullen and Walkup 1971] P. McMullen and D. W. Walkup, "A generalized lower-bound conjecture for simplicial polytopes", *Mathematika* **18** (1971), 264–273. MR Zbl

[Murai 2010] S. Murai, "On face vectors of barycentric subdivisions of manifolds", *SIAM J. Discrete Math.* **24**:3 (2010), 1019–1037. MR Zbl

[Murai and Nevo 2014] S. Murai and E. Nevo, "On r-stacked triangulated manifolds", *J. Algebraic Combin.* **39**:2 (2014), 373–388. MR Zbl

[Murai and Novik 2016] S. Murai and I. Novik, "Face numbers of manifolds with boundary", preprint, 2016. To appear in *Int. Math. Res. Not. IMRN*. arXiv

[Murai and Yanagawa 2014] S. Murai and K. Yanagawa, "Squarefree P-modules and the **cd**-index", *Adv. Math.* **265** (2014), 241–279. MR Zbl

[Novik 1998] I. Novik, "Upper bound theorems for homology manifolds", *Israel J. Math.* **108** (1998), 45–82. MR Zbl

[Novik and Swartz 2009a] I. Novik and E. Swartz, "Gorenstein rings through face rings of manifolds", *Compos. Math.* **145**:4 (2009), 993–1000. MR Zbl

[Novik and Swartz 2009b] I. Novik and E. Swartz, "Socles of Buchsbaum modules, complexes and posets", *Adv. Math.* **222**:6 (2009), 2059–2084. MR Zbl

[Pachner 1987] U. Pachner, "Konstruktionsmethoden und das kombinatorische Homöomorphieproblem für Triangulationen kompakter semilinearer Mannigfaltigkeiten", *Abh. Math. Sem. Univ. Hamburg* **57** (1987), 69–86. MR Zbl

[Pachner 1991] U. Pachner, "P.L. homeomorphic manifolds are equivalent by elementary shellings", *European J. Combin.* **12**:2 (1991), 129–145. MR Zbl

[Schenzel 1981] P. Schenzel, "On the number of faces of simplicial complexes and the purity of Frobenius", *Math. Z.* **178**:1 (1981), 125–142. MR Zbl

[Schenzel 1982] P. Schenzel, *Dualisierende Komplexe in der lokalen Algebra und Buchsbaum–Ringe*, Lecture Notes in Mathematics **907**, Springer, 1982. MR Zbl

[Stanley 1993] R. P. Stanley, "A monotonicity property of h-vectors and h^*-vectors", *European J. Combin.* **14**:3 (1993), 251–258. MR Zbl

[Stanley 1996] R. P. Stanley, *Combinatorics and commutative algebra*, 2nd ed., Progress in Mathematics **41**, Birkhäuser, Boston, 1996. MR Zbl

[Stückrad and Vogel 1986] J. Stückrad and W. Vogel, *Buchsbaum rings and applications: an interaction between algebra, geometry and topology*, Springer, 1986. MR Zbl

[Suzuki 1981] N. Suzuki, "The Koszul complex of Buchsbaum modules", pp. 15–25 in *Commutative Algebra and Algebraic Geometry* (Kyoto University, 1981), vol. 446, RIMS Kokyuroku, 1981. Zbl

[Swartz 2014] E. Swartz, "Thirty-five years and counting", preprint, 2014. arXiv

Communicated by Victor Reiner
Received 2016-03-08 Accepted 2016-10-08

s-murai@ist.osaka-u.ac.jp *Department of Pure and Applied Mathematics, Graduate School of Information Science and Technology, Osaka University, Suita 565-0871, Japan*

novik@math.washington.edu *Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350, United States*

yoshida@math.chs.nihon-u.ac.jp *Department of Mathematics, College of Humanities and Sciences, Nihon University, Setagaya-ku, Tokyo 156-8550, Japan*

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.657

Analytic functions on tubes of nonarchimedean analytic spaces

Florent Martin
Appendix by Christian Kappen and Florent Martin

Let k be a discretely valued nonarchimedean field. We give an explicit description of analytic functions whose norm is bounded by a given real number r on tubes of reduced k-analytic spaces associated to special formal schemes (including k-affinoid spaces as well as open polydiscs). As an application we study the connectedness of these tubes. In the discretely valued case, this generalizes a result of Siegfried Bosch. We use as a main tool a result of Aise Johan de Jong relating formal and analytic functions on special formal schemes and a generalization of de Jong's result which is proved in the joint appendix with Christian Kappen.

1. Introduction

Let us temporarily consider a nonarchimedean nontrivially valued field k. We work with k-analytic spaces, which were introduced by Vladimir Berkovich [1990; 1993] (and we always consider strictly k-affinoid and strictly k-analytic spaces). If X is a k-affinoid space, its ring of analytic functions $\mathcal{A}$ is a k-affinoid algebra which has nice algebraic properties. A k-analytic space is connected if and only if its ring of global analytic functions contains no nontrivial idempotents. In concrete situations, if X is a k-affinoid space, one can expect to use the nice algebraic properties of its k-affinoid algebra $\mathcal{A}$ to study the connectedness of X. For more general k-analytic spaces, it might be difficult to deal with their ring of global analytic functions. For instance, the ring of global analytic functions of the open unit disc is not Noetherian. The starting point of this work is to use generic fibres of special formal schemes to overcome this difficulty in certain situations.

A result of Siegfried Bosch. Our motivation is a generalization as well as a new proof of a result due to Siegfried Bosch [1977] in the discretely valued case. Let us consider a k-affinoid algebra $\mathcal{A}$ and let X be the associated k-affinoid space. Let x be a rigid point of X, $\mathfrak{m}_x$ the associated maximal ideal of $\mathcal{A}$ and $\tilde{x}$ the image of x under

MSC2010: primary 14G22; secondary 13F25, 13F40.
Keywords: affinoid spaces, tubes, semiaffinoid, reduction.

the reduction map $\mathrm{red}: X \to \widetilde{X}$ (for the definition and properties of the reduction map, we refer to [Bosch et al. 1984, Section 7.1.5] for rigid spaces and to [Berkovich 1990, Section 2.4] for k-analytic spaces). Let $X_+(x) := \mathrm{red}^{-1}(\tilde{x})$. Following Pierre Berthelot's terminology [1996, Définitions 1.1.2], we call $X_+(x)$ the tube of $\tilde{x}$ in X. Bosch proves that if X is distinguished and equidimensional, $X_+(x)$ is connected. This connectedness result is a corollary of the main result of [Bosch 1977, Theorem 5.8], which asserts that if X is distinguished and equidimensional, then

$$\Gamma(X_+(x), \mathcal{O}_X^\circ) \simeq (\mathcal{A}^\circ)^{\wedge(t\cdot\mathcal{A}^\circ + \mathring{\mathfrak{m}}_x)} \simeq \varprojlim_n \mathcal{A}^\circ/(t\cdot\mathcal{A}^\circ + \mathring{\mathfrak{m}}_x)^n,$$

where $t \in k$ with $0 < |t| < 1$, $\mathcal{O}_X^\circ$ denotes the sheaf of analytic functions f such that $|f|_{\sup} \le 1$, $\mathring{\mathfrak{m}}_x := \mathfrak{m}_x \cap \mathcal{A}^\circ$ and $\wedge$ denotes the completion with respect to an ideal.

Analytic functions and formal functions. For the rest of the article, we assume that k is discretely valued, with nontrivial valuation. We denote its valuation ring by R and we fix a uniformizer π. Following [Berkovich 1996, Section 1], we say that an adic R-algebra A is a *special R-algebra* if it is isomorphic to a quotient of $R\langle T_1, \dots, T_m\rangle[[S_1, \dots, S_n]]$ equipped with the $(\pi, S_1, \dots, S_n)$-adic topology. Let $\mathfrak{X} := \mathrm{Spf}(A)$ be its associated formal R-scheme. Following a construction due to Berthelot [1996, 0.2.6] for rigid spaces and extended to k-analytic spaces by Berkovich [1996, Section 1], one can associate to $\mathfrak{X}$ a k-analytic space denoted by $\mathfrak{X}_\eta$ called its generic fibre. For instance, if $A = R\langle T_1, \dots, T_m\rangle[[S_1, \dots, S_n]]$, then $\mathfrak{X}_\eta \simeq E^m \times B^n$, where we denote by E^m the m-dimensional closed unit polydisc and by B^n the n-dimensional open unit polydisc. Following the terminology introduced by Christian Kappen [2010; 2012], we say that $\mathcal{A} := A \otimes_R k$ is a *semiaffinoid k-algebra*. Up to canonical isomorphism, the k-analytic space $\mathfrak{X}_\eta$ depends only on the semiaffinoid k-algebra $\mathcal{A}$ and we call it a *semiaffinoid k-analytic space* (this should not be confused with semiaffinoid k-spaces in [Kappen 2010; 2012]). So we can functorially associate to a semiaffinoid k-algebra a k-analytic space. If A is R-flat, one gets a natural injection $A \to \Gamma(\mathfrak{X}_\eta, \mathcal{O}_{\mathfrak{X}_\eta}^\circ)$. When A is in addition normal, it was proven by A.J. de Jong [1995, Theorem 7.4.1] that

$$A \simeq \Gamma(\mathfrak{X}_\eta, \mathcal{O}_{\mathfrak{X}_\eta}^\circ).$$

For the applications we have in mind, we need the following generalization, which was already stated without proof in [de Jong 1995, Remark 7.4.2].

Theorem 2.1. *Let A be a reduced special R-algebra which is R-flat and integrally closed in $A \otimes_R k$, and let X be the associated k-analytic space. Then $A \simeq \Gamma(X, \mathcal{O}_X^\circ)$.*

Let us mention that if A is a special R-algebra which is R-flat and integrally closed in $A \otimes_R k$, then A is automatically reduced (see the argument in [Kappen 2012, Remark 2.7]). Hence one can remove the assumption that A is reduced in the above theorem if necessary. Theorem 2.1 easily follows from the following result, which is proved in the appendix with Christian Kappen.

Theorem A.8. *Let $\mathcal{A}$ be a reduced semiaffinoid k-algebra, and let X be the associated k-analytic space. Then $\mathcal{A} \simeq \{f \in \Gamma(X, \mathcal{O}_X) \mid |f|_{\sup} < \infty\}$.*

Main result. Actually, the tube $X_+(x)$ is a semiaffinoid k-analytic space. Therefore, it seemed very natural to us to look for a generalization and a more direct proof of Bosch's result [1977, Theorem 5.8] using special formal R-schemes, semiaffinoid k-algebras and de Jong's result as well as its generalization (Theorem 2.1). This is the content of this article. Let us fix X a reduced semiaffinoid k-analytic space. Let $f_1, \dots, f_n \in \Gamma(X, \mathcal{O}_X^\circ)$ and let

$$U := \{x \in X \mid |f_i(x)| < 1 \ \forall i = 1, \dots, n\}.$$

Theorem 3.1 and Proposition 5.13. *With the above notations,*

$$\Gamma(X, \mathcal{O}_X^\circ)^{\wedge(f_1,\dots,f_n)} \simeq \Gamma(U, \mathcal{O}_X^\circ). \tag{1}$$

More generally, for any positive real number r, $\Gamma(X, \mathcal{O}_X^{\le r})^{\wedge(f_1,\dots,f_n)} \simeq \Gamma(U, \mathcal{O}_X^{\le r})$, where $\mathcal{O}_X^{\le r}$ is the sheaf of analytic functions f such that $|f| \le r$.

In Section 4 we associate to a semiaffinoid k-analytic space X its *canonical reduction* $\widetilde{X}$, which is a $\tilde{k}$-scheme of finite type, and a canonical reduction map $\mathrm{red} : X \to \widetilde{X}$. If X is a k-affinoid space, then red coincides with the reduction map of [Berkovich 1990, Section 2.4]. We prove the following result:

Corollary 4.14. *Let $Z \subset \widetilde{X}$ be a connected Zariski closed subset. Then* $\mathrm{red}^{-1}(Z)$ *is connected.*

We want to stress that our results hold under the assumption that k is discretely valued, whereas this assumption is not made in [Bosch 1977]. We conjecture that Theorem 3.1 holds for any nontrivially valued nonarchimedean field k and for any reduced k-affinoid space (with $(f_1, \dots, f_n)$ replaced by $(t, f_1, \dots, f_n)$ for some $t \in k^*$ with $|t| < 1$). We do not see how this could be done using the techniques of [Bosch 1977]. We believe that quasiaffinoid k-algebras (which are a generalization of special R-algebras and semiaffinoid k-algebras to arbitrary nonarchimedean nontrivially valued fields [Lipshitz and Robinson 2000]) might be a good framework to tackle this. We also want to stress that our results are more general and the proofs simpler than in [Bosch 1977] regarding the following points.

– The proof we give of (1) is pretty short: one has to use Theorem 2.1 and a certain compatibility between integral closure and tensor product for excellent rings (whose use was suggested to us by Ofer Gabber).

– The explicit description of the rings $\Gamma(U, \mathcal{O}_X^\circ)$ given in Theorem 3.1 holds for any tube, whereas in [Bosch 1977] this was proved only for tubes over closed points.

– For a positive real number r, we extend Bosch's result to analytic functions f such that $|f|_{\sup} \leq r$ (Proposition 5.13).

– Unlike in [Bosch 1977], we do not assume that X is distinguished or equidimensional. Hence, Corollary 4.14 answers positively in the discretely valued case the question raised by Jérôme Poineau [2014, Remarque 2.9]. Let us point out that using Bosch's result, Antoine Ducros [2003, Lemma 3.1.2] proved that when k is algebraically closed and X is equidimensional and reduced, the tube of a connected Zariski closed subset of $\widetilde{X}$ in X is connected; afterwards Poineau [2014, Théorème 2.8] (still relying on Bosch's result) proved the same statement for any k and assuming only that X is equidimensional.

– Our results hold not only for k-affinoid spaces, but also for semiaffinoid k-analytic spaces.

Organization of the paper. In Section 2, we give general facts about semiaffinoid k-analytic spaces. In Section 3, we prove the main result of the article, Theorem 3.1. In Section 4, we define and study the canonical reduction of semiaffinoid k-analytic spaces, and apply it to study the connectedness of tubes. In Section 5, we prove Proposition 5.13, which is the graded version of Theorem 3.1. In Section 6, we grasp additional remarks about semiaffinoid k-analytic spaces. The joint appendix with Christian Kappen aims to prove Theorem A.8.

2. Semiaffinoid k-analytic spaces

Following [Kappen 2012, Definition 2.2], we say that a k-algebra $\mathcal{A}$ is a semiaffinoid k-algebra if it is of the form $A \otimes_R k$ for some special R-algebra A. Equivalently, a semiaffinoid k-algebra is a quotient of $R\langle T_1, \ldots, T_m\rangle[\![S_1, \ldots, S_n]\!] \otimes_R k$ for some given integers m and n. The category of semiaffinoid k-algebras is defined as the category whose objects are semiaffinoid k-algebras and whose morphisms are k-algebra morphisms. In particular, the category of k-affinoid algebras is a full subcategory of the category of semiaffinoid k-algebras. If $\mathcal{A}$ is a semiaffinoid k-algebra, a *special R-model* of $\mathcal{A}$ is an R-flat special R-algebra A such that there exists an isomorphism of k-algebras $A \otimes_R k \simeq \mathcal{A}$. If A is an R-flat special R-algebra, one has a natural inclusion $A \to \mathcal{A} := A \otimes_R k$ and through this inclusion, A is identified with a special R-model of $\mathcal{A}$. One can define a functor from the category of semiaffinoid k-algebras to the category of k-analytic spaces [Kappen 2012, Section 2C1]. If $\mathcal{A}$ is a semiaffinoid k-algebra, its associated k-analytic space X is called a semiaffinoid k-analytic space. For any special R-model A of $\mathcal{A}$, one has a natural isomorphism $X \simeq \mathrm{Spf}(A)_\eta$. If $f \in \mathcal{A}$, we set $|f|_{\sup} := \sup\{|f(x)| \mid x \in X\}$ and we define

$$\mathcal{A}^\circ = \{f \in \mathcal{A} \mid |f|_{\sup} \leq 1\}.$$

It is proved in the Appendix (Theorem A.8) that if the semiaffinoid k-algebra $\mathcal{A}$ is reduced, one has an isomorphism

$$\mathcal{A} \simeq \{f \in \Gamma(X, \mathcal{O}_X) \mid |f|_{\sup} < \infty\}.$$

We are particularly interested in the following consequence of this result.

Theorem 2.1. *Let A be a reduced special R-algebra with associated k-analytic space X. Let us assume that A is R-flat and integrally closed in $A \otimes_R k$. Then $A \simeq \Gamma(X, \mathcal{O}_X^\circ)$.*

Proof. Let $\mathcal{A} = A \otimes_R k$. By definition, $\mathcal{A}$ is a reduced semiaffinoid k-algebra, so thanks to Theorem A.8, one has $\mathcal{A}^\circ \simeq \Gamma(X, \mathcal{O}_X^\circ)$. According to [Kappen 2012, Corollaries 2.10 and 2.11] $\mathcal{A}^\circ$ is a special R-algebra which contains A, and the inclusion $A \subset \mathcal{A}^\circ$ is integral. Since by assumption A is integrally closed in $\mathcal{A}$, A is also integrally closed in $\mathcal{A}^\circ$. So $A \simeq \mathcal{A}^\circ \simeq \Gamma(X, \mathcal{O}_X^\circ)$. □

Corollary 2.2. *Let $\mathcal{A}$ be a reduced semiaffinoid k-algebra with associated k-analytic space X.*

(i) *The R-algebra $\Gamma(X, \mathcal{O}_X^\circ)$ is a reduced special R-algebra.*

(ii) *If A is a reduced special R-algebra such that $A \otimes_R k \simeq \mathcal{A}$, then $\Gamma(X, \mathcal{O}_X^\circ)$ is isomorphic to the integral closure of A in $A \otimes_R k$.*

Proof. Let A be a reduced special R-model of $\mathcal{A}$, so that $A \otimes_R k \simeq \mathcal{A}$. Let A' be the integral closure of A in $\mathcal{A}$. Since A is excellent (see [Valabrega 1975; 1976]), A' is a finite A-algebra, so A' is a reduced special R-algebra. Since $A' \otimes_R k \simeq A \otimes_R k \simeq \mathcal{A}$, the k-analytic spaces attached to A and A' are both isomorphic to X. So thanks to Theorem 2.1, $\Gamma(X, \mathcal{O}_X^\circ) \simeq A'$, which is a special R-algebra. This proves (i) and (ii). □

Example 2.3. The two statements of Corollary 2.2 do not hold for nonreduced semiaffinoid k-algebras. For instance, if $\mathcal{A} = (R[\![T_1, T_2]\!] \otimes_R k)/(T_2^2)$ with associated k-analytic space X, then any element of $\Gamma(X, \mathcal{O}_X^\circ)$ is of the form $f(T_1) + g(T_1)T_2$, where $f(T_1) \in R[\![T_1]\!]$ and $g(T_1)$ is an arbitrary analytic function on the open unit disc. We can choose $g(T_1)$ such that $g(T_1) \notin R[\![T_1]\!] \otimes_R k$ (in mixed characteristic, one could take $g = \log$). This gives a counterexample to the statements (i) and (ii) of Corollary 2.2 (for (i), see [Kappen 2012, Remark 2.7]).

Corollary 2.4. *The functor*

$$\{\textit{semiaffinoid } k\textit{-algebras}\} \to \{k\textit{-analytic spaces}\}$$

is faithful and its restriction

$$\{\textit{reduced semiaffinoid } k\textit{-algebras}\} \to \{k\textit{-analytic spaces}\}$$

is fully faithful.

Proof. If $f : X \to Y$ is a morphism of k-analytic spaces induced by a morphism of semiaffinoid k-algebras $\mathcal{B} \to \mathcal{A}$ then f induces the diagram

$$\begin{tikzcd} \mathcal{B} \arrow[r] \arrow[d] & \mathcal{A} \arrow[d] \\ \Gamma(Y, \mathcal{O}_Y) \arrow[r] & \Gamma(X, \mathcal{O}_X) \end{tikzcd}$$

whose vertical arrows are injective. This proves that the functor is faithful.

To prove that the restriction of the functor to reduced semiaffinoid k-algebras is full, let us fix a reduced semiaffinoid k-algebra $\mathcal{A}$ and let X be its associated k-analytic space. Let Y be the k-analytic space associated to another semiaffinoid k-algebra $\mathcal{B}$ and let us consider a presentation

$$\mathcal{B} = (R\langle T_1, \dots, T_m\rangle [\![S_1, \dots, S_n]\!] \otimes_R k)/I$$

giving rise to a closed immersion $Y \hookrightarrow E^m \times B^n$. If $f : X \to Y$ is a morphism of k-analytic spaces then using the composition

$$X \xrightarrow{f} Y \hookrightarrow E^m \times B^n,$$

one gets functions $f_1, \dots, f_m \in \Gamma(X, \mathcal{O}_X^\circ)$ and $f_{m+1}, \dots, f_{m+n} \in \Gamma(X, \mathcal{O}_X^\circ)$ with $|f_i(x)| < 1$ for all $x \in X$ and $i \in \{m+1, \dots, m+n\}$ such that the functions $f_1, \dots, f_{m+n}$ induce the morphism $X \to E^m \times B^n$. Thanks to Theorem A.8, $f_i \in \mathcal{A}^\circ$ for $i = 1, \dots, m+n$. Thanks to [Kappen 2012, Theorem 2.13] there is a unique morphism of semiaffinoid k-algebras $R\langle T_1, \dots, T_m\rangle [\![S_1, \dots, S_n]\!] \otimes_R k \to \mathcal{A}$ sending T_i to f_i for $i = 1, \dots, m$ and sending S_i to f_{m+i} for $i = 1, \dots, n$. It factorizes through $\mathcal{B}$ and gives the desired morphism of semiaffinoid k-algebras $\mathcal{B} \to \mathcal{A}$. □

Example 2.5. The functor from semiaffinoid k-algebras to k-analytic spaces is not fully faithful. Indeed, as in Example 2.3, consider $\mathcal{A} = (R[\![T_1, T_2]\!] \otimes_R k)/(T_2^2)$ with associated k-analytic space X. Let $g(T_1) \in k[\![T_1]\!]$ be a power series which converges on the open unit disc, but such that $g(T_1) \notin R[\![T_1]\!] \otimes_R k$. Then $T_2 g(T_1) \in \Gamma(X, \mathcal{O}_X^\circ)$ and hence it defines a morphism of k-analytic spaces $\phi : X \to \mathbb{A}_k^{1,\mathrm{an}}$ whose image is the origin of $\mathbb{A}_k^{1,\mathrm{an}}$. In particular it also defines a morphism $X \to E$, where E is the closed unit disc over k, which is a k-affinoid space (with affinoid algebra $\mathcal{B} = k\langle T\rangle$). But ϕ is not induced by a morphism of semiaffinoid k-algebras $\mathcal{B} \to \mathcal{A}$.

Remark 2.6. Theorem 2.1 was already stated in [de Jong 1995, Remark 7.4.2] without proof. Theorem A.8 is also stated in [Nicaise 2009, Lemma 2.14] but its proof is obtained as a corollary of [de Jong 1995, Remark 7.4.2]. Corollary 2.4 is also stated in [Kappen 2012, Remark 2.56] without proof.

Remark 2.7. Let $\mathcal{A}$ be a semiaffinoid k-algebra with associated k-analytic space X. Let $\mathcal{A}_{\mathrm{red}} := \mathcal{A}/(\mathrm{nil}\,\mathcal{A})$ be the reduced ring associated to $\mathcal{A}$. Then $\mathcal{A}_{\mathrm{red}}$ is a reduced semiaffinoid k-algebra. Let Y be the k-analytic space associated with $\mathcal{A}_{\mathrm{red}}$. Then $\mathcal{A} \to \mathcal{A}_{\mathrm{red}}$ induces a closed immersion of k-analytic spaces $Y \to X$ which is a bijection of sets. Let B be a special R-model of $\mathcal{A}_{\mathrm{red}}$. Since $B \subset B \otimes_R k \simeq \mathcal{A}_{\mathrm{red}}$, we deduce that B is reduced. Hence since $Y \simeq \mathrm{Spf}(B)_\eta$, by [de Jong 1995, Proposition 7.2.4(c)], Y is reduced. Hence $Y \simeq X_{\mathrm{red}}$.

3. Analytic functions and formal functions on tubes

I am very grateful to Ofer Gabber for having pointed out the reference to Proposition 6.14.4 of [EGA IV$_2$ 1965], which greatly simplifies the proof of the following statement.

Theorem 3.1. *Let X be a k-analytic space associated with a reduced semiaffinoid k-algebra $\mathcal{A}$. Let $I \subset \Gamma(X, \mathcal{O}_X^\circ)$ be an ideal and let $U = \{x \in X \mid |f(x)| < 1 \ \forall f \in I\}$. Then*
$$\Gamma(X, \mathcal{O}_X^\circ)^{\wedge I} \simeq \Gamma(U, \mathcal{O}_X^\circ).$$

Proof. Let us set $A := \Gamma(X, \mathcal{O}_X^\circ)$. Thanks to Corollary 2.2(i), A is a reduced special R-algebra. Let us choose some functions $f_1, \dots, f_n \in A$ such that $(f_1, \dots, f_n) = I$. Then
$$A^{\wedge I} \simeq A[\![\rho_1, \dots, \rho_n]\!]/(f_i - \rho_i)_{i=1,\dots,n}$$
is a special R-algebra. Let us set $\mathfrak{Y} = \mathrm{Spf}(A^{\wedge I})$. Then by [de Jong 1995, Lemma 7.2.5] (or see also [Lipshitz and Robinson 2000, Theorem 5.3.5 and Proposition 5.3.6]), the induced morphism of k-analytic spaces $\mathfrak{Y}_\eta \to X$ identifies $\mathfrak{Y}_\eta$ with U as an analytic domain of X. So U is the k-analytic space associated to the semiaffinoid k-algebra $\mathcal{B} := A^{\wedge I} \otimes_R k$ and by definition, $A^{\wedge I}$ is a special R-model of $\mathcal{B}$. Thanks to Theorem 2.1, it only remains to show that $A^{\wedge I}$ is integrally closed in $A^{\wedge I} \otimes_R k \simeq \mathcal{B}$.

Since A is an excellent ring (this follows from [Valabrega 1975, Proposition 7] when $\mathrm{char}(k) = p > 0$ and from [Valabrega 1976, Theorem 9] when $\mathrm{char}(k) = 0$), the morphism $\mathrm{Spec}(A^{\wedge I}) \to \mathrm{Spec}(A)$ is regular [EGA IV$_2$ 1965, Scholie 7.8.3(v)], so in particular, $\mathrm{Spec}(A^{\wedge I}) \to \mathrm{Spec}(A)$ is a normal morphism (see [EGA IV$_2$ 1965, Définition 6.8.1] for the definitions of normal and regular morphisms of schemes).

Since $A = \Gamma(X, \mathcal{O}_X^\circ)$, it follows that A is integrally closed in $A \otimes_R k$. So thanks to [EGA IV$_2$ 1965, Proposition 6.14.4], $A^{\wedge I}$ is integrally closed in $A^{\wedge I} \otimes_R k$. Finally, by Theorem 2.1,
$$A^{\wedge I} \simeq \Gamma(U, \mathcal{O}_X^\circ). \qquad \square$$

4. Reduction and connectedness

Reduction. Let $\mathcal{A}$ be a semiaffinoid k-algebra and let X be its associated k-analytic space. We set

$$\mathcal{A}^\circ = \{f \in \mathcal{A} \mid |f|_{\sup} \leq 1\}, \qquad \check{\mathcal{A}} = \{f \in \mathcal{A} \mid \forall x \in X\ |f(x)| < 1\},$$
$$\mathcal{A}^{\circ\circ} = \{f \in \mathcal{A} \mid |f|_{\sup} < 1\}, \qquad \widetilde{\mathcal{A}} = \mathcal{A}^\circ/\check{\mathcal{A}}, \qquad \widetilde{\mathcal{A}}^+ = \mathcal{A}^\circ/\mathcal{A}^{\circ\circ}.$$

When $\mathcal{A}$ is a k-affinoid algebra, thanks to the maximum modulus principle [Bosch et al. 1984, Proposition 6.2.1.4], $\check{\mathcal{A}} = \mathcal{A}^{\circ\circ}$, $\widetilde{\mathcal{A}} = \widetilde{\mathcal{A}}^+$ and in this case, $\widetilde{\mathcal{A}}$ corresponds to the reduction of $\mathcal{A}$ as defined in [Bosch et al. 1984, Section 6.3]. For a general semiaffinoid k-algebra, the maximum modulus principle does not hold[1] (consider for instance $S \in R[\![S]\!] \otimes_R k$), so in general, one has a strict inclusion $\mathcal{A}^{\circ\circ} \subset \check{\mathcal{A}}$ and $\widetilde{\mathcal{A}}$ is a strict quotient of $\widetilde{\mathcal{A}}^+$. If $\mathcal{A} = R\langle T_1, \ldots, T_m\rangle[\![S_1, \ldots, S_n]\!] \otimes_R k$, then $\widetilde{\mathcal{A}} = \tilde{k}[T_1, \ldots, T_m]$ and $\widetilde{\mathcal{A}}^+ = \tilde{k}[T_1, \ldots, T_m][\![S_1, \ldots, S_n]\!]$.

For a k-analytic space X, recall that we defined the subsheaf $\mathcal{O}_X^\circ \subset \mathcal{O}_X$ of analytic functions f such that $|f(x)| \leq 1$ for all x. Likewise, we denote by $\check{\mathcal{O}}_X \subset \mathcal{O}_X$ the subsheaf of analytic functions f such that $|f(x)| < 1$ for all x.

Lemma 4.1. *Let $\mathcal{A}$ be a semiaffinoid algebra and $\mathcal{A}_{\mathrm{red}}$ the associated reduced semiaffinoid k-algebra. Let X be the k-analytic space associated with $\mathcal{A}$. By Remark 2.7, the k-analytic space associated with $\mathcal{A}_{\mathrm{red}}$ can be identified with X_{red}. Then all the natural maps in the following commutative squares are isomorphisms of R-algebras*

$$\begin{tikzcd} \mathcal{A}^\circ/\check{\mathcal{A}} \arrow[r, "\alpha"] \arrow[d, "\phi"'] & (\mathcal{A}_{\mathrm{red}})^\circ/(\check{\mathcal{A}}_{\mathrm{red}}) \arrow[d, "\phi_{\mathrm{red}}"] \\ \Gamma(X, \mathcal{O}_X^\circ)/\Gamma(X, \check{\mathcal{O}}_X) \arrow[r, "\beta"'] & \Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}}^\circ)/\Gamma(X_{\mathrm{red}}, \check{\mathcal{O}}_{X_{\mathrm{red}}}) \end{tikzcd} \tag{2}$$

$$\begin{tikzcd} \mathcal{A}^\circ/\mathcal{A}^{\circ\circ} \arrow[r] \arrow[d] & (\mathcal{A}_{\mathrm{red}})^\circ/(\mathcal{A}_{\mathrm{red}})^{\circ\circ} \arrow[d] \\ \Gamma(X, \mathcal{O}_X^\circ)/\Gamma(X, \mathcal{O}_X)_{<1} \arrow[r] & \Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}}^\circ)/\Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}})_{<1} \end{tikzcd} \tag{3}$$

where $\Gamma(X, \mathcal{O}_X)_{<1} = \{f \in \Gamma(X, \mathcal{O}_X) \mid |f|_{\sup} < 1\}$, and similarly $\Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}})_{<1}$.

Proof. Using Theorem A.8, we get isomorphisms $(\mathcal{A}_{\mathrm{red}})^\circ \simeq \Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}}^\circ)$ and

$$\check{\mathcal{A}}_{\mathrm{red}} \simeq \Gamma(X_{\mathrm{red}}, \check{\mathcal{O}}_{X_{\mathrm{red}}}),$$

which imply ϕ_{red} is an isomorphism. To prove β is an isomorphism, we first remark that X is a quasi-Stein space in the sense of [Kiehl 1967, Definition 2.3]. Moreover, we have an isomorphism of ringed spaces $(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}}) \simeq (X, \mathcal{O}_X/(\operatorname{rad}\mathcal{O}_X))$, where $\operatorname{rad}\mathcal{O}_X$ is the nilradical of $\mathcal{O}_X$. Since the latter is a coherent $\mathcal{O}_X$-ideal sheaf by

[1] Actually, $\mathcal{A}$ is a k-affinoid algebra if and only if the maximum modulus principle holds [Lipshitz and Robinson 2000, Proposition 5.3.8].

[Bosch et al. 1984, Corollary 9.5.4], it follows from Theorem B for quasi-Stein spaces [Kiehl 1967, Satz 2.4.2] that $\Gamma(X, \mathcal{O}_X) \to \Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}})$ is surjective. Hence $\Gamma(X, \mathcal{O}_X^\circ) \to \Gamma(X_{\mathrm{red}}, \mathcal{O}_{X_{\mathrm{red}}}^\circ)$ is also surjective, and so β is surjective. The injectivity of β follows easily from the fact that X and X_{red} are in natural bijection. So β is an isomorphism. Since $\mathcal{A}^\circ \to (\mathcal{A}_{\mathrm{red}})^\circ$ is surjective, we deduce similarly that α is an isomorphism. Since (2) commutes, we deduce that ϕ is also an isomorphism. The proof for (3) is analogous. $\square$

Lemma 4.2. *Let $\mathcal{A}$ be a semiaffinoid k-algebra. Then $\widetilde{\mathcal{A}}$ is a reduced $\tilde{k}$-algebra of finite type and $\widetilde{\mathcal{A}}^+$ is a reduced $\tilde{k}$-algebra which is a quotient of some*

$$\tilde{k}[T_1, \dots, T_m][\![S_1, \dots, S_n]\!].$$

Proof. By Lemma 4.1, $\widetilde{\mathcal{A}} \simeq \widetilde{\mathcal{A}_{\mathrm{red}}}$, so we can replace $\mathcal{A}$ by $\mathcal{A}_{\mathrm{red}}$ and assume that $\mathcal{A}$ is reduced. By [Kappen 2012, Corollary 2.11], under this assumption $\mathcal{A}^\circ$ is a special R-algebra, so there is an isomorphism

$$\mathcal{A}^\circ \simeq R\langle T_1, \dots, T_m\rangle[\![S_1, \dots, S_n]\!]/I$$

for some ideal I of $R\langle T_1, \dots, T_m\rangle[\![S_1, \dots, S_n]\!]$. The ideal $\check{\mathcal{A}}$ then contains the ideal generated by the image of $(\pi, S_1, \dots, S_n)$ modulo I. Hence $\widetilde{\mathcal{A}}$ is a quotient of $\tilde{k}[T_1, \dots, T_m]$, proving the first part of the lemma. Likewise $\mathcal{A}^{\circ\circ}$ contains the ideal (π), hence we get a surjective map

$$\tilde{k}[T_1, \dots, T_m][\![S_1, \dots, S_n]\!] \to \widetilde{\mathcal{A}}^+. \qquad \square$$

Remark 4.3. If $\mathcal{A}$ is a reduced semiaffinoid k-algebra, then $\check{\mathcal{A}}$ is the biggest ideal of definition of the special R-algebra $\mathcal{A}^\circ$ (see [Kappen 2012, Remark 2.8]). Likewise, one can prove easily that $\mathcal{A}^{\circ\circ} = \mathrm{rad}(\pi)$.

4.4. Let A be an arbitrary special R-algebra and let $\mathfrak{X} := \mathrm{Spf}(A)$. We define $\mathfrak{X}_s := \mathrm{Spec}(A/J)$, where J is the biggest ideal of definition of A. Then there is a *specialization map* $\mathrm{sp}_{\mathfrak{X}} : \mathfrak{X}_\eta \to \mathfrak{X}_s$ which is defined in [de Jong 1995, 7.1.10] on the subset of rigid points $\mathfrak{X}^{\mathrm{rig}}$ and in general in [Berkovich 1996, §1, p. 371] (beware that in [Berkovich 1996] the map $\mathrm{sp}_{\mathfrak{X}}$ is called the reduction map).

Definition 4.5. Let X be a semiaffinoid k-analytic space coming from a semiaffinoid k-algebra $\mathcal{A}$. We set $\widetilde{X} = \mathrm{Spec}(\widetilde{\mathcal{A}})$. We call $\widetilde{X}$ the canonical reduction of X. According to Lemma 4.2, it is a reduced $\tilde{k}$-scheme of finite type. If Y is a semiaffinoid k-analytic space and $\varphi : X \to Y$ is a morphism of k-analytic spaces, then we get an associated morphism $\varphi^* : \Gamma(Y, \mathcal{O}_Y^\circ) \to \Gamma(X, \mathcal{O}_X^\circ)$. Hence by Lemma 4.1, we can functorially associate a morphism $\tilde{\varphi} : \widetilde{X} \to \widetilde{Y}$. If φ comes from a morphism of semiaffinoid k-algebras $\psi : \mathcal{B} \to \mathcal{A}$, then $\tilde{\varphi}$ is induced by $\tilde{\psi} : \widetilde{\mathcal{B}} \to \widetilde{\mathcal{A}}$.

4.6. As in [Berkovich 1990, Section 2.4] one can define a canonical reduction map $\mathrm{red} : X \to \widetilde{X}$ in the following way. If $x \in X$, we obtain an associated map $\chi_x : \mathcal{A} \to \mathcal{H}(x)$, which gives rise to a map $\tilde{\chi}_x : \widetilde{\mathcal{A}} \to \widetilde{\mathcal{H}(x)}$. Then

$$\mathrm{red} : X \to \widetilde{X}, \qquad x \mapsto \mathrm{Ker}(\tilde{\chi}_x).$$

To stress the dependence in X, we might write red_X instead of red. Similarly as in [Berkovich 1990, Corollary 2.4.3] one can check that the canonical reduction map red is anticontinuous, i.e., the inverse image of an open set is closed.

Remark 4.7. Identifying X_{red} with the k-analytic space associated with $\mathcal{A}_{\mathrm{red}}$ (see Remark 2.7), and using Lemma 4.1, we get a commutative diagram

$$\begin{tikzcd} X_{\mathrm{red}} \arrow[r] \arrow[d, "\mathrm{red}_{X_{\mathrm{red}}}"'] & X \arrow[d, "\mathrm{red}_X"] \\ \widetilde{X_{\mathrm{red}}} \arrow[r, equal] & \widetilde{X} \end{tikzcd}$$

Remark 4.8. Let A be a special R-model of the semiaffinoid k-algebra $\mathcal{A}$, with associated k-analytic space X. Let J be the biggest ideal of definition A. The injection $A \to \mathcal{A}^\circ$ induces an injective morphism of $\tilde{k}$-algebras $\varphi : A/J \to \widetilde{\mathcal{A}}$, which induces a morphism $\iota : \widetilde{X} \to \mathfrak{X}_s$. If $\mathfrak{X} := \mathrm{Spf}(A)$ we get a commutative diagram

$$\begin{tikzcd} & X \simeq \mathfrak{X}_\eta \arrow[dl, "\mathrm{red}"'] \arrow[dr, "\mathrm{sp}_{\mathfrak{X}}"] & \\ \widetilde{X} \arrow[rr, "\iota"'] & & \mathfrak{X}_s \end{tikzcd}$$

If $\mathcal{A}$ is reduced, then $A := \mathcal{A}^\circ$ is a special R-model and in that case $J = \check{\mathcal{A}}$ (see Remark 4.3). Hence in that case, φ and ι are isomorphisms. In general, since $\widetilde{\mathcal{A}}$ is a finitely generated $\tilde{k}$-algebra, we can find $f_1, \ldots, f_m \in \mathcal{A}^\circ$ such that $\widetilde{f_1}, \ldots, \widetilde{f_m}$ generate $\widetilde{\mathcal{A}}$. By [Kappen 2012, Corollary 2.12], $A' := A[f_1, \ldots, f_m] \subset \mathcal{A}^\circ$ is still a special R-model of $\mathcal{A}$, and by construction, if J' is the biggest ideal of definition of A', the map $\varphi : A'/J' \to \widetilde{\mathcal{A}}$ is surjective. Since φ is injective anyway, it is an isomorphism. In conclusion, if X is a semiaffinoid k-analytic space, one can always find a special R-scheme $\mathfrak{X} = \mathrm{Spf}(A)$ which is a model of X such that $\widetilde{X} \simeq \mathfrak{X}_s$ and such that red can be identified with $\mathrm{sp}_{\mathfrak{X}}$.

Corollary 4.9. *Let X be a semiaffinoid k-analytic space. Let Z be a Zariski closed subset of $\widetilde{X}$ and let $Y := \mathrm{red}^{-1}(Z)$. Then Y is a semiaffinoid k-analytic space and the naturally induced map $\widetilde{Y} \to \widetilde{X}$ is a closed immersion with image Z.*

Proof. Let $\mathcal{A}$ be the semiaffinoid k-algebra of X. Using Remark 4.7, we can replace $\mathcal{A}$ by $\mathcal{A}_{\mathrm{red}}$ and assume that $\mathcal{A}$ is reduced. Under this assumption, by [Kappen 2012, Corollary 2.11], $\mathcal{A}^\circ$ is a special R-algebra, and $X \simeq \mathrm{Spf}(\mathcal{A}^\circ)_\eta$. Let I be an ideal of $\mathcal{A}^\circ$ such that $Z = V(\widetilde{I})$, where $\widetilde{I} = \{\tilde{f} \mid f \in I\}$. Then as already seen in the proof of Theorem 3.1, $Y \simeq \mathrm{Spf}((\mathcal{A}^\circ)^{\wedge I})_\eta$, hence Y is the k-analytic space associated to the semiaffinoid k-algebra $\mathcal{B} := (\mathcal{A}^\circ)^{\wedge I} \otimes_R k$. Then according to Theorem 3.1, $\mathcal{B}^\circ = (\mathcal{A}^\circ)^{\wedge I}$. Since $\check{\mathcal{A}}$ is an ideal of definition of $\mathcal{A}^\circ$ [Kappen 2012, Remark 2.8], $\check{\mathcal{A}} + I$ is also an ideal of definition of $\mathcal{B}^\circ$. Since $\check{\mathcal{B}}$ is the biggest ideal of definition of $\mathcal{B}^\circ$ [Kappen 2012, Remark 2.8], it follows that

$$\widetilde{\mathcal{B}} = \mathcal{B}^\circ/\check{\mathcal{B}} = \big((\mathcal{A}^\circ)^{\wedge I}/(\check{\mathcal{A}} + I)\big)_{\mathrm{red}} = (\widetilde{\mathcal{A}}/I)_{\mathrm{red}}. \qquad \square$$

Remark 4.10. In general, when U is an affinoid domain of the k-affinoid space X, it is difficult to describe the induced canonical reduction map $\widetilde{U} \to \widetilde{X}$. However, it is proved in [Bosch et al. 1984, Proposition 7.2.6.3] that when U is the tube of a principal open subset of $\widetilde{X}$ (i.e., when $U = \mathrm{red}^{-1}(D(\tilde{f}))$ for some $f \in \mathcal{A}^\circ$), then $\widetilde{U} \to \widetilde{X}$ is an open immersion with image $D(\tilde{f})$. The above corollary is the counterpart of [Bosch et al. 1984, Proposition 7.2.6.3] for Zariski closed subsets of the canonical reduction.

Lemma 4.11. *Let A be an R-flat special R-algebra and let $\mathfrak{X} = \mathrm{Spf}(A)$. Then the specialization map $\mathrm{sp}_{\mathfrak{X}} : \mathfrak{X}_\eta \to \mathfrak{X}_s$ is surjective.*

Proof. This proof is strongly inspired by the proof of Lemme 1.2 in [Poineau 2008]. Let $\tilde{x} \in \mathfrak{X}_s$, and let r be the transcendence degree of the field extension $\tilde{k}(\tilde{x})/\tilde{k}$. Let K be the completion of $k(U_1, \dots, U_r)$ with respect to the Gauss norm. Then K is a discretely valued nonarchimedean field extension of k, and $\widetilde{K} \simeq \tilde{k}(U_1, \dots, U_r)$. Hence $\tilde{x}$ is the image of a closed point $\tilde{y} \in \mathfrak{X}_s \times_{\mathrm{Spec}(\tilde{k})} \mathrm{Spec}(\widetilde{K})$ with respect to the canonical map $\mathfrak{X}_s \times_{\mathrm{Spec}(\tilde{k})} \mathrm{Spec}(\widetilde{K}) \to \mathfrak{X}_s$.

Using base change extension induced by the inclusion $R \to K^\circ$ (see [de Jong 1995, 7.2.6]),

$$\mathfrak{X}' := \mathfrak{X} \times_{\mathrm{Spf}(R)} \mathrm{Spf}(K^\circ)$$

is a special formal scheme over K° of the form $\mathfrak{X}' = \mathrm{Spf}(A')$, where $A' := A \widehat{\otimes}_R K^\circ$. Let J be the biggest ideal of definition of A. Then JA' is an ideal of definition of A' and $J' := \mathrm{rad}(JA')$ is the biggest ideal of definition of A'. There is a well-defined morphism of $\tilde{k}$-algebras $\alpha : A/J \otimes_{\tilde{k}} \widetilde{K} \to A'/J'$ defined by $\tilde{a} \otimes \tilde{\lambda} \mapsto \widetilde{a \otimes \lambda}$ for $a \in A$, $\lambda \in K^\circ$ and where $\tilde{\cdot}$ stands for the various residue maps. We claim that α induces an isomorphism

$$(A/J \otimes_{\tilde{k}} \widetilde{K})_{\mathrm{red}} \simeq A'/J'. \tag{4}$$

Indeed by construction α is surjective, so it remains to prove that $\ker\alpha$ is the nilradical of $A/J \otimes_{\tilde{k}} \widetilde{K}$. There is also a well-defined surjective map

$$\beta : A' \to A/J \otimes_{\tilde{k}} \widetilde{K}, \qquad a \otimes \lambda \mapsto \tilde{a} \otimes \tilde{\lambda},$$

and by construction $\alpha \circ \beta : A' \to A'/J'$ is the quotient map. So for some $a_i \in A$ and $\lambda_i \in K^\circ$, let $\tilde{z} = \sum_i \widetilde{a_i} \otimes \widetilde{\lambda_i} \in \ker \alpha$ and $z := \sum_i a_i \otimes \lambda_i \in A'$. Then $\beta(z) = \tilde{z}$, hence $z \in \ker \alpha \circ \beta$, hence $z \in J'$. Since $J' := \mathrm{rad}(JA')$, we have $z^n \in JA'$ for some $n \in \mathbb{N}^*$. Hence $\beta(z^n) = 0$ and $\tilde{z}^n = 0$, thus proving (4). We get a natural composite morphism

$$\iota : (\mathfrak{X}')_s \xrightarrow{\iota_1} \mathfrak{X}_s \times_{\mathrm{Spec}(\tilde{k})} \mathrm{Spec}(\widetilde{K}) \xrightarrow{\iota_2} \mathfrak{X}_s,$$

where ι_1 is induced by (4) and hence is bijective, and ι_2 is the canonical map. Thus we can identify $\tilde{y}$ with a closed point of $(\mathfrak{X}')_s$. We also get a commutative diagram of sets

$$\begin{tikzcd} (\mathfrak{X}')_\eta \arrow[r, "\alpha"] \arrow[d, "\mathrm{sp}_{\mathfrak{X}'}"'] & (\mathfrak{X})_\eta \arrow[d, "\mathrm{sp}_{\mathfrak{X}}"] \\ (\mathfrak{X}')_s \arrow[r, "\iota"'] & \mathfrak{X}_s \end{tikzcd}$$

Thanks to [Kappen 2012, Lemma 2.3 and Remark 2.5] we know that $\mathrm{sp}_{\mathfrak{X}'}$ induces a surjective map from the set of rigid points $(\mathfrak{X}')^{\mathrm{rig}}$ to the set of closed points of $(\mathfrak{X}')_s$. Hence we can find $y \in (\mathfrak{X}')_\eta$ such that $\mathrm{sp}_{\mathfrak{X}'}(y) = \tilde{y}$. In conclusion, if $x := \alpha(y)$ we get $\mathrm{sp}_{\mathfrak{X}}(x) = \tilde{x}$, proving the surjectivity of $\mathrm{sp}_{\mathfrak{X}}$. □

Corollary 4.12. *Let X be a semiaffinoid k-analytic space. Then the canonical reduction map* $\mathrm{red} : X \to \widetilde{X}$ *is surjective.*

Proof. This follows from Lemma 4.11 and Remark 4.8. □

Connected components. In this subsection, we consider a semiaffinoid k-algebra $\mathcal{A}$ with associated k-analytic space X. It follows from Theorem A.8 and Remark 2.7 that $\mathrm{Spec}(\mathcal{A})$ is connected if and only if X is connected. Indeed by Remark 2.7, we can assume that $\mathcal{A}$ is reduced, and we conclude since the connected components of X are in correspondence with the set of idempotents of $\Gamma(X, \mathcal{O}_X)$, which themselves are equal to the set of idempotents of $\{f \in \Gamma(X, \mathcal{O}_X) \mid |f| < \infty\} = \mathcal{A}$. From the Noetherianity of $\mathcal{A}$ it follows that $\mathcal{A}$ can be uniquely decomposed as

$$\mathcal{A} \simeq \mathcal{A}_1 \times \cdots \times \mathcal{A}_n, \tag{5}$$

where each $\mathcal{A}_i$ is a semiaffinoid k-algebra such that $\mathrm{Spec}(\mathcal{A}_i)$ is connected. If we denote by X_i the k-analytic space associated to $\mathcal{A}_i$ it follows that

$$X = X_1 \amalg \cdots \amalg X_n \tag{6}$$

is the decomposition of X in connected components. These remarks easily imply the following.

Lemma 4.13. *A semiaffinoid k-analytic space X is connected if and only if $\widetilde{X}$ is connected.*

Proof. The equations (5) and (6) imply that $\widetilde{X} \simeq \coprod \widetilde{X}_i$. So if X is not connected, $\widetilde{X}$ is also not connected. Conversely, since $\mathrm{red}: X \to \widetilde{X}$ is surjective (Corollary 4.12) and anticontinuous, if one has a decomposition $\widetilde{X} = U_1 \amalg U_2$ in two nonempty closed-open sets, then $X = \mathrm{red}^{-1}(U_1) \amalg \mathrm{red}^{-1}(U_2)$ is a decomposition of X in nonempty closed-open sets. □

Corollary 4.14. *Let X be a semiaffinoid k-analytic space. Let Z be a Zariski closed subset of $\widetilde{X}$ and let $Y := \mathrm{red}^{-1}(Z)$. Then Y is connected if and only if Z is connected.*

Proof. This follows from Corollary 4.9 and from Lemma 4.13. □

Using Theorem 2.1, one checks that if A is a reduced special R-algebra which is integrally closed in $A \otimes_R k$, then there is a one-to-one correspondence between the connected components of $\mathrm{Spec}(A)$ and the connected components of $\mathrm{Spf}(A)_\eta$.

Example 4.15. In general, if A is a reduced special R-algebra with associated k-analytic space X, the connected components of $\mathrm{Spec}(A)$ do not coincide with the connected components of X, as the example $A = \mathbb{Z}_p\langle T\rangle/(T^2 + pT)$ shows.

5. Analytic functions and formal functions on tubes: a graded version

In this section, we fix $\mathcal{A}$ a reduced semiaffinoid k-algebra. Let us recall that this implies that $\mathcal{A}^\circ$ is a special R-algebra [Kappen 2012, Corollary 2.11].

5.1. For $r \in \mathbb{R}_+^*$ we set

$$\mathcal{A}_r^\circ = \{f \in \mathcal{A} \mid |f|_{\sup} \le r\}, \qquad \mathcal{A}_r^{\circ\circ} = \{f \in \mathcal{A} \mid |f|_{\sup} < r\}, \qquad \widetilde{\mathcal{A}}_r^+ = \mathcal{A}_r^\circ/\mathcal{A}_r^{\circ\circ}.$$

By definition, $\mathcal{A}_1^\circ = \mathcal{A}^\circ$, $\mathcal{A}_1^{\circ\circ} = \mathcal{A}^{\circ\circ}$ and $\widetilde{\mathcal{A}}_1^+ = \widetilde{\mathcal{A}}^+$. We also set

$$\rho(\mathcal{A}) = \{|f|_{\sup} \mid f \in \mathcal{A},\ f \neq 0\} \subset \mathbb{R}_+^*.$$

By [Kappen 2010, Proposition 1.2.5.9], $\rho(\mathcal{A}) \subset \sqrt{|k^*|}$. We denote by G the subgroup of $\sqrt{|k^*|}$ generated by $\rho(\mathcal{A})$.

Lemma 5.2. *Let $r \in \mathbb{R}_+^*$. Then $\mathcal{A}_r^\circ$, $\mathcal{A}_r^{\circ\circ}$ and $\widetilde{\mathcal{A}}_r^+$ are finitely generated $\mathcal{A}^\circ$-modules.*

Proof. Let us pick some $\lambda \in k^*$ such that $|\lambda| r \le 1$. Then

$$\mathcal{A}_r^\circ \to \mathcal{A}^\circ{}_{|\lambda| r}, \qquad f \mapsto \lambda f$$

is an isomorphism of $\mathcal{A}^\circ$-modules. So, replacing r by $|\lambda| r$, we can assume that $r \le 1$. Then $\mathcal{A}_r^\circ$ is an ideal of $\mathcal{A}^\circ$, hence it is a finitely generated $\mathcal{A}^\circ$-module because $\mathcal{A}^\circ$ is Noetherian. We conclude since $\mathcal{A}_r^{\circ\circ}$ is a submodule, and $\widetilde{\mathcal{A}}_r^+$ a quotient, of $\mathcal{A}_r^\circ$. □

Lemma 5.3. *The index $[G : |k^*|]$ is finite. As a consequence, $G \simeq \mathbb{Z}$.*

Proof. Let $\kappa := \sup\{\lambda \mid \lambda < 1 \text{ and } \lambda \in \rho(\mathcal{A})\}$. We claim that κ is actually a maximum, that is to say, there exists $\lambda < 1$ with $\lambda \in \rho(\mathcal{A})$ such that $\kappa = \lambda$ (this implies in particular that $\kappa < 1$). Indeed, if κ was not a maximum, we could find an increasing sequence $(\lambda_n)_{n\in\mathbb{N}}$ with $\lambda_n < 1$ in $\rho(\mathcal{A})$. Then $I_n = \{f \in \mathcal{A} \mid |f|_{\sup} \le \lambda_n\}$ would be an infinite increasing sequence of ideals of $\mathcal{A}^\circ$, which is Noetherian. This would be a contradiction. Hence, $\kappa < 1$ and $\kappa \in \rho(\mathcal{A})$.

Since $\kappa \in \rho(\mathcal{A})$, we can write $\kappa = |\pi|^{a/b}$ with a and b relatively prime and $b > 0$. One can easily prove using Bézout's theorem that $a = 1$. Hence $\kappa = |\pi|^{1/b}$. Likewise, using again Bézout's theorem, one can prove that if $|\pi|^{\alpha/\beta} \in \rho(\mathcal{A})$ with $\beta > 0$ and α and β coprime, then $\beta \le b$. Hence, if $v := \operatorname{lcm}(1, 2, \dots, b)$, one has $\rho(\mathcal{A}) \subset |\pi|^{\mathbb{Z}/v}$. □

Remark 5.4. If $\|\cdot\|$ is a k-Banach algebra norm on $\mathcal{A}$, according to [Kappen 2010, Lemma 1.2.5.8], for any $f \in \mathcal{A}$, $|f|_{\sup} = \lim_{n\in\mathbb{N}} \sqrt[n]{\|f^n\|}$. The notation $\rho(\mathcal{A})$ that we have introduced is compatible with [Berkovich 1990, Section 1.3], where the spectral radius of an element of a Banach algebra $\mathcal{A}$ is defined to be $\rho(f) = \lim_{n\in\mathbb{N}} \sqrt[n]{\|f^n\|}$. Let us recall from [Kappen 2010, Lemma 1.2.5.4] that a surjective morphism of semiaffinoid k-algebras

$$R\langle T_1, \dots, T_m\rangle[\![S_1, \dots, S_n]\!] \otimes_R k \to \mathcal{A}$$

induces a k-Banach norm on $\mathcal{A}$ by taking the residue seminorm of the Gauss norm on

$$R\langle T_1, \dots, T_m\rangle[\![S_1, \dots, S_n]\!] \otimes_R k.$$

Example 5.5. In general, $\rho(\mathcal{A})$ is not a subgroup, and not even a monoid. For instance if $\mathcal{A} = \mathbb{Q}_p(\sqrt{p}) \times \mathbb{Q}_p(\sqrt[3]{p})$ then $\rho(\mathcal{A}) = p^{\mathbb{Z}/2} \cup p^{\mathbb{Z}/3}$, which is not a monoid.

The following definition is inspired by [Temkin 2004, Section 3].

Definition 5.6. We define the *total reduction ring of* $\mathcal{A}$ as

$$\widetilde{\mathcal{A}}^+_{\mathrm{tot}} := \bigoplus_{r\in G} \widetilde{\mathcal{A}}^+_r.$$

Corollary 5.7. *The total reduction ring $\widetilde{\mathcal{A}}^+_{\mathrm{tot}}$ of $\mathcal{A}$ is excellent and reduced.*

Proof. By Lemma 5.3, $[G : |k^*|]$ is finite, so we can introduce $n := [G : |k^*|] \in \mathbb{N}$. Let $r_1, \dots, r_n \in G$ be some representatives of the classes of $G/|k^*|$. For each $i \in \{1, \dots, n\}$, by Lemma 5.2, we can find a finite family $(f_{i,j})_{j\in J}$ which is a finite set of generators of the $\mathcal{A}^\circ$-module $\widetilde{\mathcal{A}}^+_{r_i}$. Let us denote by $\tilde{\pi}$ the image of π in $\widetilde{\mathcal{A}}^+_{1/|\pi|}$, and likewise by $\widetilde{(1/\pi)}$ the image of $1/\pi$ in $\widetilde{\mathcal{A}}^+_{1/|\pi|}$. We get that

$$\bigoplus_{r\in G} \widetilde{\mathcal{A}}^+_r = \widetilde{\mathcal{A}}^+\big[\tilde{\pi}, \widetilde{\big(\tfrac{1}{\pi}\big)}, f_{i,j}\big].$$

Hence $\widetilde{\mathcal{A}}^+_{\mathrm{tot}}$ is a finitely generated $\widetilde{\mathcal{A}}^+$-algebra. Since $\widetilde{\mathcal{A}}^+$ is excellent, it follows that $\widetilde{\mathcal{A}}^+_{\mathrm{tot}}$ is also excellent. For the reducedness, consider a nonzero element $\tilde{f} \in \widetilde{\mathcal{A}}^+_r$ for some $f \in \mathcal{A}$ with $|f|_{\sup} = r$. Then for any integer $k > 0$, the associated element $\widetilde{f^k} = \tilde{f}^k \in \widetilde{\mathcal{A}}^+_{r^k}$ is also nonzero since $|f^k|_{\sup} = |f|^k_{\sup} = r^k$. $\square$

Remark 5.8. Let $r \in \rho(\mathcal{A})$. Since $\rho(\mathcal{A})$ is discrete in $\mathbb{R}^*_+$ (see Lemma 5.3), there exists a real number $s \in \rho(\mathcal{A})$ which is the biggest element of $\rho(\mathcal{A})$ such that $s < r$. It follows that $\mathcal{A}^{\circ\circ}_r = \mathcal{A}^\circ_s$.

We fix I an ideal of $\mathcal{A}^\circ$. If M is an $\mathcal{A}^\circ$-module, we denote by $M^{\wedge I}$ the completion of M with respect to the I-adic topology. So

$$M^{\wedge I} \simeq \varprojlim_n M/I^n M.$$

Lemma 5.9. *Let $r \in \rho(\mathcal{A})$. There is a short exact sequence of $\mathcal{A}^\circ$-modules*

$$0 \to (\mathcal{A}^{\circ\circ}_r)^{\wedge I} \to (\mathcal{A}^\circ_r)^{\wedge I} \to (\widetilde{\mathcal{A}}^+_r)^{\wedge I} \to 0. \tag{7}$$

Proof. By definition of $\widetilde{\mathcal{A}}^+_r$, there is a short exact sequence

$$0 \to \mathcal{A}^{\circ\circ}_r \to \mathcal{A}^\circ_r \to \widetilde{\mathcal{A}}^+_r \to 0 \tag{8}$$

of finitely generated $\mathcal{A}^\circ$-modules. So the I-adic completion of (8) remains exact; see [Matsumura 1989, Theorems 8.7 and 8.8]. $\square$

5.10. Let I be an ideal of $\mathcal{A}^\circ$ and let us set $U = \{x \in X \mid |f(x)| < 1 \ \forall f \in I\}$. Let us denote by $\mathcal{B}$ its associated semiaffinoid k-algebra, which can be defined as

$$\mathcal{B} = (\mathcal{A}^\circ[\![S_1, \dots, S_n]\!]/(f_i - S_i)_{i=1,\dots,n}) \otimes_R k,$$

where $I = (f_1, \dots, f_n)$. According to Theorem 3.1, one has $\mathcal{B}^\circ \simeq (\mathcal{A}^\circ)^{\wedge I}$. We denote by Y the k-analytic space associated with $\mathcal{B}$. If $g \in \mathcal{B}$, set $|g|_{\sup} = \sup_{y \in Y} |g(y)|$.

Remark 5.11. Let $r \in |k^*|$ and let $\lambda \in k$ with $|\lambda| = r^{-1}$. Multiplication by λ induces an isomorphism of $\mathcal{A}^\circ$-modules $\mathcal{A}^\circ_r \xrightarrow{\times\lambda} \mathcal{A}^\circ$. Completing with respect to I one gets an isomorphism of $(\mathcal{A}^\circ)^{\wedge I}$-modules $(\mathcal{A}^\circ_r)^{\wedge I} \simeq (\mathcal{A}^\circ)^{\wedge I}$. Finally, using Theorem 3.1 we get an isomorphism of $(\mathcal{A}^\circ)^{\wedge I}$-modules $\mathcal{B}^\circ_r \simeq (\mathcal{A}^\circ_r)^{\wedge I}$ obtained as the composition

$$(\mathcal{A}^\circ_r)^{\wedge I} \xrightarrow{\times\lambda} (\mathcal{A}^\circ)^{\wedge I} \to \mathcal{B}^\circ \xrightarrow{\times\lambda^{-1}} \mathcal{B}^\circ_r.$$

More generally, if $r \in \rho(\mathcal{A})$, we can find $s \in |k^*|$ with $r \leq s$, leading to an inclusion of $\mathcal{A}^\circ$-modules $\mathcal{A}^\circ_r \to \mathcal{A}^\circ_s$. Then completing with respect to I we get an inclusion of $(\mathcal{A}^\circ)^{\wedge I}$-modules $(\mathcal{A}^\circ_r)^{\wedge I} \to (\mathcal{A}^\circ_s)^{\wedge I}$. Using the above identifications, we can assimilate $(\mathcal{A}^\circ_r)^{\wedge I}$ as a $\mathcal{B}^\circ$-submodule of $\mathcal{B}^\circ_s$, hence as a $\mathcal{B}^\circ$-submodule of $\mathcal{B}$.

Lemma 5.12. *Let $r \in |k^*|$.*

(i) *Let $g \in \mathcal{B}_r^\circ \simeq (\mathcal{A}_r^\circ)^{\wedge I}$ (see Remark 5.11). Let $\tilde{g} \in (\widetilde{\mathcal{A}}_r^+)^{\wedge I}$ be the image of g by the reduction map of the short exact sequence (7) of Lemma 5.9. Then $\tilde{g} = 0$ if and only if $|g|_{\sup} < r$. Equivalently, $\tilde{g} \neq 0$ if and only if $|g|_{\sup} = r$.*

(ii) *There is a natural isomorphism $\mathcal{B}_r^{\circ\circ} \simeq (\mathcal{A}_r^{\circ\circ})^{\wedge I}$.*

Proof. Using the same arguments as in Remark 5.11, we can assume that $r = 1$. We then consider the short exact sequence of Lemma 5.9 for $r = 1$:

$$0 \to (\mathcal{A}^{\circ\circ})^{\wedge I} \to (\mathcal{A}^\circ)^{\wedge I} \to (\widetilde{\mathcal{A}}^+)^{\wedge I} \to 0.$$

Let us then consider $g \in \mathcal{B}^\circ \simeq (\mathcal{A}^\circ)^{\wedge I}$ and let us assume that $\tilde{g} = 0$. This implies that $g \in (\mathcal{A}^{\circ\circ})^{\wedge I}$. By Remark 5.8, there exists $s < 1$ such that $(\mathcal{A}^{\circ\circ})^{\wedge I} = (\mathcal{A}_s^\circ)^{\wedge I}$ for some $s < 1$. It follows that $|g|_{\sup} \le s < 1$. Conversely, let us assume that $|g|_{\sup} < 1$. There exists an integer $d \in \mathbb{N}$ such that $|g^d|_{\sup} \le |\pi|$. Hence, according to Remark 5.11, $g^d \in \mathcal{B}_{|\pi|}^\circ \simeq (A_{|\pi|}^\circ)^{\wedge I} \subset (\mathcal{A}^{\circ\circ})^{\wedge I}$. This proves (i), and (ii) follows from (i). $\square$

We can now generalize Theorem 3.1 to an arbitrary $r \in \rho(\mathcal{A})$.

Proposition 5.13. *We use the notations of 5.10.*

(i) *There is an inclusion $\rho(\mathcal{B}) \subset \rho(\mathcal{A})$.*

(ii) *Let $r \in \rho(\mathcal{A})$. There are isomorphisms $\mathcal{B}_r^\circ \simeq (\mathcal{A}_r^\circ)^{\wedge I}$, $\mathcal{B}_r^{\circ\circ} \simeq (\mathcal{A}_r^{\circ\circ})^{\wedge I}$ and $\widetilde{\mathcal{B}}_r^+ \simeq (\widetilde{\mathcal{A}}_r^+)^{\wedge I}$.*

(iii) *There is a natural isomorphism $\widetilde{\mathcal{B}}_{\mathrm{tot}}^+ \simeq \bigoplus_{r\in G} (\widetilde{\mathcal{A}}_r^+)^{\wedge I}$.*

Proof. Let $g \in \mathcal{B}$ be a nonzero element. Then there exists $s \in \rho(\mathcal{A})$ such that $g \in \mathcal{B}_s^\circ$. Since $\rho(\mathcal{A})$ is discrete by Lemma 5.3, we can then define the smallest element $r \in \rho(\mathcal{A})$ such that $g \in \mathcal{B}_r^\circ$. By Remark 5.11, for each $r \in \rho(\mathcal{A})$, we can naturally identify $(\mathcal{A}_r^\circ)^{\wedge I}$ with a $\mathcal{B}^\circ$-submodule of $\mathcal{B}$, and under these identifications, $\mathcal{B} = \bigcup_{r\in\rho(\mathcal{A})} (\mathcal{A}_r^\circ)^{\wedge I}$. Since $\rho(\mathcal{A})$ is discrete, we can then define the smallest element $r \in \rho(\mathcal{A})$ such that $g \in (\mathcal{A}_r^\circ)^{\wedge I}$. We then consider the short exact sequence (7):

$$0 \to (\mathcal{A}_r^{\circ\circ})^{\wedge I} \to (\mathcal{A}_r^\circ)^{\wedge I} \to (\widetilde{\mathcal{A}}_r^+)^{\wedge I} \to 0.$$

The minimality of r and Remark 5.8 imply that $g \notin (\mathcal{A}_r^{\circ\circ})^{\wedge I}$. It follows that $\tilde{g} \neq 0$, where $\tilde{g} \in (\widetilde{\mathcal{A}}_r^+)^{\wedge I}$ denotes the reduction of g in $(\widetilde{\mathcal{A}}_r^+)^{\wedge I}$. Thanks to Lemma 5.3 we can pick some $d \in \mathbb{N}^*$ such that $r^d \in |k^*|$. Then $g^d \in (\mathcal{A}_{r^d}^\circ)^{\wedge I}$. Since $(\widetilde{\mathcal{A}}_{\mathrm{tot}}^+)$ is reduced and excellent (Corollary 5.7), it follows that $(\widetilde{\mathcal{A}}_{\mathrm{tot}}^+)^{\wedge I}$ is also reduced. Hence $\widetilde{g^d} = \tilde{g}^d \neq 0$ in $(\widetilde{\mathcal{A}}_{r^d}^+)^{\wedge I}$. Since $r^d \in |k^*|$, Lemma 5.12 implies that $|g^d|_{\sup} = r^d$. So $|g|_{\sup} = r$. This proves (i) as well as (ii) and (iii). $\square$

We also obtain the following generalization of [Bosch and Lütkebohmert 1985, Lemma 2.1].

Corollary 5.14. *Let X be a k-affinoid space, and let $Z \subset \widetilde{X}$ be a Zariski closed subset. Then $\widehat{\widetilde{X}_{/Z}}$, the formal completion of $\widetilde{X}$ along Z, depends intrinsically on the k-analytic space* $\mathrm{red}^{-1}(Z)$.

Proof. Let us denote by $\mathcal{B}$ the semiaffinoid k-algebra of $\mathrm{red}^{-1}(Z)$. The space $\widehat{\widetilde{X}_{/Z}}$ is the formal scheme associated to the adic $\tilde{k}$-algebra $(\widetilde{\mathcal{A}}^{+})^{\wedge I}$. Thanks to the short exact sequence (7) for $r = 1$,

$$(\widetilde{\mathcal{A}}^{+})^{\wedge I} \simeq \mathcal{A}^{\circ \wedge I}/\mathcal{A}^{\circ\circ \wedge I}.$$

Thanks to Theorem 3.1 and Lemma 5.12(ii), one gets that

$$(\widetilde{\mathcal{A}}^{+})^{\wedge I} \simeq \mathcal{B}^{\circ}/\mathcal{B}^{\circ\circ} = \widetilde{\mathcal{B}}^{+},$$

which depends intrinsically on $\mathrm{red}^{-1}(Z)$ since $\mathcal{B}$ does. □

Remark 5.15. More generally, if X is the k-analytic space associated to the reduced semiaffinoid k-algebra $\mathcal{A}$, then $\widetilde{\mathcal{A}}^{+}$ is naturally an adic algebra with an ideal of definition given by $\check{\mathcal{A}}$. This adic algebra is isomorphic to a quotient of $\tilde{k}[T_i][\![S_j]\!]$. Let us set $\mathcal{X} := \mathrm{Spf}(\widetilde{\mathcal{A}}^{+})$. Then $|\mathcal{X}| = \widetilde{X}$. Let Z be a Zariski closed subset of $\widetilde{X}$, Y the k-analytic space defined by $Y = \mathrm{red}^{-1}(Z)$ and $\mathcal{B}$ its associated semiaffinoid k-algebra. One shows similarly that the inclusion of analytic domains $Y \to X$ induces an isomorphism $\widehat{\mathcal{X}_{/Z}} \simeq \mathrm{Spf}(\widetilde{\mathcal{B}}^{+})$, where we denote by $\widehat{\mathcal{X}_{/Z}}$ the completion of $\mathcal{X}$ along Z. In particular, $\widehat{\mathcal{X}_{/Z}}$ depends intrinsically on the k-analytic space $\mathrm{red}^{-1}(Z)$.

6. Additional remarks

Finite morphisms.

Proposition 6.1. *Let $\varphi : \mathcal{B} \to \mathcal{A}$ be a finite morphism of semiaffinoid k-algebras, with $\mathcal{A}$ reduced. Then $\varphi^{\circ} : \mathcal{B}^{\circ} \to \mathcal{A}^{\circ}$ is finite.*

Proof. Since $\mathcal{B}^{\circ} \to (\mathcal{B}_{\mathrm{red}})^{\circ}$ is finite, we can also assume that $\mathcal{B}$ is reduced. Let $f_1, \ldots, f_n$ be elements of $\mathcal{A}$ such that

$$\mathcal{A} = \mathcal{B}[f_1, \ldots, f_n]. \tag{9}$$

Each $f \in \{f_1, \ldots, f_n\}$ satisfies a unitary polynomial equation with coefficients in $\mathcal{B}$ of the form

$$f^d + b_{d-1} f^{d-1} + \cdots + b_1 f + b_0 = 0. \tag{10}$$

Then for $m \in \mathbb{N}^*$, multiplying by π^{md}, (10) becomes

$$(\pi^m f)^d + \pi^m b_{d-1} (\pi^m f)^{d-1} + \cdots + \pi^{m(d-1)} b_1 (\pi^m f) + \pi^{md} b_0 = 0.$$

But for m big enough, all the coefficients $\pi^m b_{d-1}, \ldots, \pi^{m(d-1)} b_1, \pi^{md} b_0$ appearing in the above equation belong to $\mathcal{B}^{\circ}$. Hence, for m big enough, $\pi^m f_i$ satisfies a unitary polynomial equation with coefficients in $\mathcal{B}^\circ$. So replacing each f_i by $\pi^m f_i$ (which will not change (9)), we can assume that the f_i belong to $\mathcal{A}^\circ$ and are integral over $\mathcal{B}^\circ$. So, thanks to (9), $\mathcal{B}^\circ[f_1, \dots, f_n]$ is a special R-model of $\mathcal{A}$. According to [Kappen 2012, Corollary 2.10], $\mathcal{A}^\circ$ is finite over $\mathcal{B}^\circ[f_1, \dots, f_n]$, and hence also over $\mathcal{B}^\circ$. □

We conjecture that for an arbitrary nonarchimedean nontrivially valued field k, a similar statement holds for quasiaffinoid k-algebras (see [Lipshitz and Robinson 2000, Remark 2.1.8] for the definition of a quasiaffinoid k-algebra).

Corollary 6.2. *Let $\varphi : \mathcal{A} \to \mathcal{B}$ be a finite morphism of semiaffinoid k-algebras. The associated morphisms $\tilde{\varphi} : \widetilde{\mathcal{A}} \to \widetilde{\mathcal{B}}$, $\tilde{\varphi}^+ : \widetilde{\mathcal{A}}^+ \to \widetilde{\mathcal{B}}^+$, $\tilde{\varphi}_{\mathrm{tot}} : \widetilde{\mathcal{A}}^+_{\mathrm{tot}} \to \widetilde{\mathcal{B}}^+_{\mathrm{tot}}$ are finite.*

Proof. Since the above morphisms do not change if one replaces $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{A}_{\mathrm{red}}$ and $\mathcal{B}_{\mathrm{red}}$, we can assume that $\mathcal{A}$ and $\mathcal{B}$ are reduced. The first two points then follow from Proposition 6.1. To prove that $\tilde{\varphi}_{\mathrm{tot}}$ is finite, one has to use that $\tilde{\varphi}^+$ is finite, and then argue as in the proof of Corollary 5.7. □

The nonaffine case.

Lemma 6.3. *Let A be a reduced special R-algebra which is integrally closed in the semiaffinoid k-algebra $A \otimes_R k$. Let $\mathfrak{U}$ be a formal open affine subset of $\mathfrak{X} = \mathrm{Spf}(A)$. Then $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$ is also a reduced special R-algebra which is integrally closed in the semiaffinoid k-algebra $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \otimes_R k$ and $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \simeq \Gamma(\mathrm{sp}_{\mathfrak{X}}^{-1}(\mathfrak{U}), \mathcal{O}^\circ_{\mathfrak{X}_\eta})$.*

Proof. We first assume that $\mathfrak{U}$ is a principal formal open subset of the form $\mathfrak{U} = \mathfrak{D}(f)$ for some $f \in A$. Let J be an ideal of definition of A. Then $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \simeq A_{\{f\}}$, where $A_{\{f\}} \simeq \widehat{A_f}$ is the completion of the localization A_f with respect to the ideal JA_f. The composition morphism $A \to A_f \to \widehat{A_f}$ is regular. Indeed, $A \to A_f$ is regular since it is a localization, and $A_f \to \widehat{A_f}$ is regular since it is the completion of an excellent ring. It follows that the morphism $A \to \widehat{A_f}$ is regular since regular morphisms are stable under composition [EGA IV$_2$ 1965, Proposition 6.8.3]. Using [EGA IV$_2$ 1965, Proposition 6.14.4], we conclude that $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \simeq \widehat{A_f}$ is integrally closed in $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \otimes_R k$.

Let us now assume that $\mathfrak{U}$ is an arbitrary formal open affine subset of $\mathfrak{X}$. It follows from [de Jong 1995, §7, p. 74–75] that $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$ is a special R-algebra. Moreover, we can cover $\mathfrak{U}$ by some principal formal open subsets of the form $\mathfrak{D}(f_i)$ for some $f_i \in A$. Let us now consider an element $g \in \Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \otimes_R k$ which is integral over $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$. This means that g satisfies an equation

$$g^d + b_{d-1} g^{d-1} + \cdots + b_1 g + b_0 = 0 \tag{11}$$

for some $b_j \in \Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$. But for each principal formal open subset $\mathfrak{D}(f_i)$, we can restrict (11) to $\Gamma(\mathfrak{D}(f_i), \mathcal{O}_{\mathfrak{X}}) \otimes_R k$. We then get that $g|_{\mathfrak{D}(f_i)} \in \Gamma(\mathfrak{D}(f_i), \mathcal{O}_{\mathfrak{X}}) \otimes_R k$ is integral over $\Gamma(\mathfrak{D}(f_i), \mathcal{O}_{\mathfrak{X}})$, so by the first part of the proof, $g|_{\mathfrak{D}(f_i)} \in \Gamma(\mathfrak{D}(f_i), \mathcal{O}_{\mathfrak{X}})$.

Since the $\mathfrak{D}(f_i)$ form a covering of $\mathfrak{U}$, we deduce that $g \in \Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$, which proves that $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$ is integrally closed in $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \otimes_R k$. Finally, the equality $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \simeq \Gamma(\mathrm{sp}_{\mathfrak{X}}^{-1}(\mathfrak{U}), \mathcal{O}^{\circ}_{\mathfrak{X}_\eta})$ now follows from Theorem 2.1. □

The following result extends [Bosch et al. 1984, Proposition 7.2.6.3] from affinoid to semiaffinoid k-algebras.

Corollary 6.4. *Let X be a semiaffinoid k-analytic space. Let $f \in \Gamma(X, \mathcal{O}^{\circ}_X)$ and let $Y = \{x \in X \mid |f(x)| = 1\}$. Then Y is a semiaffinoid k-analytic space and the associated map $\widetilde{Y} \to \widetilde{X}$ is the Zariski open embedding of the principal open subset $D(\tilde{f}) \subset \widetilde{X}$.*

Proof. It follows from general properties of semiaffinoid k-analytic spaces that the inclusion $Y \to X$ is induced by a morphism of semiaffinoid k-algebras (see [de Jong 1995, Proposition 7.2.1(a)]). Using Remark 4.7, we can easily assume that X and Y are reduced. Let $\mathcal{A}$ be the reduced semiaffinoid k-algebra associated with X. By [Kappen 2012, Corollary 2.11], $\mathcal{A}^{\circ}$ is a special R-algebra. Let $\mathfrak{X} := \mathrm{Spf}(\mathcal{A}^{\circ})$. Let $\mathfrak{D}(f) \subset \mathfrak{X}$ be the principal formal open subset associated with f. Hence $Y = \mathrm{sp}_{\mathfrak{X}}^{-1}(\mathfrak{D}(f))$. By Lemma 6.3, we have $\Gamma(\mathfrak{D}(f), \mathcal{O}_{\mathfrak{X}}) \simeq \mathcal{B}^{\circ}$, where $\mathcal{B}$ is the semiaffinoid k-algebra associated with Y. Let $\mathfrak{J}$ be the biggest ideal sheaf of definition of $\mathfrak{X}$ (see [EGA I 1960, Proposition 10.5.4] for the definition and the properties of $\mathfrak{J}$). By [EGA I 1960, 10.5.2] there is an isomorphism of schemes $(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}/\mathfrak{J}) \simeq \mathrm{Spec}(\widetilde{\mathcal{A}}) = \widetilde{X}$. Moreover, by [EGA I 1960, Corollaire 10.5.5], $\mathfrak{J}|_{\mathfrak{U}}$ is also the biggest ideal sheaf of definition of the formal scheme $\mathfrak{U}$. Hence $\Gamma(\mathfrak{U}, \mathfrak{J})$ is the biggest ideal of definition of the adic ring $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \simeq \mathcal{B}^{\circ}$, and as a consequence $\Gamma(\mathfrak{U}, \mathfrak{J}) \simeq \check{\mathcal{B}}$ by [Kappen 2012, Remark 2.8]. Hence we can conclude, since

$$\widetilde{\mathcal{B}} = \mathcal{B}^{\circ}/\check{\mathcal{B}} \simeq \Gamma(\mathfrak{D}(f), \mathcal{O}_{\mathfrak{X}}/\mathfrak{J}) \simeq \widetilde{\mathcal{A}}[\tilde{f}^{-1}].$$ □

Lemma 6.5. *Let $\mathfrak{X}$ be a special formal scheme over R. The following are equivalent.*

(i) *Any formal open affine subscheme of $\mathfrak{X}$ is isomorphic to* $\mathrm{Spf}(A)$, *where A is a reduced special R-algebra integrally closed in $A \otimes_R k$.*

(ii) *There exists a covering by formal open affine subschemes $\mathfrak{U}_i = \mathrm{Spf}(A_i)$, where for each i, A_i is a reduced special R-algebra which is integrally closed in $A_i \otimes_R k$.*

Proof. That (i) implies (ii) is clear. To prove the converse implication, let $\mathfrak{U}$ be a formal open affine subset of $\mathfrak{X}$. Then $\mathfrak{U}$ is covered by finitely many $\mathfrak{U} \cap \mathfrak{U}_i$. For each i we can find a finite covering $\{\mathfrak{U}_{i,j}\}$ of $\mathfrak{U} \cap \mathfrak{U}_i$ by formal open affine subsets. Thanks to Lemma 6.3, all the $\mathfrak{U}_{i,j}$ satisfy the expected property. We are then reduced to the situation where $\mathfrak{X}$ is affine and is covered by finitely many formal open affine $\mathfrak{U}_i$'s which all satisfy the expected property. Let us then consider a function f in the integral closure of $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$ in $\Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}}) \otimes_R k$. Then for all i, $f|_{\mathfrak{U}_i}$ is in the integral closure of $\Gamma(\mathfrak{U}_i, \mathcal{O}_{\mathfrak{X}})$ in $\Gamma(\mathfrak{U}_i \mathcal{O}_{\mathfrak{X}}) \otimes_R k$, which is by assumption $\Gamma(\mathfrak{U}_i, \mathcal{O}_{\mathfrak{X}})$. So $f \in \Gamma(\mathfrak{U}, \mathcal{O}_{\mathfrak{X}})$. $\square$

Appendix: Bounded functions on reduced semiaffinoid *k*-spaces

by Christian Kappen and Florent Martin

In this appendix, we assume that k is a nontrivially discretely valued nonarchimedean field, and we let R denote its valuation ring. We fix a reduced semiaffinoid k-algebra $\mathcal{A}$, and we let X denote the associated rigid analytic k-space.

Projective limits. Let us start with a reminder on derived functors of projective limits of abelian groups. Let

$$\cdots \xrightarrow{\sigma_3} G_2 \xrightarrow{\sigma_2} G_1 \xrightarrow{\sigma_1} G_0$$

be a projective system of abelian groups indexed by $\mathbb{N}$, and let

$$\varphi : \prod_{n\in\mathbb{N}} G_n \to \prod_{n\in\mathbb{N}} G_n, \qquad (a_n)_n \mapsto (a_n - \sigma_{n+1}(a_{n+1}))_n.$$

Then $\ker\varphi \simeq \varprojlim G_n$. Let us denote by $\varprojlim^i$ the i-th derived functor of $\varprojlim$. According to [Weibel 1994, Corollary 3.5.4], one has the following descriptions:

$$\begin{aligned} &{\varprojlim}^1 G_n \simeq \operatorname{coker}\varphi, \\ &{\varprojlim}^i G_n = 0 \qquad\qquad \text{for } i > 1. \end{aligned}$$

A flatness result. Let us fix a presentation $\mathcal{A} \simeq (k \otimes_R R\langle T_1, \dots, T_m\rangle[[S_1, \dots, S_n]])/I$ and let us equip $\mathcal{A}$ with the associated k-Banach algebra norm $\|\cdot\|$ as in [Kappen 2010, Section 1.2.5]. For a real number $\varepsilon \in \sqrt{|k^*|}$ such that $0 < \varepsilon < 1$, we set

$$X_\varepsilon := \{x \in X \mid |S_i(x)| \leq \varepsilon,\ i = 1, \dots, n\}.$$

Then X_ε is an affinoid k-space (depending on the chosen presentation of $\mathcal{A}$), and we let $\mathcal{A}_\varepsilon$ denote the associated affinoid k-algebra. It comes with a natural presentation

$$\mathcal{A}_\varepsilon \simeq k\langle T_1, \dots, T_m, \varepsilon^{-1}S_1, \dots, \varepsilon^{-1}S_n\rangle/I$$

and with an associated k-Banach algebra norm $\|\cdot\|_\varepsilon$ such that if $\varepsilon \leq \varepsilon'$, the restriction morphism $\mathcal{A}_{\varepsilon'} \to \mathcal{A}_\varepsilon$ is contractive, that is to say, for $f \in \mathcal{A}_{\varepsilon'}$ we have $\|f\|_\varepsilon \leq \|f\|_{\varepsilon'}$.

Let us now fix an increasing sequence of positive real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \varepsilon_n = 1$ and such that for all $n \in \mathbb{N}$, we have that $\varepsilon_n \in \sqrt{|k^*|}$. We set $\mathcal{A}_n := \mathcal{A}_{\varepsilon_n}$. We denote by X_n the k-affinoid space associated to $\mathcal{A}_n$. For $n \in \mathbb{N}$ we denote by $\tau_n : \mathcal{A} \to \mathcal{A}_n$ the associated canonical map, and for $m \geq n \in \mathbb{N}$ we denote by $\sigma_{m,n} : \mathcal{A}_m \to \mathcal{A}_n$ the restriction morphism. By the above remark, each $\sigma_{m,n}$ is a contractive morphism with respect to the norms $\|\cdot\|_{\varepsilon_m}$ and $\|\cdot\|_{\varepsilon_n}$. The sequence

$(\mathcal{A}_n)_{n\in\mathbb{N}}$ and the restriction morphisms form a projective system of abelian groups. The next lemma is inspired by [Bosch 1977, Satz 2.1], and the proof is verbatim the same as in [Bosch 1977]. For the convenience of the reader, we recall it.

Lemma A.1. *In the above setting*, $\varprojlim^1 \mathcal{A}_n = 0$.

Proof. We have to show that the map

$$\varphi : \prod_{n\in\mathbb{N}} \mathcal{A}_n \to \prod_{n\in\mathbb{N}} \mathcal{A}_n, \qquad (a_n)_n \mapsto (a_n - \sigma_{n+1,n}(a_{n+1}))_n,$$

is surjective. So let us consider a sequence $(g_n) \in \prod_{n\in\mathbb{N}} \mathcal{A}_n$, and let us find a sequence $(f_n) \in \prod_{n\in\mathbb{N}} \mathcal{A}_n$ satisfying the conditions

$$g_n = f_n - \sigma_{n+1,n}(f_{n+1}) \quad \forall n \in \mathbb{N}. \tag{12}$$

Step 1. Let us first assume that for all $n \in \mathbb{N}$, g_n lies in the image of the restriction map $\tau_n : \mathcal{A} \to \mathcal{A}_n$. For each n, let us then choose $G_n \in \mathcal{A}$ such that $g_n = \tau_n(G_n)$. We define inductively a sequence $F_n \in \mathcal{A}$ via

$$\begin{cases} F_0 := 0, \\ F_{n+1} := F_n - G_n \quad \text{for } n \geq 0. \end{cases}$$

Setting $f_n := \tau_n(F_n)$, we obtain a solution $(f_n)_n$ of (12).

Step 2. Let us now pick some arbitrary $(g_n) \in \prod_n \mathcal{A}_n$. For each $n \in \mathbb{N}$, the image of $\mathcal{A}$ in $\mathcal{A}_n$ is dense with respect to the topology induced by $\|\cdot\|_{\varepsilon_n}$. Hence for all $n \in \mathbb{N}$, there exists $h_n \in \mathcal{A}$ such that $\|g_n - \tau_n(h_n)\|_{\varepsilon_n} \leq 2^{-n}$. For $n \in \mathbb{N}$, we have $g_n = \tau_n(h_n) + (g_n - \tau_n(h_n))$. By Step 1, there exists $(H_n)_{n\in\mathbb{N}} \in \prod_n \mathcal{A}_n$ such that $\varphi((H_n)_{n\in\mathbb{N}}) = (\tau_n(h_n))_{n\in\mathbb{N}}$. Hence it remains to prove that $(g_n - \tau_n(h_n))_{n\in\mathbb{N}} \in \operatorname{im}(\varphi)$. Replacing g_n by $g_n - \tau_n(h_n)$, we can thus assume that

$$\|g_n\|_{\varepsilon_n} \leq 2^{-n} \quad \forall n \in \mathbb{N}.$$

Since the morphisms $\sigma_{m,n}$ are contractive, for each $m \geq n$ we have $\|\sigma_{m,n}(g_m)\|_{\varepsilon_n} \leq \|g_m\|_{\varepsilon_m} \leq 2^{-m}$. Hence, for each $n \in \mathbb{N}$, since $\mathcal{A}_n$ is a k-Banach algebra, it makes sense to define

$$f_n := \sum_{m\geq n} \sigma_{m,n}(g_m).$$

Finally, we have

$$\begin{aligned} f_n - \sigma_{n+1,n}(f_{n+1}) &= \sum_{m\geq n} \sigma_{m,n}(g_m) - \sigma_{n+1,n}\Bigg(\sum_{m\geq n+1} \sigma_{m,n+1}(g_m)\Bigg) \\ &= \sum_{m\geq n} \sigma_{m,n}(g_m) - \sum_{m\geq n+1} \sigma_{m,n}(g_m) = g_n, \end{aligned}$$

which proves that $\varphi((f_n)_n) = (g_n)$. □

For any $\mathcal{A}$-module M, we set $M_n := M \otimes_{\mathcal{A}} \mathcal{A}_n$.

Definition A.2. We let Θ denote the functor

$$\Theta : \{\text{finitely generated } \mathcal{A}\text{-modules}\} \to \{\Gamma(X, \mathcal{O}_X)\text{-modules}\}, \qquad M \mapsto \varprojlim_n M_n.$$

The statement and the proof of the following result are again copied almost verbatim from [Bosch 1977] (see however Remark A.4 below).

Lemma A.3. *The functor Θ has the following properties.*

(i) *The functor Θ is exact.*

(ii) *For any finitely generated $\mathcal{A}$-module M, there is a natural isomorphism*

$$\tau : M \otimes_{\mathcal{A}} \Gamma(X, \mathcal{O}_X) \simeq \Theta(M).$$

Proof. Let us first show that Θ is exact. By the local theory of uniformly rigid spaces as developed in [Kappen 2010], the rings $\mathcal{A}_n$ are flat over $\mathcal{A}$. Hence, it suffices to show that $\varprojlim^1 M_n$ vanishes for all finitely generated $\mathcal{A}$-modules M. Thus, let M be a finitely generated $\mathcal{A}$-module, and let $F \to M \to 0$ be a finite presentation of M. Since the higher derivatives of the projective limit functor vanish, $\varprojlim^1 F_n$ maps onto $\varprojlim^1 M_n$. Since $\varprojlim^1$ commutes with finite direct sums, Lemma A.1 shows that $\varprojlim^1 F_n = 0$. The claim follows. Let us now prove the second statement. Since $\Gamma(X, \mathcal{O}_X)$ is naturally isomorphic to $\varprojlim \mathcal{A}_n$, one has a natural morphism

$$\tau : M \otimes_{\mathcal{A}} \Gamma(X, \mathcal{O}_X) \to \varprojlim M_n.$$

By definition, τ is an isomorphism when $M = \mathcal{A}$, and more generally τ is an isomorphism when M is finite and free. In general, since $\mathcal{A}$ is Noetherian, and M is finitely generated, there is an exact sequence

$$F_1 \to F_2 \to M \to 0,$$

where F_1 and F_2 are finite free $\mathcal{A}$-modules. Using exactness of Θ, one obtains the exact diagram

$$\begin{tikzcd}
{} \arrow[r] & F_2 \otimes_{\mathcal{A}} \Gamma(X, \mathcal{O}_X) \arrow[r] \arrow[d, "\wr"] & F_1 \otimes_{\mathcal{A}} \Gamma(X, \mathcal{O}_X) \arrow[r] \arrow[d, "\wr"] & M \otimes_{\mathcal{A}} \Gamma(X, \mathcal{O}_X) \arrow[r] \arrow[d, "\tau"] & 0 \\
{} \arrow[r] & \Theta(F_2) \arrow[r] & \Theta(F_1) \arrow[r] & \Theta(M) \arrow[r] & 0
\end{tikzcd}$$

and it follows that τ is an isomorphism. □

Remark A.4. The statements of Lemma A.3 do not hold for general $\mathcal{A}$-modules. Likewise, Satz 2.1 and Korollar 2.2 from [Bosch 1977] do not hold for general A-modules either, although this is not explicitly mentioned there. Indeed, in Example A.5 below, we give a counterexample to Satz 2.1 and Korollar 2.2 of

[Bosch 1977] involving modules which are not finitely generated. We want to stress that the results of [Bosch 1977, Section 2] are not affected by this observation: using the notations of [Bosch 1977], a correct replacement of [Bosch 1977, Satz 2.1 and Korollar 2.2] is to say that the functor θ is exact on the category of finitely generated $A\langle\zeta\rangle$-modules and that τ is an isomorphism for finitely generated $A\langle\zeta\rangle$-modules.

Example A.5. We use the notations of Satz 2.1 and Korollar 2.2 of [Bosch 1977], and we consider $A = k$. Moreover, we assume that $\zeta = (\zeta_1)$; that is, ζ is made of only one variable. So θ is the functor sending a $k\langle\zeta\rangle$-module M to the $k\langle\langle\zeta\rangle\rangle$-module

$$\theta(M) = \varprojlim_{n\in\mathbb{N}} M \otimes_{k\langle\zeta\rangle} k\langle\varepsilon_n^{-1}\zeta\rangle,$$

and for each M, $\tau = \tau_M$ is the natural map

$$\tau_M : M \otimes_{k\langle\zeta\rangle} k\langle\langle\zeta\rangle\rangle \to \theta(M).$$

Let us write $\mathcal{A} = k\langle\zeta\rangle$ and $\mathcal{A}_j = k\langle\varepsilon_j^{-1}\zeta\rangle$, and let us consider the $\mathcal{A}$-modules

$$M' = \bigoplus_{j\in\mathbb{N}} \mathcal{A}, \qquad M = \bigoplus_{j\in\mathbb{N}} \mathcal{A}_j, \qquad M'' = \bigoplus_{j\in\mathbb{N}} (\mathcal{A}_j/\mathcal{A}),$$

which form a natural short exact sequence of $\mathcal{A}$-modules

$$0 \to M' \to M \to M'' \to 0.$$

We denote be $(e_j)_{j\in\mathbb{N}}$ the canonical basis of M. For simplicity of notation, we also denote by $(e_j)_{j\in\mathbb{N}}$ the canonical bases of M' and M''. We claim that both the natural morphism $\tau_{M''}$ for M'' and the induced map $\theta(M) \to \theta(M'')$ are not surjective, contrary to the statements of Satz 2.1 and Korollar 2.2 of [Bosch 1977]. To this end, let us choose, for each $j \in \mathbb{N}$, a function $f_j \in \mathcal{A}$ such that f_j is invertible in $\mathcal{A}$ and such that for any $m > j$, f_j is not invertible in $\mathcal{A}_m$, by picking an element $t \in k$ as well as some positive integers a, b such that

$$\varepsilon_{j+1} \geq |t|^{a/b} > \varepsilon_j$$

and by setting $f_j := \zeta^b - t^a \in \mathcal{A} = k\langle\zeta\rangle$. Let us now consider the element $g \in \prod_n M''_n$ which is defined by giving, for each n, the element

$$g_n := \sum_{j=0}^{n-1} [f_j^{-1} \otimes 1] \cdot e_j$$

of the $\mathcal{A}_n$-module

$$M''_n = \bigoplus_{j\in\mathbb{N}} (\mathcal{A}_j/\mathcal{A}) \otimes_{\mathcal{A}} \mathcal{A}_n \cong \bigoplus_{j\in\mathbb{N}} (A_j \otimes_{\mathcal{A}} \mathcal{A}_n)/\mathcal{A}_n,$$

where $[\cdot]$ denotes the formation of the residue class and where we have used flatness of $\mathcal{A}_n$ over $\mathcal{A}$ to establish the above isomorphism. Then

$$g \in \varprojlim_{n\in\mathbb{N}} M''_n,$$

because in M''_n, we have that $[f_n^{-1}\otimes 1] = [1\otimes f_n^{-1}] = 0$. Let us now consider the natural map

$$\tau_{M''} : M'' \otimes_{\mathcal{A}} \varprojlim_n \mathcal{A}_n \to \varprojlim_n M''_n.$$

For each element h in the image of this map, there exists a j_0 such that for all $j > j_0$, the j-th component of h_n is zero for all n. On the other hand, for each n, all of the summands $[f_j^{-1}\otimes 1]\cdot e_j$ with $j < n$ defining g_n are nonzero. Indeed, if there was an element $h \in \mathcal{A}_n$ with

$$f_j^{-1}\otimes 1 = 1\otimes h \quad \text{in } \mathcal{A}_j \otimes_{\mathcal{A}} \mathcal{A}_n,$$

then the same equality would hold in the completed tensor product, which is $\mathcal{A}_n$, and f_j^{-1} would thus extend to $\mathcal{A}_n$, which is not the case. We have shown that $g \notin \operatorname{im}\tau_{M''}$ and thus established our claim that $\tau_{M''}$ is not surjective. The same statement regarding the structure of g shows our second claim, namely that g does not lie in the image of the natural map $\theta(M) \to \theta(M'')$. Indeed, it suffices to remark that

$$\begin{aligned}\theta(M) &= \varprojlim_n \Big(\bigoplus_j (\mathcal{A}_j \otimes_{\mathcal{A}} \mathcal{A}_n)\Big) = \varinjlim_i \varprojlim_n \Big(\bigoplus_{j\le i} (\mathcal{A}_j \otimes_{\mathcal{A}} \mathcal{A}_n)\Big)\\ &= \bigoplus_j \Big(\varprojlim_n (\mathcal{A}_j \otimes_{\mathcal{A}} \mathcal{A}_n)\Big),\end{aligned}$$

which follows from the fact that the transition maps $\mathcal{A}_{n+1} \to \mathcal{A}_n$ are injective and that the $\mathcal{A}_j$ are flat over $\mathcal{A}$.

Proposition A.6. *The $\mathcal{A}$-module $\Gamma(X, \mathcal{O}_X)$ is faithfully flat.*

Proof. Flatness follows from Lemma A.3. For faithfully flatness, let us consider a maximal ideal $\mathfrak{m}$ of $\mathcal{A}$. By the Nullstellensatz for semiaffinoid k-algebras, $\mathfrak{m}$ corresponds to a rigid point x of X. Since the X_n cover X, for n big enough one has $x \in X_n$. Hence $\mathfrak{m}' := \{f \in \Gamma(X, \mathcal{O}_X) \mid f(x) = 0\}$ is a maximal ideal of $\Gamma(X, \mathcal{O}_X)$ such that $\mathfrak{m}' \cap \mathcal{A} = \mathfrak{m}$. $\square$

Lemma A.7. *The following are equivalent.*

(i) *The semiaffinoid k-algebra $\mathcal{A}$ is normal.*

(ii) *The special R-algebra $\mathcal{A}^\circ$ of power-bounded functions in $\mathcal{A}$ is normal.*

Proof. Let us first remark that $\mathrm{Quot}(\mathcal{A}) = \mathrm{Quot}(\mathcal{A}^\circ)$. Let us first prove that (i) $\Rightarrow$ (ii). Let $f \in \mathrm{Quot}(\mathcal{A}^\circ)$ be an element which satisfies an equation $f^n + \sum_{i=0}^{n-1} a_i f^i = 0$ with $a_i \in \mathcal{A}^\circ$. Then $f \in \mathcal{A}$ by (i), and since the a_i are in $\mathcal{A}^\circ$, it follows that $f \in \mathcal{A}^\circ$. Let us now prove that (ii) $\Rightarrow$ (i). Let $f \in \mathrm{Quot}(\mathcal{A})$ be an element which satisfies an equation $f^n + \sum_{i=0}^{n-1} a_i f^i = 0$ with $a_i \in \mathcal{A}$. Then there exists an integer m such that for all i, $\pi^m a_i \in \mathcal{A}^\circ$. Using the same argument as in the proof of Proposition 6.1, it follows that $\pi^m f$ satisfies a unitary equation with coefficients in $\mathcal{A}^\circ$, hence $\pi^m f \in \mathcal{A}^\circ$ by (ii), hence $f \in \mathcal{A}$. $\square$

Theorem A.8. *Let $\mathcal{A}$ be a reduced semiaffinoid k-algebra, and X its associated rigid analytic k-space. Then $\mathcal{A} \simeq \{f \in \Gamma(X, \mathcal{O}_X) \mid |f|_{\sup} < \infty\}$.*

Proof. If $\mathcal{A}$ is normal, then $\mathcal{A}^\circ$ is a normal special R-algebra by Lemma A.7, and [de Jong 1995, Theorem 7.4.1] shows that $\mathcal{A}^\circ \simeq \Gamma(X, \mathcal{O}_X^\circ)$. The theorem follows in that case. In general, let $\mathcal{B}$ denote the normalization of $\mathcal{A}$. According to [Valabrega 1975; 1976], $\mathcal{B}$ is a semiaffinoid k-algebra. Let X' denote the rigid analytic k-space associated to $\mathcal{B}$ and $p : X' \to X$ the induced morphism, and let us consider the induced commutative diagram

$$\begin{tikzcd} \mathcal{A} \arrow[r] \arrow[d] & \mathcal{B} \arrow[d] \\ \Gamma(X, \mathcal{O}_X) \arrow[r, "p^*"'] & \Gamma(X', \mathcal{O}_{X'}) \end{tikzcd}$$

Let us first observe that

$$\Gamma(X', \mathcal{O}_{X'}) \simeq \Gamma(X, \mathcal{O}_X) \otimes_{\mathcal{A}} \mathcal{B}.$$

Indeed, since $\mathcal{A} \to \mathcal{B}$ is finite, for each n the morphism $\mathcal{A}_n \to \mathcal{A}_n \otimes_{\mathcal{A}} \mathcal{B}$ is also finite. Since $\mathcal{A}_n$ is a k-affinoid algebra, it follows that $\mathcal{A}_n \otimes_{\mathcal{A}} \mathcal{B}$ is also a k-affinoid algebra [Bosch et al. 1984, Proposition 6.1.1.6]. If X'_n denotes the k-affinoid space associated to $\mathcal{A}_n \otimes_{\mathcal{A}} \mathcal{B}$, then X'_n is an affinoid domain of X' and the X'_n cover X'. It follows that

$$\Gamma(X', \mathcal{O}_{X'}) \simeq \varprojlim_{n \in \mathbb{N}} \Gamma(X'_n, \mathcal{O}_{X'}) \simeq \varprojlim_{n \in \mathbb{N}} (\mathcal{A}_n \otimes_{\mathcal{A}} \mathcal{B}) \simeq \Gamma(X, \mathcal{O}_X) \otimes_{\mathcal{A}} \mathcal{B},$$

where the second equality follows from the fact that X'_n is a k-affinoid space, and the third equality follows from Lemma A.3(ii).

Let now $f \in \Gamma(X, \mathcal{O}_X)$ be a bounded function. Then $p^*(f) \in \Gamma(X', \mathcal{O}_{X'}) \simeq \Gamma(X, \mathcal{O}_X) \otimes_{\mathcal{A}} \mathcal{B}$ is also bounded on X', and according to what we have shown in the first part of the proof, $p^*(f)$ comes from an element $b \in \mathcal{B}$.

Finally, $\mathcal{A} \subset \mathcal{B}$ is a sub-$\mathcal{A}$-module of $\mathcal{B}$ because $\mathcal{A}$ is reduced. Since $\Gamma(X, \mathcal{O}_X)$ is flat over $\mathcal{A}$ (Proposition A.6), it follows that $\Gamma(X, \mathcal{O}_X)$ is a submodule of

$\Gamma(X', \mathcal{O}_{X'}) \simeq \Gamma(X, \mathcal{O}_X) \otimes_{\mathcal{A}} \mathcal{B}$. Since $\Gamma(X, \mathcal{O}_X)$ is even faithfully flat over $\mathcal{A}$, we conclude that

$$\Gamma(X, \mathcal{O}_X) \cap \mathcal{B} = \mathcal{A}.$$

Indeed, this follows from [Bourbaki 1998, Section I.3.5, Proposition 10(ii)], which asserts that if $C \to C'$ is a faithfully flat ring morphism, if N is a C-module and if $N' \subset N$ is a sub-C-module, then $(N' \otimes_C C') \cap N = N'$. Since $f \in \Gamma(X, \mathcal{O}_X) \cap \mathcal{B}$, it follows that $f \in \mathcal{A}$. $\square$

Example A.9. Theorem A.8 does not hold if we do not assume the semiaffinoid k-algebra $\mathcal{A}$ to be reduced. For instance, as in Example 2.3, let

$$\mathcal{A} := R[\![T_1, T_2]\!]/(T_2^2) \otimes_R k,$$

and let X be the associated rigid k-space. Let $f(T_1) \in k[\![T_1]\!]$ be a formal power series which converges on the open unit disc, but such that $f(T_1) \notin R[\![T_1]\!] \otimes_R k$. Then $f(T_1)T_2 \in \Gamma(X, \mathcal{O}_X^\circ) \setminus \mathcal{A}$.

Acknowledgements

We thank the referee for a very careful reading leading to a great improvement of this text. The first author would like to thank Ofer Gabber for pointing out a result of [EGA IV$_2$ 1965] which greatly simplified the proof of the main result. He also would like to thank Antoine Ducros, Jérôme Poineau, Andrea Pulita, Matthieu Romagny and Michael Temkin for useful comments and conversations about this work. The first author acknowledges support from SFB 1085 Higher invariants funded by the Deutsche Forschungsgesellschaft (DFG).

References

[Berkovich 1990] V. G. Berkovich, *Spectral theory and analytic geometry over non-Archimedean fields*, Mathematical Surveys and Monographs **33**, American Mathematical Society, Providence, RI, 1990. MR Zbl

[Berkovich 1993] V. G. Berkovich, "Étale cohomology for non-Archimedean analytic spaces", *Inst. Hautes Études Sci. Publ. Math.* 78 (1993), 5–161. MR Zbl

[Berkovich 1996] V. G. Berkovich, "Vanishing cycles for formal schemes, II", *Invent. Math.* **125**:2 (1996), 367–390. MR Zbl

[Berthelot 1996] P. Berthelot, "Cohomologie rigide et cohomologie rigide à supports propres, première partie", preprint, Institut de recherche mathématique de Rennes, 1996, available at https://perso.univ-rennes1.fr/pierre.berthelot/publis/Cohomologie_Rigide_I.pdf.

[Bosch 1977] S. Bosch, "Eine bemerkenswerte Eigenschaft der formellen Fasern affinoider Räume", *Math. Ann.* **229**:1 (1977), 25–45. MR Zbl

[Bosch and Lütkebohmert 1985] S. Bosch and W. Lütkebohmert, "Stable reduction and uniformization of abelian varieties, I", *Math. Ann.* **270**:3 (1985), 349–379. MR Zbl

[Bosch et al. 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis: A systematic approach to rigid analytic geometry*, Grundlehren der Mathematischen Wissenschaften **261**, Springer, Berlin, 1984. MR Zbl

[Bourbaki 1998] N. Bourbaki, *Commutative algebra, Chapters 1–7*, Springer, Berlin, 1998. MR Zbl

[Ducros 2003] A. Ducros, "Parties semi-algébriques d'une variété algébrique p-adique", *Manuscripta Math.* **111**:4 (2003), 513–528. MR Zbl

[EGA I 1960] A. Grothendieck, "Eléments de géométrie algébrique, I: Le langage des schémas", *Inst. Hautes Études Sci. Publ. Math.* **4** (1960), 5–228. MR Zbl

[EGA IV$_2$ 1965] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, II", *Inst. Hautes Études Sci. Publ. Math.* **24** (1965), 5–231. MR Zbl

[de Jong 1995] A. J. de Jong, "Crystalline Dieudonné module theory via formal and rigid geometry", *Inst. Hautes Études Sci. Publ. Math.* 82 (1995), 5–96. MR Zbl

[Kappen 2010] C. Kappen, *Uniformly rigid spaces and Néron models of formally finite type*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster, 2010, available at http://nbn-resolving.de/urn:nbn:de:hbz:6-78489606551. Zbl

[Kappen 2012] C. Kappen, "Uniformly rigid spaces", *Algebra Number Theory* **6**:2 (2012), 341–388. MR Zbl

[Kiehl 1967] R. Kiehl, "Theorem A und Theorem B in der nichtarchimedischen Funktionentheorie", *Invent. Math.* **2** (1967), 256–273. MR Zbl

[Lipshitz and Robinson 2000] L. Lipshitz and Z. Robinson, "Rings of separated power series", pp. 3–108 in *Rings of separated power series and quasi-affinoid geometry*, Astérisque **264**, Société Mathématique de France, Paris, 2000. MR Zbl

[Matsumura 1989] H. Matsumura, *Commutative ring theory*, 2nd ed., Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press, 1989. MR Zbl

[Nicaise 2009] J. Nicaise, "A trace formula for rigid varieties, and motivic Weil generating series for formal schemes", *Math. Ann.* **343**:2 (2009), 285–349. MR Zbl

[Poineau 2008] J. Poineau, "Un résultat de connexité pour les variétés analytiques p-adiques: privilège et noethérianité", *Compos. Math.* **144**:1 (2008), 107–133. MR Zbl

[Poineau 2014] J. Poineau, "Sur les composantes connexes d'une famille d'espaces analytiques p-adiques", *Forum Math. Sigma* **2** (2014), e14, 21. MR Zbl

[Temkin 2004] M. Temkin, "On local properties of non-Archimedean analytic spaces, II", *Israel J. Math.* **140** (2004), 1–27. MR Zbl

[Valabrega 1975] P. Valabrega, "On the excellent property for power series rings over polynomial rings", *J. Math. Kyoto Univ.* **15**:2 (1975), 387–395. MR Zbl

[Valabrega 1976] P. Valabrega, "A few theorems on completion of excellent rings", *Nagoya Math. J.* **61** (1976), 127–133. MR Zbl

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics **38**, Cambridge University Press, 1994. MR Zbl

Communicated by Brian Conrad
Received 2016-05-11 Revised 2016-12-14 Accepted 2017-01-11

florent.martin@mathematik.uni-regensburg.de
Fakultät für Mathematik, Universität Regensburg, D-93040 Regensburg, Germany

christian.kappen@uni-due.de *Universität Duisburg-Essen, Institut für Experimentelle Mathematik, Ellernstraße 29, D-45326 Essen, Germany*

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.685

Automatic sequences and curves over finite fields

Andrew Bridy

We prove that if $y=\sum_{n=0}^{\infty}\boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$ is an algebraic power series of degree d, height h, and genus g, then the sequence $\boldsymbol{a}$ is generated by an automaton with at most $q^{h+d+g-1}$ states, up to a vanishingly small error term. This is a significant improvement on previously known bounds. Our approach follows an idea of David Speyer to connect automata theory with algebraic geometry by representing the transitions in an automaton as twisted Cartier operators on the differentials of a curve.

1. Introduction

Our starting point is the following well-known theorem of finite automata theory.

Theorem 1.1 [Christol 1979; Christol et al. 1980]. *The power series*

$$y=\sum_{n=0}^{\infty}\boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$$

is algebraic over $\mathbb{F}_q(x)$ if and only if the sequence $\boldsymbol{a}$ is q-automatic.

Christol's theorem establishes a dictionary between automatic sequences and number theory in positive characteristic. The purpose of this paper is to investigate how complexity translates across this dictionary. A secondary purpose is to demonstrate an intimate connection between automatic sequences and the algebraic geometry of curves.

We take the complexity of a sequence $\boldsymbol{a}$ to be *state complexity*. Let $N_q(\boldsymbol{a})$ denote the number of states in a minimal q-automaton that generates $\boldsymbol{a}$ in the reverse-reading convention (this will be defined precisely in Section 2). If y is algebraic over $k(x)$, let the *degree* $\deg y$ be the usual field degree $[k(x)[y]:k(x)]$ and the *height* $h(y)$ be the minimal x-degree of a bivariate polynomial $f(x,T)\in k[x,T]$ such that $f(x,y)=0$. The *genus* of y will be the genus of the normalization of the projective closure of the affine plane curve defined by the minimal polynomial of y.

MSC2010: primary 11B85; secondary 11G20, 14H05, 14H25.
Keywords: automatic sequences, formal power series, algebraic curves, finite fields.

We bound the complexity of $\boldsymbol{a}$ in terms of the degree, height, and genus of y. In Section 2 we review some known lower bounds. Our main result is the following upper bound.

Theorem 1.2. *Let $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$ be algebraic over $\mathbb{F}_q(x)$ of degree d, height h, and genus g. Then*

$$N_q(\boldsymbol{a}) \le (1+o(1))q^{h+d+g-1}.$$

The $o(1)$ term tends to 0 for large values of q, h, d, or g.

All previous upper bounds are much larger. These are usually stated in terms of the q-kernel of $\boldsymbol{a}$, which can be described as the orbit of a certain semigroup acting on the series y and is in bijection with a minimal automaton that outputs $\boldsymbol{a}$ (see Theorem 2.2). The best previous bound is due to Fresnel, Koskas, and de Mathan [2000, Theorem 2.2] who show that

$$N_q(\boldsymbol{a}) \le q^{qd(h(2d^2-2d+1)+C)}, \tag{1.3}$$

for some $C = C(q)$ that they do not seem to compute exactly. Adamczewski and Bell [2012, p. 383] prove a bound that is roughly $q^{d^4h^2p^{5d}}$, where $p = \operatorname{char}\mathbb{F}_q$. Earlier, Derksen [2007, Proposition 6.5] showed a special case of the bound of Adamczewski and Bell for rational functions. Harase [1988; 1989] proved a larger bound using essentially the same technique. It should be noted that some of these results hold in more generality than the setting of this paper; the techniques of Adamczewski, Bell, and Derksen apply to power series in several variables over infinite ground fields of positive characteristic (with appropriate modifications).

In Proposition 3.14 we show that Theorem 1.2 is qualitatively sharp for the power series expansions of rational functions, in that it is sharp if we replace the $o(1)$ term by 0. There are some special cases where the bound can be improved. For example, an easy variation of our main argument shows that if $y\,dx$ is a holomorphic differential on the curve defined by the minimal polynomial of y, then $N_q(\boldsymbol{a}) \le q^{d+g-1}$ (see Example 4.3). It is also possible to give a coarser estimate that is independent of g, which shows that Theorem 1.2 compares favorably to the work of Fresnel et al., even when the genus is as large as possible:

Corollary 1.4. *Under the hypotheses of Theorem 1.2, $N_q(\boldsymbol{a}) \le (1+o(1))q^{hd}$.*

Proof. Let X be the curve defined by the minimal polynomial of y. Observe that d is the degree of the map $\pi_x : X \to \mathbb{P}^1$ that projects on the x-coordinate. Likewise, h is the degree of the projection map $\pi_y : X \to \mathbb{P}^1$. Therefore $g \le (d-1)(h-1)$ by Castelnuovo's inequality (actually, a special case originally due to Riemann, see [Stichtenoth 2009, Corollary 3.11.4]). $\square$

Though the reverse-reading convention is the natural one to use in the context of algebraic series, our approach also gives an upper bound on the state complexity of forward-reading automata via a dualizing argument. Let $N_q^f(\boldsymbol{a})$ denote the minimal number of states in a forward-reading automaton that generates the sequence $\boldsymbol{a}$, and let y be as in Theorem 1.2.

Theorem 1.5. $$N_q^f(\boldsymbol{a}) \le q^{h+2d+g-1}.$$

To our knowledge, there are no previous bounds on forward-reading complexity in this context except for the well-known observation that $N_q^f(\boldsymbol{a}) \le q^{N_q(\boldsymbol{a})}$ (see Proposition 2.4).

The key idea in our argument is to recast finite automata in the setting of algebraic geometry. If $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n$ is algebraic, then it lies in the function field of a curve X, and a minimal reverse-reading automaton that generates $\boldsymbol{a}$ embeds into the differentials of X in a natural way. This brings to bear the machinery of algebraic curves, and in particular the Riemann–Roch theorem. This idea was introduced by David Speyer [2010] and used to give a new proof of the "algebraic implies automatic" direction of Christol's theorem. Building on Speyer's work, we improve the complexity bound implicit in his proof.

The paper is organized as follows. Section 2 reviews the theory of finite automata and automatic sequences. Section 3 introduces the connection with algebraic geometry, leading to the proof of Theorem 1.2. We also discuss the problem of state complexity growth as the field varies: if K is a number field and $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in K[\![x]\!]$ is algebraic over $K(x)$, then the state complexity of the reduced sequences $\boldsymbol{a}_{\mathfrak{p}}$ varies with the prime $\mathfrak{p}$ in a way controlled by the algebraic nature of y. Section 4 illustrates some detailed examples of our method.

2. Automata, sequences, and representations

2A. *Finite automata and automatic sequences.* A comprehensive introduction to finite automata and automatic sequences can be found in the book of Allouche and Shallit [2003]. We give a brief overview of the theory for the convenience of the reader.

A *finite automaton* or DFAO (*deterministic finite automaton with output*) M consists of a finite set Σ known as the *input alphabet*, a finite set Δ known as the *output alphabet*, a finite set of *states* Q, a distinguished *initial state* $q_0 \in Q$, a *transition function* $\delta : Q \times \Sigma \to Q$, and an *output function* $\tau : Q \to \Delta$.

Let Σ^* (the *Kleene closure* of Σ) be the monoid of all finite-length words over Σ under the operation of juxtaposition, including an empty word as the identity element. The function δ can be prolonged to a function $\delta : Q \times \Sigma^* \to Q$ by inductively defining $\delta(q_i, wa) = \delta(\delta(q_i, w), a)$ for $w \in \Sigma^*$ and $a \in \Sigma$. Therefore a DFAO M induces a map $f_M : \Sigma^* \to \Delta$ defined by $f_M(w) = \tau(\delta(q_0, w))$ under the

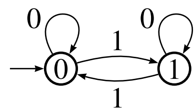


Figure 1. Thue–Morse 2-DFAO T.

forward-reading convention. If we let w^R denote the reverse of the word w, then the *reverse-reading* convention is $f_M(w) = \tau(\delta(q_0, w^R))$. A function $f : \Sigma^* \to \Delta$ is a *finite-state function* if $f = f_M$ for some DFAO M. A DFAO is *minimal* if it has the smallest number of states among automata that induce the same function f_M.

A helpful way of visualizing a DFAO M is through its *transition diagram*. This is a directed graph with vertex set Q and directed edges that join q to $\delta(q, a)$ for each $q \in Q$ and $a \in \Sigma$. The initial state is marked by an incoming arrow with no source. The states are labeled by their output $\tau(q)$. Figure 1 shows the transition diagram of the Thue–Morse DFAO T with $\Sigma = \Delta = \{0, 1\}$, where $f_T(w) = 1$ if and only if $w \in \{0, 1\}^*$ contains an odd number of 1s.

Let $p \geq 2$ be an integer (not necessarily prime) and let $(n)_p$ denote the base-p expansion of the integer $n \geq 0$. A sequence $\boldsymbol{a}$ is *p-automatic* if there exists a DFAO M with input alphabet $\Sigma_p = \{0, 1, \ldots, p-1\}$ such that $\boldsymbol{a}(n) = f_M((n)_p)$; we say that M *generates* $\boldsymbol{a}$. It is known that a sequence is p-automatic with respect to the forward-reading convention if and only if it is p-automatic with respect to the reverse-reading convention [Allouche and Shallit 2003, Theorem 5.2.3]. (We prove a quantitative version of this fact in Proposition 2.4.) A DFAO with input alphabet Σ_p is called a p-DFAO.

Let $\boldsymbol{a}$ be a p-automatic sequence. As in the introduction, the *forward-reading complexity* $N_p^f(\boldsymbol{a})$ is the number of states in a minimal forward-reading p-DFAO that generates $\boldsymbol{a}$, and the *reverse-reading complexity* $N_p(\boldsymbol{a})$ is the number of states in a minimal reverse-reading p-DFAO that generates $\boldsymbol{a}$.

Remark 2.1. Base-p expansions are only unique if we disallow leading zeros — for example, the binary strings 11 and 011 both represent the integer 3. This creates a minor ambiguity in the minimality of a generating DFAO for a sequence, as it may be the case that a larger DFAO is needed if we require that the same output is produced for *every possible* base-p expansion of an integer (it is not even *a priori* clear that both definitions of "p-automatic" are equivalent; see [Allouche and Shallit 2003, Theorem 5.2.1]). Throughout this paper we enforce the stricter requirement that the generating DFAO gives the same output regardless of leading zeros. (This is necessary for minimality to translate correctly from automata to curves.)

The canonical example of an automatic sequence is the 2-automatic Thue–Morse sequence

$$\boldsymbol{a} = 01101001\ldots,$$

where $\boldsymbol{a}(n) = f_T((n)_2)$ for the Thue–Morse automaton T. The term $\boldsymbol{a}(n)$ is the parity of the sum of the bits in the binary expansion of n. Note that T generates $\boldsymbol{a}$ under both the forward-reading and reverse-reading conventions, and T is obviously minimal, so $N_2(\boldsymbol{a}) = N_2^f(\boldsymbol{a}) = 2$.

There are many characterizations of a sequence that are equivalent to being p-automatic. We mention two, which are relevant to computing state complexity. The first is due to Eilenberg and relies on the notion of the p-kernel of $\boldsymbol{a}$, which is defined to be the set of sequences $n \mapsto \boldsymbol{a}(p^i n + j)$ for all $i \geq 0$ and $0 \leq j \leq p^i - 1$. The second actually holds in more generality than automatic sequences: it is an easy adaptation of the Myhill–Nerode theorem to the DFAO model.

Theorem 2.2 (Eilenberg). *For $p \geq 2$, the p-kernel of $\boldsymbol{a}$ is finite if and only if $\boldsymbol{a}$ is p-automatic. Moreover, $N_p(\boldsymbol{a})$ is precisely the size of the p-kernel of $\boldsymbol{a}$.*

Proof. See [Eilenberg 1974, Proposition V.3.3] or [Allouche and Shallit 2003, Proposition 6.6.2]. See [Derksen 2007, Proposition 4.9] for the claim of minimality (this minimality is in the strict sense of Remark 2.1, as a smaller DFAO may exist otherwise). □

Let $f : \Sigma^* \to \Delta$ be any function. For $x, y \in \Sigma^*$, define $x \sim y$ to mean $f(xz) = f(yz)$ for all $z \in \Sigma^*$. Then $\sim$ is an equivalence relation on Σ^*, called the Myhill–Nerode equivalence relation.

Theorem 2.3 (Myhill–Nerode). *The equivalence relation $\sim$ has finitely many equivalence classes if and only if f is a finite-state function. The number of equivalence classes of $\sim$ is the minimal number of states in a forward-reading* DFAO *M such that $f = f_M$.*

Proof. See [Allouche and Shallit 2003, Theorem 4.1.8 and p. 149]. □

2B. ***Automata from p-representations.*** Another way of characterizing p-automatic sequences is by p-representations. Let $\boldsymbol{a}$ be a sequence taking values in a field k. A *p-representation* of $\boldsymbol{a}$ consists of a finite-dimensional vector space V over k, a vector $v \in V$, a morphism of monoids $\phi : \Sigma_p^* \to \operatorname{End}(V)$, and a linear functional $\lambda \in V^* = \operatorname{Hom}(V, k)$, such that for any positive integer n,

$$\boldsymbol{a}(n) = \lambda\phi((n)_p)v.$$

A sequence that admits a p-representation is known as *p-regular* [Allouche and Shallit 2003, Chapter 16]. Equivalently, its associated power series is *recognizable* in the language of [Berstel and Reutenauer 1988] (see also [Salomaa and Soittola 1978]).

We also make the nonstandard but natural definition of a *p-antirepresentation* of $\boldsymbol{a}$, which consists of the data of a representation except that $\phi : \Sigma_p^* \to \operatorname{End}(V)$ is an antimorphism of monoids, that is, $\phi(wv) = \phi(v)\phi(w)$. Antirepresentations on V correspond to representations on the dual space V^*. In Proposition 2.4 we show that when k is finite of characteristic p, p-representations give rise to reverse-reading automata and p-antirepresentations give rise to forward-reading automata.

An obvious necessary condition for a sequence to be automatic is that it assumes finitely many values. It is not hard to show that a sequence over a field is p-automatic if and only if it is both p-regular and assumes finitely many values (see [Allouche and Shallit 2003, Theorem 16.1.5] or [Berstel and Reutenauer 1988, Theorem V.2.2]). We give a quantitative proof of this fact when k is a finite field, in which case the "finitely many values" hypothesis holds trivially. Our argument will allow us to deduce a bound on state complexity.

Proposition 2.4. *Let p be a prime or prime power and let $k = \mathbb{F}_p$. The sequence $\boldsymbol{a}$ over k is p-regular if and only if it is p-automatic. Furthermore, if $\boldsymbol{a}$ has a p-representation on a vector space V of dimension m, then $N_p^f(\boldsymbol{a})$ and $N_p(\boldsymbol{a})$ are both at most p^m.*

Proof. First assume that $\boldsymbol{a}$ is p-automatic. There exists a reverse-reading p-DFAO M such that $f_M((n)_p^R) = \boldsymbol{a}(n)$. We construct a representation for $\boldsymbol{a}$ analogous to the regular representation in group theory. Let $q_1, \dots, q_m$ be the states of M and let $V = k^m$. Let $v = e_1$, the first standard basis vector of V. For $i \in \Sigma_p$, define the matrix $\phi(i) \in k^{m\times m} \simeq \mathrm{End}(V)$ by

$$\phi(i)_{a,b} = \begin{cases} 1 & \text{if } \delta_i(q_b) = q_a, \\ 0 & \text{otherwise,} \end{cases}$$

and extend ϕ to a morphism from Σ_p^* to $k^{m\times m}$. Let λ be defined by $\lambda(e_j) = \tau(q_j)$ for each j and extended linearly to a functional $\lambda : V \to k$. This defines a p-representation of $\boldsymbol{a}$, which can be pictured as embedding the states of M into V and realizing the transition function as a set of p linear transformations.

Now assume instead that $\boldsymbol{a}$ is p-regular. Let (V, v, ϕ, λ) be a p-representation of $\boldsymbol{a}$ with $\dim V = m$. We construct a reverse-reading DFAO M as follows. The initial state q_0 is v, the set of states is $Q = \{\phi(w)v : w \in \Sigma_p^*\}$, the transition function is given by $\delta(w, i) = \phi(i)(w)$, and the output function is $\tau(w) = \lambda(w)$. It is a matter of unraveling notation to see that M outputs the sequence $\boldsymbol{a}$, and M has at most $|V| = p^m$ states.

We now construct a p-antirepresentation for the sequence $\boldsymbol{a}$. Let $\phi^T : \Sigma_p^* \to \mathrm{End}(V^*)$ be the antimorphism defined by $\phi^T(w) = \phi(w)^T$, where T denotes transpose. Now $(V^*, \lambda, \phi^T, v)$ is a p-antirepresentation of $\boldsymbol{a}$, where we identify V with $(V^*)^*$ in the natural way. If $(n)_p = c_u \cdots c_1 c_0$, then

$$\boldsymbol{a}(n) = \lambda\phi((n)_p)v = \lambda\phi(c_u)\cdots\phi(c_0)v,$$

so thinking of v as an element of $(V^*)^*$, we have

$$\boldsymbol{a}(n) = v(\lambda\phi(c_u)\cdots\phi(c_0)) = v\phi^T(c_0)\cdots\phi^T(c_u)\lambda = v\phi^T(c_u\cdots c_0)\lambda = v\phi^T((n)_p)\lambda.$$

This corresponds to a forward-reading DFAO M in the following way: let the initial state q_0 of M be λ, the set of states be $Q = \{\phi^T(w)\lambda : w \in \Sigma_p^*\}$, the transition function be $\delta(\mu, i) = \phi^T(i)(\mu)$, and the output function be $\tau(\mu) = v^T(\mu) = \mu(v)$. Taking transposes has the effect of reversing input words, so this gives a forward-reading DFAO that outputs the sequence $\boldsymbol{a}$, and it has at most p^m states, as $\dim V = \dim V^*$. □

Remark 2.5. A special case of the antirepresentation constructed in Proposition 2.4 gives a standard result of automata theory; if $\Delta = \{0, 1\}$, so that M either accepts or rejects each input string, and M has n states, then a minimal reversed automaton for M has at most 2^n states. We can identify Δ with $\mathbb{F}_2$, and the regular representation is on the vector space $\mathbb{F}_2^n$.

The representation constructed in Proposition 2.4 produces a p-DFAO where the states are identified with a subset of V and the transitions are realized as linear transformations. In general, this is not a minimal DFAO. Much of our work in the rest of the paper will be describing canonical representations that produce minimal DFAO.

Somewhat surprisingly, for any p-representation of $\boldsymbol{a}$, the forward-reading p-DFAO antirepresentation in Proposition 2.4 is minimal as long as V equals the linear span of $\{\phi(w)v : w \in \Sigma_p^*\}$. This assumption on V loses no generality, because we can always replace V with this subspace (in particular, satisfying this assumption does *not* mean the corresponding reverse-reading automaton is minimal). This observation is to our knowledge new, and we prove it in Proposition 2.6. In a sense, this is an analogue via representations of the minimization algorithm of Brzozowski [Brzozowski 1963; Shallit 2009].

Proposition 2.6. *Let $\boldsymbol{a}$ be a sequence taking values in a finite field k, and let (V, v, ϕ, λ) be a p-representation of $\boldsymbol{a}$. Assume without loss of generality that V is the k-linear span of $\{\phi(w)v : w \in \Sigma_p^*\}$. The* DFAO *$M$ corresponding to the antirepresentation in Proposition 2.4 is a minimal forward-reading p-*DFAO *that generates $\boldsymbol{a}$.*

Proof. The state set of M is $Q = \{\phi^T(w)\lambda : w \in \Sigma_p^*\}$ with initial state λ, the transition function is $\delta(\mu, i) = \phi(i)^T(\mu)$, and the output function is $\tau(\mu) = \mu(v)$ for some fixed $v \in V$. We show that the states of Q are in one-to-one correspondence with the Myhill–Nerode equivalence classes of the finite-state function f_M as in Theorem 2.3.

Let $[x]$ be the equivalence class of $x \in \Sigma_p^*$. We need to show that $[x] = [y]$ if and only if $\phi^T(x)(\lambda) = \phi^T(y)(\lambda)$. We have

$$[x] = \left\{y \in \Sigma_p^* : \tau(\phi^T(xz)\lambda) = \tau(\phi^T(yz)\lambda) \text{ for all } z \in \Sigma_p^*\right\},$$

and the computation $\tau(\phi^T(xz)\lambda) = \tau(\lambda\phi(xz)) = \lambda\phi(x)\phi(z)v$ shows that

$$\begin{aligned}[x] &= \left\{ y \in \Sigma_p^* : \lambda\phi(x)\phi(z)v = \lambda\phi(y)\phi(z)v \text{ for all } z \in \Sigma_p^* \right\} \\ &= \left\{ y \in \Sigma_p^* : \lambda\phi(x) = \lambda\phi(y) \text{ in } V^* \right\},\end{aligned}$$

because V is the span of the set $\{\phi(z)v : z \in \Sigma_p^*\}$. So $[x]$ is the precisely the set of all y such that $\phi^T(x)\lambda = \phi^T(y)\lambda$. By the Myhill–Nerode theorem, M is a minimal forward-reading DFAO that generates $\boldsymbol{a}$. □

2C. ***Power series and bounds on degree and height.*** We develop some standard machinery that is used in the proof of Christol's theorem. Let k be a perfect field of characteristic p, for example, a finite field $\mathbb{F}_{p^r}$. Let

$$y = \sum_{n=-\infty}^{\infty} \boldsymbol{a}(n)x^n \in k((x)),$$

where $\boldsymbol{a}(n) = 0$ for all sufficiently large negative n. Define

$$\Lambda_i(y) = \sum_{n=-\infty}^{\infty} \boldsymbol{a}(pn+i)^{1/p}x^n. \tag{2.7}$$

The operators Λ_i are $\mathbb{F}_p$-linear (not necessarily k-linear) endomorphisms of the field $k((x))$. They are known in this context as Cartier operators. Observe that

$$\begin{aligned} y &= \sum_{i=0}^{p-1}\sum_{n=-\infty}^{\infty} \boldsymbol{a}(pn+i)x^{pn+i} \\ &= \sum_{i=0}^{p-1} x^i \sum_{n=-\infty}^{\infty} \boldsymbol{a}(pn+i)x^{pn} \\ &= \sum_{i=0}^{p-1} x^i \left(\sum_{n=-\infty}^{\infty} \boldsymbol{a}(pn+i)^{1/p}x^n \right)^p \end{aligned} \tag{2.8}$$

and therefore

$$y = \sum_{i=0}^{p-1} x^i (\Lambda_i(y))^p. \tag{2.9}$$

If $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_p[[x]]$, it is easy to see that the p-kernel of $\boldsymbol{a}$ is in bijection with the orbit of y under the monoid generated by the Λ_i operators, as taking p-th roots fixes each element of $\mathbb{F}_p$. If $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_q[[x]]$ for $q = p^r$, then applying r-fold compositions of the Λ_i operators gives q-ary decimations of the sequence $\boldsymbol{a}$. That is, if

$$c = i_r p^{r-1} + i_{r-1}p^{r-2} + \cdots + i_2 p + i_1,$$

where each $i_j \in \{0, \cdots, p-1\}$, then

$$\Lambda_{i_1}\Lambda_{i_2}\cdots\Lambda_{i_r}(y) = \sum_{n=0}^{\infty} \boldsymbol{a}(p^r n + i_r p^{r-1} + \cdots + i_2 p + i_1)^{1/p^r} x^n$$
$$= \sum_{n=0}^{\infty} \boldsymbol{a}(qn+c)x^n, \tag{2.10}$$

because $\mathbb{F}_q$ is fixed under taking q-th roots. It follows that

$$y = \sum_{0 \le i_1, \ldots, i_r \le p-1} x^{i_r p^{r-1} + \cdots + i_2 p + i_1} (\Lambda_{i_1}\Lambda_{i_2}\cdots\Lambda_{i_r}(y))^q. \tag{2.11}$$

Remark 2.12. The operators Λ_i are usually defined by $\Lambda_i(y) = \sum_{n=-\infty}^{\infty} \boldsymbol{a}(qn+i)x^n$ when $k = \mathbb{F}_q$; see for example [Adamczewski and Bell 2012; 2013; Allouche and Shallit 2003]. With this definition, (2.11) takes on the much simpler form

$$y = \sum_{i=0}^{q-1} x^i (\Lambda_i(y))^q.$$

However, our definition fits more naturally into the geometric setting of Section 3 because it is invariant under base extension, whereas the usual definition depends on a choice of $\mathbb{F}_q$ fixed in advance.

Continue to assume that $y \in \mathbb{F}_q[\![x]\!]$, where p is prime and $q = p^r$. Define S_q to be the monoid generated by all r-fold compositions of the Λ_i operators. The q-kernel of $\boldsymbol{a}$ is in bijection with the orbit of y under S_q, which we denote $S_q(y)$. If $\boldsymbol{a}$ is a q-automatic sequence, then by Eilenberg's Theorem $S_q(y)$ is finite and $|S_q(y)| = N_q(\boldsymbol{a})$. Moreover, we have a q-representation for $\boldsymbol{a}$: V is the finite-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_q[\![x]\!]$ spanned by the power series whose coefficient sequences are in the q-kernel of $\boldsymbol{a}$, ϕ is defined so that, for $i \in \Sigma_q$, $\phi(i)$ maps $\sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n$ to $\sum_{n=0}^{\infty} \boldsymbol{a}(qn+i)x^n$ by the r-fold composition of the Λ_i operators given in (2.11), and the linear functional λ maps a power series to its constant term.

The standard proof of the "algebraic implies automatic" half of Christol's theorem [Christol et al. 1980; Christol 1979; Allouche and Shallit 2003, Theorem 12.2.5] follows from the observation that y is algebraic if and only if it lies in a finite-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_q((x))$ invariant under S_q. Given an algebraic power series y, it is easy to construct an invariant space using Ore's lemma [Allouche and Shallit 2003, pp. 355–356], which leads to the prior bounds on state complexity mentioned in the introduction, but the dimension of the space constructed is often far larger than the dimension of the linear span of $S_q(y)$. We achieve a sharper bound on the dimension by introducing some relevant machinery from algebraic geometry in the next section. First we demonstrate some easy upper bounds on height and degree in terms of reverse-reading state complexity that can be extracted from the usual proof of Christol's theorem.

Proposition 2.13. *Let* $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$. *Assume* $\boldsymbol{a}$ *is* q*-automatic and* $N_q(\boldsymbol{a}) = m$. *Then* y *is algebraic*, $\deg y \le q^m - 1$, *and* $h(y) \le mq^{m+1}$.

Proof. Let $S_q(y) = \{y_1, \dots, y_m\}$. From (2.11), for each $i \in \{1, \dots, m\}$ we have

$$y_i \in \langle y_1^q, \dots, y_m^q \rangle,$$

where the angle brackets $\langle \cdots \rangle$ indicate $\mathbb{F}_q(x)$-linear span. So

$$y_i^q \in \langle y_1^{q^2}, \dots, y_m^{q^2} \rangle,$$

and eventually

$$y_i^{q^m} \in \langle y_1^{q^{m+1}}, \dots, y_m^{q^{m+1}} \rangle.$$

Therefore

$$\{y_i, y_i^q, y_i^{q^2}, \dots, y_i^{q^m}\} \subseteq \langle y_1^{q^{m+1}}, \dots, y_m^{q^{m+1}} \rangle,$$

which forces an $\mathbb{F}_q(x)$-linear relation among $y_i, y_i^q, y_i^{q^2}, \dots, y_i^{q^m}$, that is, an algebraic equation satisfied by y_i. In particular, y is algebraic, which proves the "automatic implies algebraic" direction of Christol's theorem. If $y \neq 0$ we can cancel y to deduce $\deg y \le q^m - 1$ (if $y = 0$ this is trivially true).

Working through the chain of linear dependences shows that each $y_i^{q^k}$ can be written as a linear combination of $\{y_1^{q^{m+1}}, \dots, y_m^{q^{m+1}}\}$ with polynomial coefficients of degree at most q^{m+1}. A standard argument in linear algebra shows that there is a vanishing linear combination of $\{y_i, y_i^q, \dots, y_i^{q^m}\}$ with polynomial coefficients of degree at most mq^{m+1}. □

It is easy to construct infinite families of power series for which degree and height grow exponentially in $N_q(\boldsymbol{a})$, which we do in Examples 2.14 and 2.15. It is not clear whether the bounds of Proposition 2.13 are sharper than these families indicate.

Example 2.14. Let $y = x^n$ and let $\boldsymbol{a}$ be the sequence with a 1 in the n-th position and 0 in every other position. We have $N_q(\boldsymbol{a}) = \lceil \log_q(n) \rceil + 1$, because a q-DFAO generating $\boldsymbol{a}$ needs $\lceil \log_q(n) \rceil$ states to recognize the base-q expansion of n and one additional "trap state" that outputs zero on any input that deviates from this expansion. So $h(y)$ grows exponentially in the number of states required. For example, Figure 2 gives a 3-DFAO that outputs 1 on the word 201 and 0 otherwise.

Example 2.15. The degree bound of Proposition 2.13 is nearly sharp for those degrees that are powers of q. We argue that the unique solution in $\mathbb{F}_q[\![x]\!]$ to the Artin–Schreier equation

$$y^{q^m} - y = x,$$

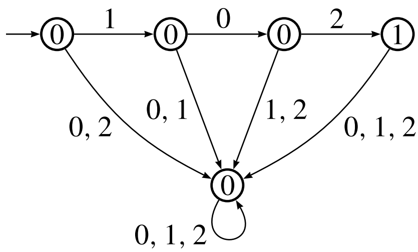


Figure 2. Minimal 3-DFAO that outputs 1 on 201 and 0 otherwise.

which is

$$y = x + x^{q^m} + x^{q^{2m}} + x^{q^{3m}} + \cdots ,$$

satisfies $N_q(\boldsymbol{a}) = m + 2$, so $\deg y = q^{N_q(\boldsymbol{a})-2}$. As usual, we identify the states of a reverse-reading q-automaton that outputs $\boldsymbol{a}$ with the orbit of y under S_q. We compute

$$\phi(0)(y) = \sum_{n=0}^{\infty} \boldsymbol{a}(qn)x^n = x^{q^{m-1}} + x^{q^{2m-1}} + x^{q^{3m-1}} + \cdots ,$$

$$\phi(1)(y) = \sum_{n=0}^{\infty} \boldsymbol{a}(qn+1)x^n = 1,$$

$$\phi(c)(y) = \sum_{n=0}^{\infty} \boldsymbol{a}(qn+c)x^n = 0,$$

for $2 \le c \le q-1$. It is clear that $S_q(1) = \{0, 1\}$, and

$$\begin{aligned}
\phi(0)^2(y) &= x^{q^{m-2}} + x^{q^{2m-2}} + x^{q^{3m-2}} + \cdots , \\
\phi(0)^3(y) &= x^{q^{m-3}} + x^{q^{2m-3}} + x^{q^{3m-3}} + \cdots , \\
&\vdots \\
\phi(0)^{m-1}(y) &= x^q + x^{q^{m+1}} + x^{q^{2m+1}} + \cdots , \\
\phi(0)^m(y) &= x + x^{q^m} + x^{q^{2m}} + \cdots = y.
\end{aligned}$$

Except for y, the power series in this list are all q-th powers, so any element of S_q that includes a Λ_i operator other than Λ_0 sends each one to zero. By Eilenberg's theorem, a minimal reverse-reading automaton that outputs $\boldsymbol{a}$ has $m+2$ states. Figure 3 depicts such an automaton for $m = 4$. Any undrawn transition arrow leads to a trap state q_T (not pictured) where $\delta(q_T, i) = q_T$ for every $i \in \Sigma_q$ and $\tau(q_T) = 0$.

A sequence is p-automatic if and only if it is p^r-automatic for any $r \ge 1$ [Allouche and Shallit 2003, Theorem 6.6.4], so it makes sense to discuss the base-p state complexity of $\boldsymbol{a}$, as well as the base-q state complexity. In fact, Christol's theorem is usually stated in the equivalent form that for any $r \ge 1$, a power series over $\mathbb{F}_{p^r}$ is

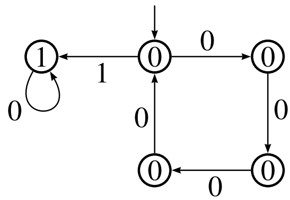


Figure 3. Minimal q-DFAO generating the coefficients of y, where $y^{q^4} - y = x$.

algebraic if and only if its coefficient sequence is p-automatic. The next proposition shows that, for a given q and p, there is no qualitative difference between base-p and base-q complexity, in the sense they are at most a multiplicative constant apart.

Proposition 2.16. *Let $\boldsymbol{a}$ be p-automatic and $q = p^r$. Then*

$$N_q(\boldsymbol{a}) \le N_p(\boldsymbol{a}) \le \frac{q-1}{p-1} N_q(\boldsymbol{a}) \quad \textit{and} \quad N_q^f(\boldsymbol{a}) \le N_p^f(\boldsymbol{a}) \le \frac{q-1}{p-1} N_q^f(\boldsymbol{a}).$$

Proof. First we handle reverse-reading complexity. Without loss of generality, assume that the output alphabet Δ of the DFAO that produces $\boldsymbol{a}$ is a subset of $\mathbb{F}_{p^N}$ for some N with $\mathbb{F}_q \subseteq \mathbb{F}_{p^N}$. The lower bound on $N_p(\boldsymbol{a})$ is clear from the fact that $S_q(y) \subseteq S_p(y)$ for $y = \sum \boldsymbol{a}(n)x^n$.

For the upper bound, let M_p be a minimal reverse-reading p-DFAO that outputs $\boldsymbol{a}$. Observe that M_p contains the q-DFAO M_q (which also outputs $\boldsymbol{a}$) as a “subDFAO”, where the transitions in M_q are achieved by following r-fold transitions inside M_p. So the states of M_p that are not in M_q comprise at most one p-ary tree of height r rooted at each state of M_q. So

$$N_p(\boldsymbol{a}) \le (1 + p + p^2 + \cdots + p^{r-1})|M_q| = \frac{p^r - 1}{p-1} N_q(\boldsymbol{a}),$$

which yields the claimed inequality.

To pass to forward-reading state complexity, follow the dualizing construction of Proposition 2.4 to embed the states of a forward-reading p-DFAO in some vector space over $\mathbb{F}_{p^N}$. Then let S_p^T and S_q^T be monoids consisting of the transposes of the operators in S_p and S_q. The same arguments as above now apply. □

3. Curves, the Cartier operator, and Christol's theorem

3A. ***Curves and the Cartier operator.*** At this point we recall some standard definitions and terminology from the algebraic geometry of curves. For an introduction to the subject, see [Hartshorne 1977, Chapter IV; Silverman 2009, Chapter II; Stichtenoth 2009].

Let k be a perfect field of characteristic p and let X/k be a smooth projective algebraic curve. Denote the function field $k(X)$ by K. Let $\Omega = \Omega_{K/k}$ be the K-vector space of (Kähler) differentials of K/k, which is one-dimensional.

Let P be a (closed) point of X, or equivalently a place of K. (Whenever we refer to points of X, we will always mean closed points.) Write $v_P(f)$ or $v_P(\omega)$ for the valuation given by the order of vanishing of $f \in K^\times$ or $\omega \in \Omega \setminus \{0\}$ at P. The valuation ring $\mathcal{O}_P$ is defined to be

$$\mathcal{O}_P = \{f \in K^\times : v_P(f) \geq 0\} \cup \{0\},$$

with maximal ideal

$$\mathfrak{m}_P = \{f \in K^\times : v_P(f) \geq 1\} \cup \{0\}.$$

The degree $\deg P$ is $\dim_k \mathcal{O}_P/\mathfrak{m}_P$. Write $\operatorname{res}_P(\omega)$ for the residue of ω at P.

Let $\operatorname{Div}_k(X)$ denote the group of k-rational divisors of X. If $f \in K^\times$, define

$$\begin{aligned}(f)_0 &= \sum_{v_P(f)>0} v_P(f)P,\\ (f)_\infty &= \sum_{v_P(f)<0} -v_P(f)P,\\ (f) &= (f)_0 - (f)_\infty.\end{aligned}$$

For $D = \sum_P n_P P \in \operatorname{Div}_k(X)$, write $D \geq 0$ if D is effective, that is, if $n_P \geq 0$ for all P. Define

$$\mathcal{L}(D) = \{f \in K^\times : (f) + D \geq 0\} \cup \{0\} \quad \text{and} \quad \Omega(D) = \{\omega \in \Omega\setminus\{0\} : (\omega) + D \geq 0\} \cup \{0\}.$$

By the Riemann–Roch theorem,

$$\dim_k \Omega(D) = \dim_k \mathcal{L}(-D) + \deg D + g - 1.$$

If D is effective, then $\mathcal{L}(-D) = \{0\}$ and

$$\dim_k \Omega(D) = \deg D + g - 1.$$

For an effective divisor D, it will be convenient to introduce the nonstandard notation $\sqrt{D}$ for the "radical" of D, that is,

$$\sqrt{D} = \sum_{v_P(D)>0} P.$$

Let $x \in K$ be a separating variable ($x \notin K^p$, equivalently $dx \neq 0$). For such an x, there is some point P of X such that $v_P(x)$ is not divisible by p. By an easy argument using valuations at P, the powers $1, x, x^2, \ldots, x^{p-1}$ are linearly independent over K^p. As $[K : K^p] = p$ by standard facts about purely inseparable extensions [Stichtenoth 2009, Proposition 3.10.2], the set $\{1, x, \dots, x^{p-1}\}$ forms a basis of K over K^p. Thus, any $\omega \in \Omega$ can be written as

$$\omega = (u_0^p + u_1^p x + \cdots + u_{p-1}^p x^{p-1})\, dx, \tag{3.1}$$

for unique $u_0, \dots, u_{p-1} \in K$. Define a map $\mathcal{C} : \Omega \to \Omega$ by

$$\mathcal{C}(\omega) = u_{p-1}\, dx. \tag{3.2}$$

It is true, but far from obvious, that $\mathcal{C}$ does not depend on the choice of x [Stichtenoth 2009, p. 183]. The operator $\mathcal{C}$ is an $\mathbb{F}_p$-linear endomorphism of Ω known as the Cartier operator. This operator is of great importance in characteristic-p algebraic geometry. It can be extended in a natural way to r-forms of higher-dimensional varieties for any r, though we do not need this for our purposes (see for example [Cartier 1957; Serre 1958]).

It follows from the definition of $\mathcal{C}$ that for any $\omega \in \Omega$,

$$\omega = \sum_{i=0}^{p-1} x^i \left(\frac{\mathcal{C}(x^{p-1-i}\omega)}{dx} \right)^p dx. \tag{3.3}$$

Comparing equations (2.9) and (3.3) motivates the following definition. For $i \in \{0, 1, \dots, p-1\}$, define the *twisted Cartier operator* $\sigma_i : \Omega \to \Omega$ by

$$\sigma_i(\omega) = \mathcal{C}(x^{p-1-i}\omega). \tag{3.4}$$

For this to make sense in an arbitrary function field K, we need to fix a distinguished separating $x \in K$ in advance (equivalently, a distinguished separable cover $X \to \mathbb{P}^1$). Having done so, if $y \in k((x)) \cap K$, it is clear that

$$\sigma_i(y\, dx) = \Lambda_i(y)\, dx, \tag{3.5}$$

so the σ_i act on differentials just as the Λ_i act on series.

Remark 3.6. Equation (3.5) is true in the differential module of any function field K that contains the Laurent series y, as long as $x \in K$ is separating. In particular, we can take $K = k(x)[y]$, as the Laurent series field $k((x))$ is a separable extension of $k(x)$. This proves that if y is algebraic, then the operators Λ_i map $k(x)[y]$ into itself. This is not at all obvious from the definition of Λ_i as an operator on formal Laurent series.

We summarize some important properties of $\mathcal{C}$ in the next proposition. These are standard (see e.g., [Stichtenoth 2009, p. 182]), but we sketch proofs for the convenience of the reader.

Proposition 3.7. *For any $\omega, \omega' \in \Omega$, $f \in K$, and any point P of X:*

(1) $\mathcal{C}(\omega + \omega') = \mathcal{C}(\omega) + \mathcal{C}(\omega')$.

(2) $\mathcal{C}(f^p\omega) = f\mathcal{C}(\omega)$.

(3) $\mathcal{C}(\omega) = 0$ *if and only if* $\omega = dg$ *for some* $g \in K$.

(4) *If* ω *is regular at* P, *then so is* $\mathcal{C}(\omega)$.

(5) *If* ω *has a pole at* P, *then* $v_P(\mathcal{C}(\omega)) \geq (v_P(\omega)+1)/p - 1$, *and equality holds if the RHS is an integer. In particular, if* $v_P(\omega) = -1$, *then* $v_P(\mathcal{C}(\omega)) = -1$.

(6) *If* $\deg P = 1$, *then* $\operatorname{res}_P(\mathcal{C}(\omega))^p = \operatorname{res}_P(\omega)$.

Proof. Statements (1) and (2) are immediate from the definition and imply that $\mathcal{C}$ is $\mathbb{F}_p$-linear.

Statement (3) follows from the fact that there is no $g \in K$ such that $dg = x^{p-1}\,dx$. If there were such a g, we would have $dg/dx = x^{p-1}$, but this is impossible because the derivative of x^p is zero. For the converse, if $u_{p-1} = 0$, set

$$g = u_0^p x + u_1^p \frac{x^2}{2} + \cdots + u_{p-2}^p \frac{x^{p-1}}{p-1},$$

and note that no denominator is zero. Then $\omega = dg$.

For statement (4), choose a uniformizer t at P, which is necessarily separating (if t is a p-th power, then $v_P(t)$ is a multiple of p and t cannot be a uniformizer at P). Let $\omega = f\,dt$. If ω is regular at P, then so is f, because $v_P(dt) = 0$. So, f can be written as a power series in t and $\mathcal{C}(\omega) = \Lambda_{p-1}(f)\,dt$. The series $\Lambda_{p-1}(f)$ is regular at P, so $\mathcal{C}(\omega)$ is also. Statements (5) and (6) follow similarly by writing f as a Laurent series in t. □

3B. ***Christol's theorem and complexity bounds.*** At this point we fix a prime p and a prime power $q = p^r$. Let $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$ be algebraic of degree d, height h, and genus g. Let X be the normalization of the projective closure of the affine curve defined by the minimal polynomial of y (after clearing denominators). Set $K = \mathbb{F}_q(X)$ and $\Omega = \Omega_{K/\mathbb{F}_q}$.

Define $\mathcal{S}_q$ to be the monoid generated by all r-fold compositions of the operators $\{\sigma_0, \ldots, \sigma_{p-1}\}$. In particular, $\mathcal{S}_p = \langle \sigma_0, \ldots, \sigma_{p-1}\rangle$. We write $\mathcal{S}_q(\omega)$ for the orbit of ω under $\mathcal{S}_q$. Note that $dx \neq 0$ because $\mathbb{F}_q(x) \subseteq K \subseteq \mathbb{F}_q((x))$, so $K/\mathbb{F}_q(x)$ is separable (the Laurent series field is a separable extension of the rational function field). Therefore (3.5) holds, so the orbits $\mathcal{S}_q(y\,dx)$ and $S_q(y)$ are in bijection. So if $\mathcal{S}_q(y\,dx)$ is finite, then $\boldsymbol{a}$ is p-automatic, and $|\mathcal{S}_q(y\,dx)| = N_q(\boldsymbol{a})$.

We now present the proof of the "algebraic implies automatic" direction of Christol's theorem due to David Speyer [2010]. As indicated above, the crux of the argument is to show that the orbit $\mathcal{S}_q(y\,dx)$ is finite. By Proposition 2.16, it loses essentially nothing to replace $\mathcal{S}_q(y\,dx)$ with the larger orbit $\mathcal{S}_p(y\,dx)$.

Proposition 3.8 (Speyer). *The sequence* $\boldsymbol{a}$ *is* q*-automatic.*

Proof. Let P be a point of X. By Proposition 3.7, if neither x nor ω has a pole at P, then $\sigma_i(\omega) = \mathcal{C}(x^{p-i-1}\omega)$ does not have a pole at P for any $i \in \{0, \dots, p-1\}$. Therefore, the only places where elements of $\mathcal{S}_p(y\,dx)$ can have poles are the finitely many poles of $y\,dx$ and of x.

Now assume that P is a pole of $y\,dx$ or of x. Let $n = v_P(y\,dx)$ and $m = v_P(x)$. Applying the inequality of Proposition 3.7 gives

$$v_P(\sigma_i(y\,dx)) \geq \frac{n+m(p-1-i)+1}{p} - 1.$$

The pole of largest order that $\sigma_i(y\,dx)$ could have at P occurs when both n and m are negative. In this case,

$$v_P(\sigma_i(y\,dx)) \geq \frac{n}{p} + \frac{m(p-1)}{p} + \frac{1}{p} - 1.$$

Applying the same reasoning again shows that

$$v_P(\sigma_j\sigma_i(y\,dx)) \geq \frac{n}{p^2} + \frac{m(p-1)}{p^2} + \frac{m(p-1)}{p} + \frac{1}{p^2} - 1,$$

and applying it k times shows that

$$\begin{aligned} v_P(\sigma_{i_k}\cdots\sigma_{i_1}(y\,dx)) &\geq \frac{n}{p^k} + m(p-1)\left(\frac{1}{p^k} + \frac{1}{p^{k-1}} + \dots + \frac{1}{p}\right) + \frac{1}{p^k} - 1 \\ &\geq n + \frac{m(p-1)}{p-1} + \frac{1}{p^k} - 1 \\ &> n + m - 1. \end{aligned}$$

Therefore $v_P(\omega) \geq n + m$ for any $\omega \in \mathcal{S}_p(y\,dx)$. (If one of $\{n, m\}$ is positive, then it follows in the same way that $v_P(\omega) \geq \min\{n, m\}$ instead.)

The differentials in $\mathcal{S}_p(y\,dx)$ have poles at only finitely many places, and the orders of these poles are bounded. So there is a finite-dimensional $\mathbb{F}_q$-vector space that contains $\mathcal{S}_p(y\,dx)$, and in particular $\mathcal{S}_p(y\,dx)$ is finite. □

The Riemann–Roch bound implicit in Proposition 3.8 gives a complexity bound that is a preliminary version of Theorem 1.2. This is Corollary 3.10, for which it will be convenient to use the language of representations. Let $v = y\,dx$, and let V and λ be as in the setup before Proposition 3.8. Let $\phi : \Sigma_q^* \to \operatorname{End}(V)$ be the unique monoid morphism defined for $c \in \Sigma_q$ by

$$\phi(c) = \sigma_{i_r}\sigma_{i_{r-1}}\cdots\sigma_{i_1},$$

where $c = i_r p^{r-1} + i_{r-1}p^{r-2} + \dots + i_2 p + i_1$ with $0 \leq i_1, \dots, i_r \leq p-1$. Then by the power series machinery of Section 2C and the bijection between $S_q(y)$ and $\mathcal{S}_q(y\,dx)$, we see that (V, v, ϕ, λ) gives a q-representation of $\boldsymbol{a}$.

Remark 3.9. If D is any divisor such that $\mathcal{S}_q(y\,dx) \subseteq \Omega(D)$, then we can identify λ with an element of $H^1(X, \mathcal{O}_X(-D))$. This is because of the natural duality isomorphism

$$H^1(X, \mathcal{O}_X(-D))^* \simeq H^0(X, \Omega_X^1(D)),$$

which is the classical statement of Serre duality for curves [Hartshorne 1977, Chapter III.7]. (The global sections of $\Omega_X^1(D)$ are exactly what we have called $\Omega(D)$.) In fact, λ has an explicit realization as a repartition (adele). See, e.g., [Stichtenoth 2009, Chapter 1.5; Serre 1958, p. 37]. Moreover, the Cartier operator on $\Omega(D)$ is the transpose (or adjoint) of the Frobenius operator on $H^1(X, \mathcal{O}_X(-D))$.

Corollary 3.10. $\max(N_q^f(\boldsymbol{a}), N_q(\boldsymbol{a})) \leq q^{h+3d+g-1}$.

Proof. The bounds on the orders of poles in the proof of Proposition 3.8 show that

$$\mathcal{S}_q(y\,dx) \subseteq \Omega((y\,dx)_\infty + (x)_\infty),$$

so we may take $V = \Omega((y\,dx)_\infty + (x)_\infty)$ in the q-representation of $\boldsymbol{a}$. By Proposition 2.4, both $N_q^f(\boldsymbol{a})$ and $N_q(\boldsymbol{a})$ are at most $|V| = q^{\dim_{\mathbb{F}_q} V}$. We have

$$\dim_{\mathbb{F}_q} V = \deg((y\,dx)_\infty + (x)_\infty) + g - 1 \leq \deg(y\,dx)_\infty + \deg(x)_\infty + g - 1.$$

Let $\pi_x, \pi_y : X \to \mathbb{P}^1$ be the projection maps from X onto the x- and y-coordinates. We have $\deg(x)_\infty = \deg \pi_x = d$ and $\deg(y)_\infty = \deg \pi_y = h$. (The easiest way to see this is by looking at the function field inclusions $\pi_x^* : \mathbb{F}_q(x) \hookrightarrow K$ and $\pi_y^* : \mathbb{F}_q(y) \hookrightarrow K$. That is, $d = [K : \mathbb{F}_q(x)]$ and $h = [K : \mathbb{F}_q(y)]$.)

The poles of dx occur at points which are poles of x, and the order of a pole of dx at P can be at most one more than the order of the pole of x at P. So $\deg(dx)_\infty$ is maximized when the poles of x are all simple, in which case $\deg(dx)_\infty = 2\deg(x)_\infty = 2d$. The fact that $\deg(y\,dx)_\infty \leq \deg(y)_\infty + \deg(dx)_\infty$ gives the upper bound. $\square$

The bound in Corollary 3.10 is superseded by Theorem 1.2 for large values of h, d, and g. However, it is simple to prove and is already much better than the previous bounds derived from Ore's Lemma.

We aim to prove Theorem 1.2 by bounding the size of the orbit $\mathcal{S}_q(y\,dx)$. As in the proof of Proposition 3.8, it will be easier to deal with the larger orbit $\mathcal{S}_p(y\,dx)$. By Proposition 2.16, this creates no essential difference in the size of the orbit. To streamline the exposition, we establish some preliminary lemmas. Lemmas 3.11 and 3.12 determine the "eventual behavior" of $y\,dx$ under $\mathcal{S}_p$. The main difficulty is in handling the orbit of $y\,dx$ under the operator σ_0 (this is related to the special role that 0 plays in nonuniqueness of base expansions). Recall that $\sqrt{D}$ is the sum of the points in the support of the divisor D, neglecting multiplicities.

Lemma 3.11. *Let* $V=\Omega((y)_\infty+(x)_\infty+\sqrt{(x)_\infty})$ *and* $W=\Omega((y)_\infty+(x)_\infty)$. *Then for any* $i\in\{1,\dots,p-1\}$, $\sigma_i(V)\subseteq W$, *and for any* $i\in\{0,\dots,p-1\}$, $\sigma_i(W)\subseteq W$.

Proof. Let $\omega\in V$. Then for any point P, $v_P(\omega)\geq -v_P((y)_\infty)-2v_P((x)_\infty)$. By Proposition 3.7,

$$\begin{aligned} v_P(\sigma_i(y\,dx)) &= v_P(\mathcal{C}(x^{p-1-i}y\,dx)) \\ &\geq \frac{-v_P((y)_\infty)+(-p-1+i)v_P((x)_\infty)+1}{p}-1 \\ &\geq \frac{-v_P((y)_\infty)+1}{p}-\frac{pv_P((x)_\infty)}{p}-1 \\ &\geq -v_P((y)_\infty)-v_P((x)_\infty), \end{aligned}$$

where we have used that $i\geq 1$. A similar calculation shows that $\sigma_i(W)\subseteq W$ for any i. □

Lemma 3.12. *Let* T *be the maximum order of any pole of* y *or zero of* x. *Then*

$$\sigma_0^\ell(y\,dx)\in\Omega\bigl(\sqrt{(y)_\infty}-(x_0)+\sqrt{(x)_0}+(x)_\infty+\sqrt{(x)_\infty}\bigr)$$

for $\ell\geq\lceil\log_p(T)\rceil$.

Proof. As in the proof of Proposition 3.8, any $\omega\in\mathcal{S}_p(y\,dx)$ can have poles only at the poles of $y\,dx$ or of x. Writing locally in Laurent series expansions shows that the poles of $y\,dx$ are all either poles of y or poles of x, and in fact $y\,dx\in\Omega((y)_\infty+(x)_\infty+\sqrt{(x)_\infty})$. For any $\omega\in\Omega$, we compute

$$\sigma_0(\omega)=\mathcal{C}\Bigl(\frac{x^p\omega}{x}\Bigr)=x\mathcal{C}\Bigl(\frac{\omega}{x}\Bigr)$$

and therefore

$$\sigma_0^n(\omega)=x\mathcal{C}^n\Bigl(\frac{\omega}{x}\Bigr)$$

for every $n\geq 1$.

Let $\alpha=(y\,dx)/x$. We have $\alpha\in\Omega((y)_\infty+(x)_0+\sqrt{(x)_\infty})$. As y is a power series in x, it must be the case that y is a regular function at every zero of x, so no point can be both a pole of y and a zero of x. Let P be a point that is either a pole of y or a zero of x. Thus $v_P(\alpha)\geq -T$. Repeatedly applying Proposition 3.7, we see that $v_P(\mathcal{C}^\ell(\alpha))\geq -1$ for $\ell\geq\log_p(T)$. Therefore $\mathcal{C}^\ell(\alpha)\in\Omega(\sqrt{(y)_\infty}+\sqrt{(x)_0}+\sqrt{(x)_\infty})$. As $\sigma_0^\ell(y\,dx)=x\mathcal{C}^\ell(\alpha)$, we conclude

$$\sigma_0^\ell(y\,dx)\in\Omega\bigl(\sqrt{(y)_\infty}+\sqrt{(x)_0}+\sqrt{(x)_\infty}-(x)_0+(x)_\infty\bigr)$$

as claimed. □

The next lemma handles the repeated action of $\mathcal{C}$ on differentials with simple poles.

Lemma 3.13. *Suppose $\omega \in K$ has simple poles at points of degrees $e_1, e_2, \dots, e_n$. Let m be the LCM of $e_1, \dots, e_n$. Then $\mathcal{C}^{rm}(\omega) - \omega$ is holomorphic (recall $q = p^r$).*

Proof. Let X' be the base change $X' = X \otimes_{\mathbb{F}_q} \mathbb{F}_{q^m}$ with base change morphism $\phi : X' \to X$. Let $K' = \phi^* K$, which is the constant field extension $\mathbb{F}_{q^m} K$. Each place P of K which is a pole of ω splits completely in the extension to K' (for example, by [Stichtenoth 2009, Theorem 3.6.3 g] each place P' lying over P has residue field equal to $\mathbb{F}_{q^m}$). Therefore the pullback $\phi^*\omega$ has simple poles at places of degree 1. So $\mathcal{C}(\phi^*\omega)$ has simple poles at the same places as $\phi^*\omega$. At each of these places P', we compute

$$\operatorname{res}_{P'}(\mathcal{C}^{rm}(\phi^*\omega)) = \operatorname{res}_{P'}(\phi^*\omega)^{(1/p)^{rm}} = \operatorname{res}_{P'}(\phi^*\omega)^{q^{-m}} = \operatorname{res}_{P'}(\phi^*\omega),$$

because the residue lies in $\mathbb{F}_{q^m}$, which is fixed under the q^m-th power map. So $\mathcal{C}^{rm}(\phi^*\omega)$ has simple poles at the same places as $\phi^*\omega$ with the same residues, and therefore $\mathcal{C}^{rm}(\phi^*\omega) - \phi^*\omega$ is holomorphic. The Cartier operator commutes with pullback, so

$$\mathcal{C}^{rm}(\phi^*\omega) - \phi^*\omega = \phi^*(\mathcal{C}^{rm}(\omega) - \omega)$$

and we conclude that $\mathcal{C}^{rm}(\omega) - \omega$ is also holomorphic. □

Using the preceding lemmas, we now prove Theorem 1.2.

Proof of Theorem 1.2. Let $V = \Omega((y)_\infty + (x)_\infty + \sqrt{(x)_\infty})$ and $W = \Omega((y)_\infty) + (x)_\infty)$. By Lemma 3.11, $\sigma_i(y\,dx) \in W$ for every $i > 0$, and W is σ_i-invariant for every i. So we have

$$\left|N_q(\boldsymbol{a})\right| = \left|\mathcal{S}_q(y\,dx)\right| \le \left|\mathcal{S}_p(y\,dx)\right| \le 1 + \left|\{\sigma_0^n(y\,dx) : n \ge 1\}\right| + \left|W\right|.$$

By Riemann–Roch, $\dim_{\mathbb{F}_q} W = \deg((y)_\infty + (x)_\infty) + g - 1 \le h + d + g - 1$. The remainder of the proof will handle the orbit of $y\,dx$ under σ_0.

Let T be the maximum order of any pole of y or zero of x. Let

$$D = \sqrt{(y)_\infty} + \sqrt{(x)_0} + \sqrt{(x)_\infty}.$$

By Lemma 3.12, for $n \ge \lceil \log_p(T) \rceil$, we have that $\sigma_0^n(y\,dx) \in \Omega(D - (x))$. Let $\alpha = x^{-1}\sigma_0^{\lceil \log_p(T) \rceil}(y\,dx)$. So $\alpha \in \Omega(D)$, that is, α has simple poles at points that are either poles of y, poles of x, or zeroes of x. We have seen that $x\mathcal{C}^n(\alpha) = \sigma_0^n(x\alpha)$. It follows that

$$\left|\{\mathcal{C}^n(\alpha) : n \ge 0\}\right| = \left|\{\sigma^n(y\,dx) : n \ge \log_p(T)\}\right|.$$

Let m be the LCM of the degrees of the points at which α has a pole. By Lemma 3.13, $\mathcal{C}^{rm}(\alpha) - \alpha$ is holomorphic. The space of holomorphic differentials is invariant under $\mathcal{C}$, so the orbit of α under $\mathcal{C}$ is contained in the set

$$\{\mathcal{C}^k(\alpha) + \eta : 0 \le k < rm \text{ and } \eta \in \Omega(0)\}.$$

This set has size at most $rm|\Omega(0)| = rmq^g$. Thus

$$\left|\{\sigma_0^n(\omega) : n \geq 1\}\right| \leq \left\lceil \log_p(T) \right\rceil + rmq^g.$$

We now need to estimate m.

Let $L(n)$ be Landau's function, that is, the largest LCM of all partitions of n, or equivalently the maximum order of an element in the symmetric group of order n. Recall from the proof of Corollary 3.10 that $\deg(y)_\infty = h$ and $\deg(x)_\infty = d$. We have

$$\sum_{v_P(y)<0} \deg P \leq \sum_{v_P(y)<0} -v_P(y) \deg P = \deg(y_\infty) = h,$$

and it follows in the same way that $\sum_{v_P(x)<0} \deg P \leq d$ and $\sum_{v_P(x)>0} \deg P \leq d$. Therefore $m \leq L(h)L(d)^2$. It is clear that $L(a)L(b) \leq L(a+b)$ for all a and b, so $m \leq L(h+2d)$. So

$$\left|\{\sigma_0^n(y\,dx) : n \geq 1\}\right| \leq \left\lceil \log_p(T) \right\rceil + rL(h+2d)q^g.$$

Therefore $|\mathcal{S}_p(y\,dx)| \leq 1 + \lceil \log_p(T) \rceil + rL(h+2d)q^g + q^{h+d+g-1}$.

It remains to show that the quantity

$$\frac{1 + \left\lceil \log_p(T) \right\rceil + rL(h+2d)q^g}{q^{h+d+g-1}}$$

decays to zero as any of q, h, d, or g grow to ∞. This follows easily from the fact that $g \geq 0$ and $h + d \geq 2$ for any algebraic curve, the simple bound on Landau's function

$$L(n) \leq \exp\bigl((1 + o(1))\sqrt{n \log n}\,\bigr)$$

from [Massias et al. 1989], and the fact that $T \leq \max(h, d)$. □

The forward-reading complexity bound of Theorem 1.5 follows as an easy corollary.

Proof of Theorem 1.5. Let $V = \Omega((y)_\infty + (x)_\infty + \sqrt{(x)_\infty})$ and let $\lambda \in V^*$ be the linear functional that maps ω to the constant term of the power series $\frac{\omega}{dx}$. We have $\dim_{\mathbb{F}_q} V \leq h + 2d + g - 1$, so

$$|N_q^f(\boldsymbol{a})| = |\mathcal{S}_q^T(\lambda)| \leq |V^*| \leq q^{h+2d+g-1}$$

by Proposition 2.4. □

We now show that Theorem 1.2 is qualitatively sharp for the power series expansions of rational functions, that is, it is sharp if we replace the "error term" $o(1)$ by zero.

Proposition 3.14. *For every prime power q and every positive integer $h \geq 1$, there exists $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_q[\![x]\!]$ with* $\deg y = 1$ *and $h(y) = h$ (and therefore $g = 0$) such that $N_q(\boldsymbol{a}) \geq q^h$.*

Proof. Let $f = x^h + c_{h-1}x^{h-1} + \cdots + c_1 x + c_0 \in \mathbb{F}_q[x]$ be any primitive polynomial, that is, such that a root of f generates $\mathbb{F}_{q^h}^{\times}$. Let $y = f^{-1} - 1 \in \mathbb{F}_q[\![x]\!]$. The coefficient sequence $\boldsymbol{a}$ of y satisfies the linear recurrence relation

$$c_0\boldsymbol{a}(n) + c_1\boldsymbol{a}(n-1) + \cdots + c_{h-1}\boldsymbol{a}(n-h+1) + \boldsymbol{a}(n-h) = 0$$

and $\boldsymbol{a}$ is eventually periodic with minimal period $q^h - 1$ [Lidl and Niederreiter 1994, Theorem 6.28]. We have $\deg y = 1$ and $h(y) = h$, so the curve X is $\mathbb{P}^1$, with $K = \mathbb{F}_q(x)$ and $g = 0$.

Note that $(y)_\infty = (f)_0$ is a single point of degree h. Let P_∞ be the pole of x, which is distinct from $(f)_0$. We compute

$$(f^{-1}\,dx) = hP_\infty - (f)_0 - 2P_\infty = (h-2)P_\infty - (f)_0,$$

so $f^{-1}\,dx$ has at most two poles: a simple pole at $(f)_0$ of degree h, and if $h = 1$, a simple pole at P_∞ of degree 1. By Lemma 3.13, $\mathcal{C}^{rh}(y\,dx) - f^{-1}\,dx = \mathcal{C}^{rh}(f^{-1}\,dx) - f^{-1}\,dx$ is holomorphic (note $\mathcal{C}(dx) = 0$ by Proposition 3.7). As X has genus 0, it carries no nonzero holomorphic differentials, so $\mathcal{C}^{rh}(y\,dx) = f^{-1}\,dx$.

Let $\boldsymbol{b}$ be the coefficient sequence of f^{-1}. The sequence $\boldsymbol{b}$ satisfies a linear recurrence relation of degree h and has period $q^h - 1$, so it must be the case that all possible strings of h elements in $\mathbb{F}_q$ except for the string $(0, \ldots, 0)$ occur in $\boldsymbol{b}$ within the first $q^h - 1$ terms. For each $0 \leq c \leq q-1$, a certain r-fold composition of Λ_i operators $s \in S_q$ gives $s(f^{-1}) = \sum_{n=0}^{\infty} \boldsymbol{b}(qn + c)x^n$, so

$$\Lambda_0^{rh} s(f^{-1}) = \sum_{n=0}^{\infty} \boldsymbol{b}(q^h n + c)x^n = \sum_{n=c}^{\infty} \boldsymbol{b}(n)x^n,$$

by the periodicity of $\boldsymbol{b}$. So there are at least $q^h - 1$ distinct power series in $S_q(f^{-1})$.

Let $V = \Omega((f)_0 + P_\infty)$. We have $(f^{-1}\,dx) \in V$, and $\dim_{\mathbb{F}_q} V = h$ by Riemann–Roch. A calculation with properties of $\mathcal{C}$ and orders of poles shows that V is σ_i-invariant for any $i \in \{0, \ldots, p-1\}$, so $\mathcal{S}_q(f^{-1}\,dx) \subseteq V$. The counting argument from the previous paragraph shows that the orbit $\mathcal{S}_q(f^{-1}\,dx)$ comprises all nonzero elements of V. (In fact, it is not hard to show that $\sigma_i|_V$ is invertible for each i, so the action of $\mathcal{S}_p$ on V is a group action with precisely two orbits: $\{0\}$ and $V \setminus \{0\}$.) Note that $y\,dx \notin V$, for if $y\,dx$ were in V, then dx would be also, but $(dx) = -2P_\infty$. So $y \notin S_q(f^{-1})$, which establishes the lower bound $|S_q(y)| = N_q(\boldsymbol{a}) \geq q^h - 1 + 1 = q^h$. ∎

3C. ***Variation mod primes.*** Let K be a number field and $y=\sum_{n=0}^{\infty}\boldsymbol{a}(n)x^n\in K[\![x]\!]$. If the prime $\mathfrak{p}$ of K is such that $v_{\mathfrak{p}}(\boldsymbol{a}(n))\geq 0$ for all n, let $\boldsymbol{a}_{\mathfrak{p}}$ denote the reduction of $\boldsymbol{a}$ mod $\mathfrak{p}$, and let $y_{\mathfrak{p}}=\sum_{n=0}^{\infty}\boldsymbol{a}_{\mathfrak{p}}(n)x^n$ be the reduced power series with coefficients in the residue field $k(\mathfrak{p})$.

Suppose y is algebraic over $K(x)$. By a theorem of Eisenstein [1852], there are only finitely many primes $\mathfrak{p}$ such that $v_{\mathfrak{p}}(\boldsymbol{a}(n))<0$ for some n (see [Schmidt 1990] for exposition). So the sequence $\boldsymbol{a}_{\mathfrak{p}}$ is defined for all but finitely many $\mathfrak{p}$, and by Christol's theorem it is $|k(\mathfrak{p})|$-automatic (it is an easy observation that the reduction mod $\mathfrak{p}$ of an algebraic function is algebraic). An extension of our main question is how the algebraic nature of y affects the complexity $N_{|k(\mathfrak{p})|}(\boldsymbol{a}_{\mathfrak{p}})$ as the prime $\mathfrak{p}$ varies. Theorem 3.15 answers this question in the case that the complexities are bounded at all primes; in this case y must have a very special form. Note that we do not need to assume that y is algebraic in the statement of the theorem. To simplify notation, we will write $N_{\mathfrak{p}}(\boldsymbol{a}_{\mathfrak{p}})$ in place of $N_{|k(\mathfrak{p})|}(\boldsymbol{a}_{\mathfrak{p}})$.

Theorem 3.15. *Let $y=\sum_{n=0}^{\infty}\boldsymbol{a}(n)x^n\in K[\![x]\!]$. Then $N_{\mathfrak{p}}(\boldsymbol{a}_{\mathfrak{p}})$ and $N_{\mathfrak{p}}^f(\boldsymbol{a}_{\mathfrak{p}})$ are bounded independently of $\mathfrak{p}$ if and only if y is a rational function with at worst simple poles that occur at roots of unity* (*except possibly for a pole at ∞, which may be of any order*).

Proof. It suffices to prove the theorem for reverse-reading complexity, as $N_{\mathfrak{p}}^f(\boldsymbol{a})\leq p^{N_{\mathfrak{p}}(\boldsymbol{a})}$ for any sequence $\boldsymbol{a}$ by Proposition 2.4. Assume that $N_{\mathfrak{p}}(\boldsymbol{a})$ is uniformly bounded for all $\mathfrak{p}$ (such that it is defined). Then the coefficient sequence $\boldsymbol{a}$ assumes a bounded number of values under reduction mod $\mathfrak{p}$ regardless of $\mathfrak{p}$, and so $\boldsymbol{a}$ assumes finitely many values in K. Let A be this finite subset of K.

Choose two primes $\mathfrak{p}$ and $\mathfrak{q}$ such that $|k(\mathfrak{p})|$ and $|k(\mathfrak{q})|$ are multiplicatively independent integers (i.e., $\operatorname{char}k(\mathfrak{p})\neq\operatorname{char}k(\mathfrak{q})$) and all elements of A are distinct both mod $\mathfrak{p}$ and mod $\mathfrak{q}$ (this is possible because only finitely many primes divide distances between distinct elements of A). The reduced power series $y_{\mathfrak{p}}\in k(\mathfrak{p})[\![x]\!]$ and $y_{\mathfrak{q}}\in k(\mathfrak{q})[\![x]\!]$ are both algebraic by Christol's theorem. So there exist injections $i_{\mathfrak{p}}:A\hookrightarrow k(\mathfrak{p})$ and $i_{\mathfrak{q}}:A\hookrightarrow k(\mathfrak{q})$ such that the sequence $\boldsymbol{b}(n)=i_{\mathfrak{p}}(\boldsymbol{a}(n))$ is $|k(\mathfrak{p})|$-automatic and the sequence $\boldsymbol{c}(n)=i_{\mathfrak{q}}(\boldsymbol{a}(n))$ is $|k(\mathfrak{q})|$-automatic. Therefore $\boldsymbol{a}$ is both $|k(\mathfrak{p})|$-automatic and $|k(\mathfrak{q})|$-automatic. By Cobham's theorem [Allouche and Shallit 2003, Theorem 11.2.2], $\boldsymbol{a}$ is an eventually periodic sequence of some period m, so y is a rational function of the form $y=f/(1-x^m)$ for some polynomial f. So the (finite) poles of y are simple and occur at roots of unity.

Conversely, assume that the (finite) poles of y are simple and occur at roots of unity. Therefore $y=f/(1-x^m)$ for some m and $f\in K[x]$, and the coefficient sequence $\boldsymbol{a}$ is eventually periodic of (possibly nonminimal) period m, that is, there is some c such that $\boldsymbol{a}(n+m)=\boldsymbol{a}(n)$ for all $n>c$. In particular, $\boldsymbol{a}$ assumes finitely many values. An easy decimation argument now shows that $N_{\mathfrak{p}}(\boldsymbol{a})$ is uniformly

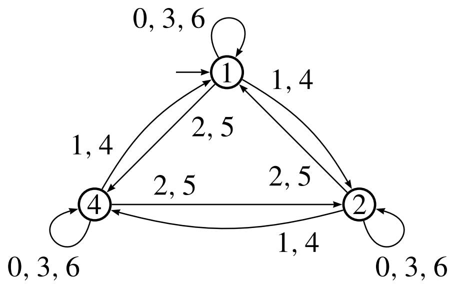


Figure 4. 7-DFAO generating powers of 2 mod 7.

bounded for all primes. Suppose $|k(\mathfrak{p})| = p^r$ and assume that $p^r > c$ (which excludes only finitely many $\mathfrak{p}$). For any $i \geq 1$ and $j \in \{0, \ldots, p^{ri} - 1\}$, the subsequences $\boldsymbol{a}(p^{ri}n + j)$ are periodic of period m beginning with the second term, and they assume the same finite set of values as $\boldsymbol{a}$. There are clearly only finitely many sequences that fit this description. So the size of the p^r-kernel of $\boldsymbol{a}$, and therefore $N_{\mathfrak{p}}(\boldsymbol{a})$, is bounded independently of p. □

4. Examples

We give three detailed examples of computing the state complexity of an automatic sequence. Examples 4.2 and 4.3 in particular show the usefulness of the algebro-geometric approach.

Example 4.1.
$$y = \frac{1}{1-2x}.$$

Let p be odd. Let $\boldsymbol{a}(n) = 2^n \bmod p$ and $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_p[\![x]\!]$. We have $y(1-2x) = 1$, so y has degree 1, height 1, and genus 0. By Theorem 1.2, $N_p(\boldsymbol{a}) \leq (1 + o_p(1))p$. We compute $\Lambda_i(y) = 2^i y$, so $S_p(y) = \{2^i y : i \geq 0\}$, and $N_p(\boldsymbol{a}) = \mathrm{ord}_p(2)$. So in fact

$$\lceil \log_2(p) \rceil \leq N_p(\boldsymbol{a}) \leq p - 1.$$

If there are infinitely many Mersenne primes, the lower bound is sharp infinitely often, and if Artin's conjecture is true, the upper bound is sharp infinitely often.

The sequence $\boldsymbol{a}$ has a one-dimensional p-representation where $\phi(i) : v \mapsto 2^i v$. Each $\phi(i)$ can be written as a (symmetric) 1×1 matrix, so the p-antirepresentation on V^* is the same as the original representation, and $N_p^f(\boldsymbol{a}) = N_p(\boldsymbol{a})$. From the automata point of view, this is the obvious fact that the same DFAO outputs $\boldsymbol{a}$ in both the forward-reading and reverse-reading conventions. The transition diagram of the DFAO for $p = 7$ is given in Figure 4.

Example 4.2.
$$y = \frac{1}{\sqrt{1-4x}}.$$

Let p be an odd prime. Let $\boldsymbol{a}(n)$ be the central binomial coefficient $\binom{2n}{n}$ reduced mod p and let $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n \in \mathbb{F}_p[\![x]\!]$. From Newton's formula for the binomial

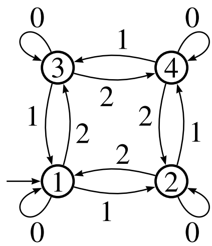


Figure 5. 5-DFAO generating the central binomial coefficients mod 5.

series, we have $y^2(1-4x)=1$. So y has degree 2, height 1, and genus 0, and $N_p(\boldsymbol{a}) \le (1+o_p(1))p^2$. We show that $N_p(\boldsymbol{a}) = N_p^f(\boldsymbol{a}) = p$. (A calculation verifies that $N_2^f(\boldsymbol{a}) = N_2(\boldsymbol{a}) = 2$ as well.)

Let X be the curve defined by $y^2(1-4x)=1$. We have $x = \frac{1}{4} - y^{-2}$, so $X = \mathbb{P}^1(\mathbb{F}_p)$, parametrized by y. Let P_0 and P_∞ be the zero and pole of y, and let $\omega = y\,dx$. A computation gives $dx = -2y^{-3}\,dy$, so

$$(\omega) = -2(P_0) \quad \text{and} \quad (x) = (P_2) + (P_{-2}) - 2(P_0).$$

For any $i \in \{0, \dots, p-1\}$,

$$v_{P_0}(x^{p-1-i}\omega) = (p-1-i)v_{P_0}(x) + v_{P_0}(\omega) > (p-1)(-2) - 2 = -2p.$$

So $v_{P_0}(\sigma_i(\omega)) = v_{P_0}(\mathcal{C}(x^{p-1-i}\omega))$ is either 0, -1, or -2, and P_0 is the only point at which $\sigma_i(\omega)$ can have a pole. A canonical divisor of X has degree -2, so by Riemann–Roch, $\Omega(2P_0)$ is one-dimensional, $\mathcal{S}_p(\omega) \subseteq \mathbb{F}_p\omega$, and $N_p(\boldsymbol{a}) \le p$.

More explicitly, as $\mathcal{S}_p(\omega)$ sits in a one-dimensional vector space we have $\sigma_i(\omega) = c_i\omega$ for some $c_i \in \mathbb{F}_p$. As $\sigma_i(\omega) = \Lambda_i(y)\,dx$ and the constant term of y is 1, c_i is equal to the constant term of $\Lambda_i(y)$, which is $\binom{2i}{i}$. So $\sigma_i(\omega) = \binom{2i}{i}\omega$. Equating coefficients gives

$$\binom{2(pn+i)}{pn+i} \equiv \binom{2i}{i}\binom{2n}{n} \pmod p.$$

(Amusingly, this gives a roundabout argument that recovers a special case of the classical theorem of Lucas on binomial coefficients mod p.)

To see that $N_p(\boldsymbol{a})$ is exactly p, note that $\boldsymbol{a}(1) = 2$, and that for any odd prime q, $\boldsymbol{a}((q+1)/2)$ is the first central binomial coefficient divisible by q. This shows that the subgroup of $\mathbb{F}_p^\times$ generated by all nonzero central binomial coefficients mod p contains all primes less than p and therefore is all of $\mathbb{F}_p^\times$, and furthermore $\boldsymbol{a}((p+1)/2) = 0$.

As in the previous example, the fact that $\mathcal{S}_p(\omega)$ lies in a one-dimensional vector space verifies that $N_p^f(\boldsymbol{a}) = N_p(\boldsymbol{a})$ for all p. The transition diagrams for the automata that output $\boldsymbol{a}$ have nicely symmetric structures. Figure 5 displays the DFAO for $p = 5$ — all undrawn transitions, which are on the inputs 3, 4, and 5, go to an undrawn trap state, which outputs zero.

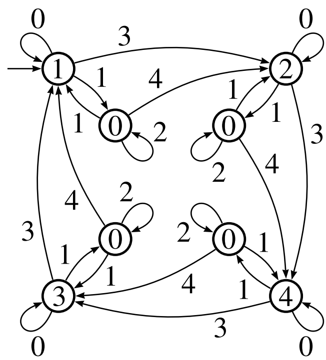


Figure 6. Reverse-reading 5-DFAO generating the coefficients of $(1-4x^3)^{-1/2}$ mod 5.

Example 4.3.
$$y = \frac{1}{\sqrt{1-4x^3}}.$$

Let $p \notin \{2, 3\}$. Let $\boldsymbol{a}$ be the coefficient sequence of the series $1/\sqrt{1-4x^3} \in \mathbb{Q}[[x]]$ reduced mod p, so that $\boldsymbol{a}(3n) = \binom{2n}{n}$ mod p and $\boldsymbol{a}(n) = 0$ if n is not a multiple of 3. Let $y = \sum_{n=0}^{\infty} \boldsymbol{a}(n)x^n$. We have $y^2(1-4x^3) = 1$, so y is of degree 2, height 3, and genus 1, and $N_p(\boldsymbol{a}) \le (1+o_p(1))p^5$. We show that $N_p(\boldsymbol{a}) = 2p-1$.

Let C be the curve defined by $Y^2(Z^3-4W^3) = Z^5$ in $\mathbb{P}^2$. This curve is singular, so define the smooth (elliptic) curve X by $Y^2Z = Z^3 - 4W^3$. The morphism $\phi : X \to C$, defined in homogeneous coordinates by

$$\phi : [W : Y : Z] \mapsto [WY : Z^2 : YZ],$$

gives the normalization of C. The forms $[Y : Z]$ and $[W : Z]$ are maps from C to $\mathbb{P}^1$, so we can consider them as elements of $\mathbb{F}_p(C)$. Let $y = \phi^*[Y : Z]$ and $x = \phi^*[W : Z]$. So we have $y^2(1-4x^3) = 1$. Let $\omega = y\,dx$ and let P_∞ be the point on X written as $[0 : 1 : 0]$ in homogeneous coordinates. The following are easy computations, where P, Q, and R are some points of X that we do not need to compute explicitly:

$$\begin{aligned}
(y) &= 3P_\infty - P - Q - R,\\
(x) &= [0:1:1] + [0:-1:1] - 2P_\infty,\\
(dx) &= -(y).
\end{aligned}$$

In particular, $(\omega) = 0$. Our usual computation for the possible orders of poles of $\sigma_i(\omega)$ shows that $\mathcal{S}_p(\omega) \subseteq \Omega(2P_\infty)$, which has dimension 2 with $\{\omega, x\omega\}$ as a basis. We use properties of $\mathcal{C}$ to compute the action of $\mathcal{S}_p$ on the basis.

As $y^2 = 1/(1-4x^3)$, we have

$$\begin{aligned}\sigma_i(\omega) &= \mathcal{C}(x^{p-i-1}y\,dx)\\ &= \mathcal{C}\Big(x^{p-i-1}\frac{y^p}{y^{p-1}}\,dx\Big)\\ &= y\mathcal{C}(x^{p-i-1}(1-4x^3)^{(p-1)/2}\,dx)\\ &= y\mathcal{C}\bigg(\sum_{k=0}^{(p-1)/2}\binom{(p-1)/2}{k}(-4)^k x^{3k+p-i-1}\,dx\bigg).\end{aligned}$$

So $\sigma_i(\omega)$ is nonzero precisely when there is some $0 \le k \le (p-1)/2$ with

$$3k+p-i-1 \equiv p-1 \pmod p,$$

that is, when $3k \equiv i \pmod p$ has a solution k with $0 \le k \le (p-1)/2$. If there is such a k, then it is unique, and $3k-i \le 3(p-1)/2 < 2p$, so either $3k = i$, in which case

$$\sigma_i(\omega) = \binom{(p-1)/2}{i/3}(-4)^{i/3}\omega,$$

or $3k = i+p$, in which case

$$\sigma_i(\omega) = \binom{(p-1)/2}{(i+p)/3}(-4)^{(i+p)/3}x\omega.$$

This shows that $\mathcal{S}_p(\omega) \subseteq \mathbb{F}_p\omega \cup \mathbb{F}_p x\omega$. Also, the fact that $\sigma_i(\omega) = \Lambda_i(y)\,dx$ proves the identity

$$\binom{(p-1)/2}{k}(-4)^k \equiv \binom{2k}{k} \pmod p$$

for all k.

With this calculation we can explicitly write the restriction of σ_i to $\Omega(2P_\infty)$ by computing its action on the basis $\{\omega, x\omega\}$. So far we have only computed the action of σ_i on ω, but for $i \ge 1$ we have $\sigma_i(x\omega) = \sigma_{i-1}(\omega)$, and $\sigma_0(x\omega) = \mathcal{C}(x^p\omega) = x\mathcal{C}(\omega) = x\sigma_{p-1}(\omega)$. We have

$$\sigma_i(\omega) = \begin{cases}\binom{2i/3}{i/3}\omega & i \equiv 0 \pmod 3,\\ \binom{2(i+p)/3}{(i+p)/3}x\omega & i \equiv -p \pmod 3,\\ 0 & i \equiv p \pmod 3,\end{cases}$$

and

$$\sigma_i(x\omega) = \begin{cases}\binom{2(i-1)/3}{(i-1)/3}\omega & i \equiv 1 \pmod 3,\\ \binom{2(i-1+p)/3}{(i-1+p)/3}x\omega & i \equiv 1-p \pmod 3,\\ 0 & i \equiv 1+p \pmod 3,\end{cases}$$

where $0, p, -p$ are distinct $\bmod p$ because $p > 3$.

Incidentally, it follows from our computation that $\mathcal{C}(\omega) = 0$ if and only if $p \equiv 2 \pmod 3$, which shows that these are precisely the primes for which the elliptic curve X is supersingular, as the classical Hasse invariant is the rank of the restriction of the Cartier operator to the space of holomorphic differentials. See [Silverman 2009, Section 5.4].

Examining the binomial coefficients that appear in the formula for $\sigma_i(\omega)$ shows that $\binom{2k}{k}$ appears as a coefficient on ω for $0 \le k \le \lfloor p/3 \rfloor$, and as a coefficient on $x\omega$ for $\lfloor p/3 \rfloor + 1 \le k \le (2p-1)/3$. As in Example 4.2, the values of $\binom{2k}{k}$ mod p for $0 \le k \le (p-1)/2$ generate the multiplicative group of $\mathbb{F}_p^\times$, and we have $\sigma_1(x\omega) = \omega$. This is already enough to show that $\mathcal{S}_p(\omega) = \mathbb{F}_p\omega \cup \mathbb{F}_p x\omega$, so $N_p(\boldsymbol{a}) = 2p - 1$.

For $p = 5$, the reverse-reading 5-DFAO that outputs $\boldsymbol{a}$ is pictured in Figure 6. As usual, any undrawn transitions lead to a trap state that outputs 0.

Acknowledgements. The author would like to thank David Speyer for first exploring this striking link between automata and geometry and for allowing an exposition of his work in [Speyer 2010]. The author also thanks Eric Bach and Jeffrey Shallit for many helpful conversations related to the topics in the paper, and the referee for a careful reading of the paper and helpful comments. Diagrams of automata were produced using the VauCanSon-G Latex package of Sylvain Lombardy and Jacques Sakarovitch.

References

[Adamczewski and Bell 2012] B. Adamczewski and J. P. Bell, "On vanishing coefficients of algebraic power series over fields of positive characteristic", *Invent. Math.* **187**:2 (2012), 343–393. MR Zbl

[Adamczewski and Bell 2013] B. Adamczewski and J. P. Bell, "Diagonalization and rationalization of algebraic Laurent series", *Ann. Sci. Éc. Norm. Supér.* (4) **46**:6 (2013), 963–1004. MR Zbl

[Allouche and Shallit 2003] J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge University Press, 2003. MR Zbl

[Berstel and Reutenauer 1988] J. Berstel and C. Reutenauer, *Rational series and their languages*, EATCS Monographs on Theoretical Computer Science **12**, Springer, 1988. MR Zbl

[Brzozowski 1963] J. A. Brzozowski, "Canonical regular expressions and minimal state graphs for definite events", pp. 529–561 in *Proc. Sympos. Math. Theory of Automata* (New York, 1962), Polytechnic Press, Brooklyn, NY, 1963. MR Zbl

[Cartier 1957] P. Cartier, "Une nouvelle opération sur les formes différentielles", *C. R. Acad. Sci. Paris* **244** (1957), 426–428. MR Zbl

[Christol 1979] G. Christol, "Ensembles presque periodiques k-reconnaissables", *Theoret. Comput. Sci.* **9**:1 (1979), 141–145. MR Zbl

[Christol et al. 1980] G. Christol, T. Kamae, M. Mendès France, and G. Rauzy, "Suites algébriques, automates et substitutions", *Bull. Soc. Math. France* **108**:4 (1980), 401–419. MR Zbl

[Derksen 2007] H. Derksen, "A Skolem–Mahler–Lech theorem in positive characteristic and finite automata", *Invent. Math.* **168**:1 (2007), 175–224. MR Zbl

[Eilenberg 1974] S. Eilenberg, *Automata, languages, and machines*, vol. A, Pure and Applied Mathematics **58**, Academic Press, New York, 1974. MR Zbl

[Eisenstein 1852] G. Eisenstein, "Über eine allgemeine Eigenschaft der Reihen-Entwicklungen aller algebraischen Funktionen", *Preuss. Akad. d. Wissensch. Berlin* (1852), 441–443. reprinted as pp. 765–767 in his *Mathematische Werke*, vol. II, Chelsea, New York, 1975. MR Zbl

[Fresnel et al. 2000] J. Fresnel, M. Koskas, and B. de Mathan, "Automata and transcendence in positive characteristic", *J. Number Theory* **80**:1 (2000), 1–24. MR Zbl

[Harase 1988] T. Harase, "Algebraic elements in formal power series rings", *Israel J. Math.* **63**:3 (1988), 281–288. MR Zbl

[Harase 1989] T. Harase, "Algebraic elements in formal power series rings, II", *Israel J. Math.* **67**:1 (1989), 62–66. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977. MR Zbl

[Lidl and Niederreiter 1994] R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*, 1st ed., Cambridge University Press, 1994. MR Zbl

[Massias et al. 1989] J.-P. Massias, J.-L. Nicolas, and G. Robin, "Effective bounds for the maximal order of an element in the symmetric group", *Math. Comp.* **53**:188 (1989), 665–678. MR Zbl

[Salomaa and Soittola 1978] A. Salomaa and M. Soittola, *Automata-theoretic aspects of formal power series*, Springer, 1978. MR Zbl

[Schmidt 1990] W. M. Schmidt, "Eisenstein's theorem on power series expansions of algebraic functions", *Acta Arith.* **56**:2 (1990), 161–179. MR Zbl

[Serre 1958] J.-P. Serre, "Sur la topologie des variétés algébriques en caractéristique p", pp. 24–53 in *Symposium internacional de topología algebraica* (Mexico City, 1956), Universidad Nacional Autónoma de México and UNESCO, Mexico City, 1958. MR Zbl

[Shallit 2009] J. Shallit, *A second course in formal languages and automata theory*, Cambridge University Press, 2009. Zbl

[Silverman 2009] J. H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics **106**, Springer, 2009. MR Zbl

[Speyer 2010] D. Speyer, "Christol's theorem and the Cartier operator", 2010, available at https://sbseminar.wordpress.com/2010/02/11/christols-theorem-and-the-cartier-operator/.

[Stichtenoth 2009] H. Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics **254**, Springer, 2009. MR Zbl

Communicated by Kiran S. Kedlaya
Received 2016-06-20 Revised 2016-11-20 Accepted 2016-12-19

andrewbridy@math.tamu.edu Department of Mathematics, Texas A&M University, Department of Mathematics, Mailstop 3368, College Station, TX 77843, United States

ALGEBRA AND NUMBER THEORY 11:3 (2017)
dx.doi.org/10.2140/ant.2017.11.713

On an analogue of the Ichino–Ikeda conjecture for Whittaker coefficients on the metaplectic group

Erez Lapid and Zhengyu Mao

In previous papers we formulated an analogue of the Ichino–Ikeda conjectures for Whittaker–Fourier coefficients of automorphic forms on quasisplit classical groups and the metaplectic group of arbitrary rank. In the latter case we reduced the conjecture to a local identity. In this paper we prove the local identity in the p-adic case, and hence the global conjecture under simplifying conditions at the archimedean places.

1. Introduction

Let G be a quasisplit group over a number field F with ring of adeles $\mathbb{A}$. In [Lapid and Mao 2015a], we formulated (under some hypotheses) a conjecture relating Whittaker–Fourier coefficients of cusp forms on $G(F)\backslash G(\mathbb{A})$ to the Petersson inner product. This conjecture is in the spirit of conjectures of Sakellaridis and Venkatesh [2012] and Ichino and Ikeda [2010], which attempt to generalize the classical

Lapid was partially supported by Grant #711733 from the Minerva Foundation.
Mao was partially supported by NSF grants DMS 1000636 and 1400063 and by a fellowship from the Simons Foundation.
MSC2010: primary 11F30; secondary 11F70.
Keywords: Whittaker coefficients, metaplectic group, automorphic forms.

work of Waldspurger [1985; 1981]. In the case of (quasisplit) classical groups, as well as the metaplectic group (i.e., the metaplectic double cover of the symplectic group), we made this conjecture explicit using the descent construction of Ginzburg, Rallis and Soudry [Ginzburg et al. 2011] and the functorial transfer of generic representations of classical groups by Cogdell, Kim, Piatetski-Shapiro and Shahidi [Cogdell et al. 2001; 2004; 2011].

In a follow-up paper [Lapid and Mao 2017], we reduced the global conjecture in the metaplectic case to a local conjectural identity. We also gave a *purely formal* argument for the case of $\widetilde{\mathrm{Sp}}_1$ (i.e., ignoring convergence issues). In this paper we prove the local identity in the p-adic case, justifying the heuristic analysis (and extending it to the general case).

Let us recall the conjecture of [Lapid and Mao 2015a] in the case of the metaplectic group $\widetilde{\mathrm{Sp}}_n(\mathbb{A})$, the double cover of $\mathrm{Sp}_n(\mathbb{A})$ with the standard (Rao) cocycle. We view $\mathrm{Sp}_n(F)$ as a subgroup of $\widetilde{\mathrm{Sp}}_n(\mathbb{A})$. For any genuine function φ on $\mathrm{Sp}_n(F)\backslash\widetilde{\mathrm{Sp}}_n(\mathbb{A})$ we consider the Whittaker coefficient

$$\widetilde{\mathcal{W}}(\tilde{\varphi}) = \widetilde{\mathcal{W}}^{\psi_{\widetilde{N}}}(\tilde{\varphi}) := (\mathrm{vol}(N'(F)\backslash N'(\mathbb{A})))^{-1} \int_{N'(F)\backslash N'(\mathbb{A})} \tilde{\varphi}(u)\psi_{\widetilde{N}}(u)^{-1}\, du.$$

Here $\psi_{\widetilde{N}}$ is a nondegenerate character on $N'(\mathbb{A})$ which is trivial on $N'(F)$, where N' is the standard maximal unipotent subgroup of Sp_n. (We view $N'(\mathbb{A})$ as a subgroup of $\widetilde{\mathrm{Sp}}_n(\mathbb{A})$.) We also consider the inner product

$$(\tilde{\varphi}, \tilde{\varphi}^\vee)_{\mathrm{Sp}_n(F)\backslash \mathrm{Sp}_n(\mathbb{A})} = (\mathrm{vol}(\mathrm{Sp}_n(F)\backslash \mathrm{Sp}_n(\mathbb{A})))^{-1} \int_{\mathrm{Sp}_n(F)\backslash \mathrm{Sp}_n(\mathbb{A})} \tilde{\varphi}(g)\tilde{\varphi}^\vee(g)\, dg$$

of two square-integrable genuine functions on $\mathrm{Sp}_n(F)\backslash\widetilde{\mathrm{Sp}}_n(\mathbb{A})$.

Another ingredient in the conjecture of [Lapid and Mao 2015a] is a regularized integral

$$\int^{\mathrm{st}}_{N'(F_S)} f(u)\, du$$

for a finite set of places S and for a suitable class of smooth functions f on $N'(F_S)$. Suffice it to say that if S consists solely of nonarchimedean places then

$$\int^{\mathrm{st}}_{N'(F_S)} f(u)\, du = \int_{N'_1} f(u)\, du$$

for any sufficiently large compact open subgroup N'_1 of $N'(F_S)$. (The definition of the regularized integral is different in the archimedean case, however in this paper we do not use regularized integrals over archimedean fields.)

The conjecture of [Lapid and Mao 2015a] is applicable for $\psi_{\widetilde{N}}$-generic representations which are not exceptional, in the sense that their theta ψ-lift to $\mathrm{SO}(2n+1)$ is cuspidal (or equivalently, their theta ψ-lift to $\mathrm{SO}(2n-1)$ vanishes; here ψ is determined by $\psi_{\widetilde{N}}$). By [Ginzburg et al. 2011, Chapter 11], which is also based on [Cogdell et al. 2001], there is a one-to-one correspondence between these representations and automorphic representations π of $\mathrm{GL}_{2n}(\mathbb{A})$ which are the isobaric sum $\pi_1 \boxplus \cdots \boxplus \pi_k$ of pairwise inequivalent irreducible cuspidal representations π_i of $\mathrm{GL}_{2n_i}(\mathbb{A})$, for $i = 1, \ldots, k$ (with $n_1 + \cdots + n_k = n$), such that $L^S\left(\frac{1}{2}, \pi_i\right) \neq 0$ and $L^S(s, \pi_i, \wedge^2)$ has a pole (necessarily simple) at $s = 1$ for all i. Here $L^S(s, \pi_i)$ and $L^S(s, \pi_i, \wedge^2)$ are the standard and exterior square (partial) L-functions, respectively. More specifically, to any such π one constructs a $\psi_{\widetilde{N}}$-generic representation $\tilde{\pi}$ of $\widetilde{\mathrm{Sp}}_n(\mathbb{A})$, which is called the $\psi_{\widetilde{N}}$-descent of π. The theta ψ-lift of $\tilde{\pi}$ is the unique irreducible generic cuspidal representation of $\mathrm{SO}(2n+1)$ which lifts to π.

Conjecture 1.1 [Lapid and Mao 2015a, Conjecture 1.3]. Assume that $\tilde{\pi}$ is the $\psi_{\widetilde{N}}$-descent of π as above. Then for any $\tilde{\varphi} \in \tilde{\pi}$ and $\tilde{\varphi}^\vee \in \tilde{\pi}^\vee$ and for any sufficiently large finite set S of places of F, we have

$$\widetilde{\mathcal{W}}^{\psi_{\widetilde{N}}}(\tilde{\varphi})\widetilde{\mathcal{W}}^{\psi_{\widetilde{N}}^{-1}}(\tilde{\varphi}^\vee) = 2^{-k}\left(\prod_{i=1}^{n} \zeta_F^S(2i)\right)\frac{L^S\left(\frac{1}{2}, \pi\right)}{L^S(1, \pi, \mathrm{sym}^2)}$$

$$\times (\mathrm{vol}(N'(\mathcal{O}_S)\backslash N'(F_S)))^{-1} \int_{N'(F_S)}^{\mathrm{st}} (\tilde{\pi}(u)\tilde{\varphi}, \tilde{\varphi}^\vee)_{\mathrm{Sp}_n(F)\backslash \mathrm{Sp}_n(\mathbb{A})} \psi_{\widetilde{N}}(u)^{-1}\, du. \quad (1\text{-}1)$$

Here $\zeta_F^S(s)$ is the partial Dedekind zeta function, $\mathcal{O}_S$ is the ring of S-integers of F, and $L^S(s, \pi, \mathrm{sym}^2)$ is the symmetric square partial L-function of π.

The main result in [Lapid and Mao 2017] is the following.

Theorem 1.2 [Lapid and Mao 2017, Theorem 6.2]. *In the above setup we have*

$$\widetilde{\mathcal{W}}^{\psi_{\widetilde{N}}}(\tilde{\varphi})\widetilde{\mathcal{W}}^{\psi_{\widetilde{N}}^{-1}}(\tilde{\varphi}^\vee) = 2^{-k}\left(\prod_{i=1}^{n} \zeta_F^S(2i)\right)\frac{L^S\left(\frac{1}{2}, \pi\right)}{L^S(1, \pi, \mathrm{sym}^2)}\left(\prod_{v\in S} c_{\pi_v}^{-1}\right)$$

$$\times (\mathrm{vol}(N'(\mathcal{O}_S)\backslash N'(F_S)))^{-1} \int_{N'(F_S)}^{\mathrm{st}} (\tilde{\pi}(u)\tilde{\varphi}, \tilde{\varphi}^\vee)_{\mathrm{Sp}_n(F)\backslash \mathrm{Sp}_n(\mathbb{A})} \psi_{\widetilde{N}}(u)^{-1}\, du, \quad (1\text{-}2)$$

where c_{π_v}, $v \in S$ are certain nonzero constants which depend only on the local representations π_v.

The main result of this paper specifies a value for the c_{π_v}.

Theorem 1.3. *In Theorem 1.2 we have $c_{\pi_v} = \epsilon\left(\frac{1}{2}, \pi_v, \psi_v\right)$ (the root number of π_v) for all finite places v.*

We also show in Proposition 8.6 that the root number of π_v equals the central sign of $\tilde{\pi}_v$. In [Ichino et al. 2017] it is shown that Theorem 1.3 implies the formal degree conjecture of Hiraga, Ichino and Ikeda [Hiraga et al. 2008] (or more precisely, its metaplectic analogue) for generic square-integrable representations of $\widetilde{\mathrm{Sp}}_n$. Conversely, in the real case (where the formal degree conjecture is a reformulation of classical results of Harish-Chandra) it is shown that $c_{\pi_v} = \epsilon\left(\frac{1}{2}, \pi_v, \psi_v\right)$ if $\tilde{\pi}_v$ is square-integrable. (Note that in the square-integrable case, matrix coefficients are integrable on $N'(F_v)$ and no regularization is necessary.) We conclude:

Corollary 1.4. *Conjecture 1.1 holds if F is totally real and $\tilde{\pi}_\infty$ is a discrete series.*

Theorem 1.3 is the culmination of the series of papers [Lapid and Mao 2014; 2015b; 2017]. More precisely, the theorem can be formulated as an identity — the main identity (MI) made explicit in Section 3F (based on [Lapid and Mao 2017]) between integrals of Whittaker functions in the induced spaces of π (in a local setting). In principle, formal manipulations using the functional equations of [Lapid and Mao 2014] reduce the identity to the results of [Lapid and Mao 2015b]. Such an argument was described heuristically for the case $n = 1$ in [Lapid and Mao 2017, §7]. However, making this rigorous (even in the case $n = 1$ and for π supercuspidal) seems nontrivial because the integrals only converge as iterated integrals. This is the main task of the present paper.

In Section 3F we reduce the theorem to the cases where π is tempered and satisfies some good properties. Here we rely on the classification, due to Matringe [2015], of the generic representations admitting a nontrivial $\mathrm{GL}_n \times \mathrm{GL}_n$-invariant functional. We also use a globalization result from [Ichino et al. 2017, Appendix A] which is based on a result of [Sakellaridis and Venkatesh 2012].

The rest of the argument is purely local. In Section 5 we start the manipulation of the left-hand side of the main identity. It is technically important to restrict oneself to certain special sections in the induced space. This is possible by a nonvanishing result on the Bessel function of generic representations proved in the Appendix. Another useful idea is to write the left-hand side of the main identity (for W special) as $B\left(W, M\left(\pi, \frac{1}{2}\right) W^\wedge, \frac{1}{2}\right)$, where $B(W, W^\vee, s)$ (for $s \in \mathbb{C}$) is an analytic family of bilinear forms on $I(\pi, s) \times I(\pi^\vee, -s)$, and $M(\pi, s)$ is the intertwining operator on $I(\pi, s)$. This relies on results of Baruch [2005], generalized to the present context in [Lapid and Mao 2013].

The reason for introducing this analytic family is that because of convergence issues, we can only apply the functional equations of [Lapid and Mao 2014] for $\operatorname{Re} s \ll 0$. This is the most delicate step, and is described in Section 7. It entails a further restriction on $W^\wedge$ (which is fortunately harmless for our purpose). The upshot is an expression (for special W and $W^\wedge$ and for $\operatorname{Re} s \ll 0$)

$$B(W, M(\pi, s) W^\wedge, s) = \int E^\psi(M_s^* W, t) E^{\psi^{-1}}(W_s^\wedge, t) \frac{dt}{|\det t|},$$

where t is integrated over a certain n-dimensional torus and E^ψ is a certain integral of W. The restriction on $W^\wedge$ ensures that the integrand is compactly supported. Moreover, by [Lapid and Mao 2015b], as a distribution in t, $E^\psi(W_s, t)$ extends to an entire function on s. Therefore, the above identity is meaningful for all s

where $M(\pi, s)$ is holomorphic. Specializing to $s=\frac{1}{2}$, the remaining assertions are that $E^{\psi}(M^*W, t)$ is constant in t (and in particular, is a function), whose value can be determined, while the integral of $E^{\psi^{-1}}(W^{\wedge}_{1/2}, t)$ factors through $M\big(\pi, \frac{1}{2}\big)W^{\wedge}$, again in an explicit way. This is the content of Corollaries 8.2 and 8.5, respectively, which follow from the representation-theoretic results established in [Lapid and Mao 2015b]. We refer the reader to Section 4 below for a more detailed sketch of the proof.

2. Notation and preliminaries

For the convenience of the reader we introduce in this section the most common notation that will be used throughout.

We fix a positive integer $\mathbf{n}$ (not to be confused with a running variable n). In the following m is either $\mathbf{n}$ or $2\mathbf{n}$. Let F be a local field of characteristic 0.

2A. ***Groups, homomorphisms and group elements.*** All algebraic groups are defined over F. We typically denote algebraic varieties (or groups) over F by boldface letters (e.g., $\mathbf{X}$) and denote their set (or group) of F-points by the corresponding plain letter (e.g., X). (In most cases $\mathbf{X}$ will be clear from the context.)

- I_m is the identity matrix in GL_m, w_m is the $m\times m$ matrix with ones on the nonprincipal diagonal and zeros elsewhere.
- For any group Q, Z_Q is the center of Q; e is the identity element of Q. We denote the modulus function of Q (i.e., the quotient of a right Haar measure by a left Haar measure) by δ_Q.
- Mat_m is the vector space of $m\times m$ matrices over F.
- $x\mapsto x^t$ is the transpose on Mat_m; $x\mapsto \check{x}$ is the twisted transpose map on Mat_m given by $\check{x}=w_m x^t w_m$; $g\mapsto g^*$ is the outer automorphism of GL_m given by $g^*=w_m^{-1}(g^t)^{-1}w_m$.
- $\mathfrak{s}_m=\{x\in\mathrm{Mat}_m : \check{x}=x\}$.
- $\mathbb{M}=\mathrm{GL}_{2\mathbf{n}}$, $\mathbb{M}'=\mathrm{GL}_{\mathbf{n}}$.
- $G=\mathrm{Sp}_{2\mathbf{n}}=\left\{g\in\mathrm{GL}_{4\mathbf{n}} : g^t\begin{pmatrix} & w_{2\mathbf{n}}\\ -w_{2\mathbf{n}} & \end{pmatrix}g=\begin{pmatrix} & w_{2\mathbf{n}}\\ -w_{2\mathbf{n}} & \end{pmatrix}\right\}$.
- $G'=\mathrm{Sp}_{\mathbf{n}}=\left\{g\in\mathrm{GL}_{2\mathbf{n}} : g^t\begin{pmatrix} & w_{\mathbf{n}}\\ -w_{\mathbf{n}} & \end{pmatrix}g=\begin{pmatrix} & w_{\mathbf{n}}\\ -w_{\mathbf{n}} & \end{pmatrix}\right\}$.
- G' is embedded as a subgroup of G via $g\mapsto\eta(g)=\mathrm{diag}(I_{\mathbf{n}}, g, I_{\mathbf{n}})$.
- $P=M\ltimes U$ and $P'=M'\ltimes U'$ are the Siegel parabolic subgroups of G and G', respectively, with the standard Levi decomposition.
- $\overline{P}=P^t$ is the opposite parabolic of P, with unipotent radical $\overline{U}=U^t$.
- We use the isomorphism $\varrho(g)=\mathrm{diag}(g, g^*)$ to identify $\mathbb{M}$ with $M\subset G$. Similarly for $\varrho' : \mathbb{M}'\to M'\subset G'$.

• We use the embeddings $\eta_{\mathbb{M}}(g) = \operatorname{diag}(g, I_{\mathbf{n}})$ and $\eta_{\mathbb{M}}^{\vee}(g) = \operatorname{diag}(I_{\mathbf{n}}, g)$ to identify $\mathbb{M}'$ with subgroups of $\mathbb{M}$. We also set $\eta_M = \varrho \circ \eta_{\mathbb{M}}$ and $\eta_M^{\vee} = \varrho \circ \eta_{\mathbb{M}}^{\vee} = \eta \circ \varrho'$.

• K is the standard maximal compact subgroup of G. (In the p-adic case it consists of the matrices with integral entries.)

• N is the standard maximal unipotent subgroup of G consisting of upper unitriangular matrices; T is the maximal torus of G consisting of diagonal matrices; $B = T \ltimes N$ is the Borel subgroup of G.

• For any subgroup X of G, we set $X' = \eta^{-1}(X)$, $X_M = X \cap M$ and $X_{\mathbb{M}} = \varrho^{-1}(X_M)$; similarly, $X'_{M'} = X' \cap M'$ and $X'_{\mathbb{M}'} = \varrho'^{-1}(X'_{M'})$.

• $\ell_{\mathbb{M}} : \mathrm{Mat}_{\mathbf{n}} \to N_{\mathbb{M}}$ is the group embedding given by $\ell_{\mathbb{M}}(x) = \begin{pmatrix} I_{\mathbf{n}} & x \\ & I_{\mathbf{n}} \end{pmatrix}$ and $\ell_M = \varrho \circ \ell_{\mathbb{M}}$.

• $\ell : \mathfrak{s}_{2\mathbf{n}} \to U$ is the isomorphism given by $\ell(x) = \begin{pmatrix} I_{2\mathbf{n}} & x \\ & I_{2\mathbf{n}} \end{pmatrix}$.

• $\widetilde{G} = \widetilde{\mathrm{Sp}}_{\mathbf{n}}$ is the metaplectic group, i.e., the nontrivial two-fold cover of G' (unless F is complex). We write elements of $\widetilde{G}$ as pairs (g, ϵ), for $g \in G$ and $\epsilon = \pm 1$, with the multiplication given by Rao's cocycle. See [Ginzburg et al. 2011].

• When $g \in G'$, we write $\tilde{g} = (g, 1) \in \widetilde{G}$. Of course, $g \mapsto \tilde{g}$ is not a group homomorphism, but we do have $\widetilde{gn} = \tilde{g}\tilde{n}$ and $\widetilde{ng} = \tilde{n}\tilde{g}$ for any $g \in G'$ and $n \in N'$.

• $\widetilde{N}$ and $\widetilde{P}$ are the inverse images of N' and P' under the canonical projection $\widetilde{G} \to G'$. We identify N' with a subgroup of $\widetilde{N}$ via $n \mapsto \tilde{n}$.

• $\xi_m = (0, \dots, 0, 1) \in F^m$.

• $\mathcal{P}$ is the mirabolic subgroup of $\mathbb{M}$ consisting of the elements g such that $\xi_{2\mathbf{n}} g = \xi_{2\mathbf{n}}$.

• $E = \operatorname{diag}(1, -1, \dots, 1, -1) \in \mathbb{M}$. H is the centralizer of $\varrho(E)$ in G. It is isomorphic to $\mathrm{Sp}_{\mathbf{n}} \times \mathrm{Sp}_{\mathbf{n}}$.

• $H_{\mathbb{M}}$ is then the centralizer of E in $\mathbb{M}$. It is isomorphic to $\mathrm{GL}_{\mathbf{n}} \times \mathrm{GL}_{\mathbf{n}}$.

• $w'_0 = \begin{pmatrix} & w_{\mathbf{n}} \\ -w_{\mathbf{n}} & \end{pmatrix} \in G'$ represents the longest Weyl element of G'.

• $w_U = \begin{pmatrix} & I_{2\mathbf{n}} \\ -I_{2\mathbf{n}} & \end{pmatrix} \in G$ represents the longest M-reduced Weyl element of G.

• $w'_{U'} = \begin{pmatrix} & I_{\mathbf{n}} \\ -I_{\mathbf{n}} & \end{pmatrix} \in G'$ represents the longest M'-reduced Weyl element of G'.

• $w_0^{\mathbb{M}} = w_{2\mathbf{n}} \in \mathbb{M}$ represents the longest Weyl element of $\mathbb{M}$; $w_0^M = \varrho(w_0^{\mathbb{M}})$.

• $w_0^{\mathbb{M}'} = w_{\mathbf{n}} \in \mathbb{M}'$ represents the longest Weyl element of $\mathbb{M}'$; $w_0^{M'} = \varrho'(w_0^{\mathbb{M}'})$.

• $w_{2\mathbf{n},\mathbf{n}} = \begin{pmatrix} & I_{\mathbf{n}} \\ I_{\mathbf{n}} & \end{pmatrix} \in \mathbb{M}$, $w'_{2\mathbf{n},\mathbf{n}} = \begin{pmatrix} & I_{\mathbf{n}} \\ w_0^{\mathbb{M}'} & \end{pmatrix} \in \mathbb{M}$.

• $\gamma = w_U \eta(w'_{U'})^{-1} = \begin{pmatrix} & I_{\mathbf{n}} & & \\ & & & I_{\mathbf{n}} \\ -I_{\mathbf{n}} & & & \\ & & I_{\mathbf{n}} & \end{pmatrix} \in G$.

• $\mathfrak{d} = \operatorname{diag}(1, -1, \dots, (-1)^{\mathbf{n}-1}) \in \mathrm{Mat}_{\mathbf{n}}$, $\boldsymbol{\epsilon}_1 = \ell_M((-1)^{\mathbf{n}}\mathfrak{d}) \in N_M$, $\boldsymbol{\epsilon}_2 = \ell_{\mathbb{M}}(\mathfrak{d}) \in N_{\mathbb{M}}$, $\boldsymbol{\epsilon}_3 = w'_{2\mathbf{n},\mathbf{n}} \boldsymbol{\epsilon}_2 \in \mathbb{M}$, $\boldsymbol{\epsilon}_4 = \ell_{\mathbb{M}}\big(-\frac{1}{2}\mathfrak{d} w_0^{\mathbb{M}'}\big) \in N_{\mathbb{M}}$.

• V and $V^\sharp$ are the unipotent radicals of the standard parabolic subgroup of G with Levi $\mathrm{GL}_1^{\mathbf{n}} \times \mathrm{Sp}_{\mathbf{n}}$ and $\mathrm{GL}_1^{\mathbf{n}-1} \times \mathrm{Sp}_{\mathbf{n}+1}$, respectively. Thus, $N = \eta(N') \ltimes V$, $V^\sharp$ is normal in V and $V/V^\sharp$ is isomorphic to the Heisenberg group of dimension $2\mathbf{n}+1$ (see below). Also, $V = V_M \ltimes V_U$, where $V_U = V \cap U = \left\{\ell\left(\left(\begin{smallmatrix} x & y \\ & \check{x} \end{smallmatrix}\right)\right) : x \in \mathrm{Mat}_{\mathbf{n}},\, y \in \mathfrak{s}_{\mathbf{n}}\right\}$.

• $V_- = V_M^\sharp \ltimes V_U$. (Recall $V_M^\sharp = V^\sharp \cap M$ by our convention.)

• $V_\gamma = V \cap \gamma^{-1} N \gamma = \eta(w'_{U'}) V_M \eta(w'_{U'})^{-1} = \eta_M(N'_{\mathbb{M}'}) \ltimes \left\{\ell\left(\left(\begin{smallmatrix} x & \\ & \check{x} \end{smallmatrix}\right)\right) : x \in \mathrm{Mat}_{\mathbf{n}}\right\} \subset V_-$.

• $V_+ \subset V$ is the image under ℓ_M of the space of $\mathbf{n} \times \mathbf{n}$ matrices whose rows are zero except possibly for the last one. Thus, $V = V_+ \ltimes V_-$. For $c = \ell_M(x) \in V_+$ we denote by $\underline{c} \in F^{\mathbf{n}}$ the last row of x.

• $N^\sharp = V_- \rtimes \eta(N')$ is the stabilizer in N of the character ψ_U defined below.

• $N_{\mathbb{M}}^\flat = (N_{\mathbb{M}}^\sharp)^*$.

• J is the subspace of $\mathrm{Mat}_{\mathbf{n}}$ consisting of the matrices whose first column is zero.

• $\bar{R} = \left\{\left(\begin{smallmatrix} I_{\mathbf{n}} & \\ x & n^t \end{smallmatrix}\right) : x \in J,\, n \in N'_{\mathbb{M}'}\right\}$.

• $T'' = Z_{\mathbb{M}} \times \eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}) = \{\mathrm{diag}(t_1, \ldots, t_{2\mathbf{n}}) : t_1 = \cdots = t_{\mathbf{n}}\} \subset T_{\mathbb{M}}$.

2B. ***Characters.*** We fix a nontrivial unitary character ψ of F. For each of the unipotent groups X listed below we assign a character ψ_X of X as follows:

$$\psi_{N_{\mathbb{M}}}(u) = \psi(u_{1,2} + \cdots + u_{2\mathbf{n}-1,2\mathbf{n}}) \quad \text{(nondegenerate)},$$

$$\psi_{N_M} \circ \varrho = \psi_{N_{\mathbb{M}}},$$

$$\psi_{N'_{\mathbb{M}'}}(u') = \psi(u'_{1,2} + \cdots + u'_{\mathbf{n}-1,\mathbf{n}}) \quad \text{(nondegenerate)},$$

$$\psi_{N'_{M'}} \circ \varrho' = \psi_{N'_{\mathbb{M}'}} \quad (\text{i.e., } \psi_{N'_{M'}}(n) = \psi_{N_M}(\gamma \eta(n) \gamma^{-1})),$$

$$\psi_{N'}(nu) = \psi_{N'_{M'}}(n) \psi\left(\tfrac{1}{2} u_{\mathbf{n},\mathbf{n}+1}\right)^{-1}, \quad n \in N'_{M'},\ u \in U',$$

$$\psi_{\widetilde{N}} \text{ is the extension of } \psi_{N'} \text{ to a genuine character of } \widetilde{N},$$

$$\psi_N(nu) = \psi_{N_M}(n), \quad n \in N_M,\ u \in U, \quad \text{(degenerate)},$$

$$\psi_U(\ell(v)) = \psi\left(\tfrac{1}{2}(v_{\mathbf{n},\mathbf{n}+1} - v_{2\mathbf{n},1})\right),$$

$$\psi_{V_-}(vu) = \psi_{N_M}(v)^{-1} \psi_U(u), \quad v \in V_M^\sharp,\ u \in V_U,$$

$$\psi_{N^\sharp}(\eta(n)v) = \psi_{N'}(n) \psi_{V_-}(v), \quad n \in N',\ v \in V_-,$$

$$\psi_{N_{\mathbb{M}}^\sharp} = \psi_{N^\sharp} \circ \varrho,$$

$$\psi_{N_{\mathbb{M}}^\flat}(m) = \psi_{N_{\mathbb{M}}^\sharp}(m^*), \quad m \in N_{\mathbb{M}}^\flat,$$

$$\psi_{\overline{U}}(\ell(v)^t) = \psi(v_{1,1}), \quad v \in \mathfrak{s}_{2\mathbf{n}}.$$

2C. ***Other notation.*** • We use the notation $a \ll_d b$ to mean that $a \le cb$ with $c > 0$ a constant depending on d.

- For any $g \in G$, define $\nu(g) \in \mathbb{R}_{>0}$ by $\nu(u\varrho(m)k) = |\det m|$ for any $u \in U$, $m \in \mathbb{M}$, $k \in K$. Let $\nu'(g) = \nu(\eta(g))$ for $g \in G'$.

- $\mathcal{CSGR}(Q)$ is the set of compact open subgroups of a topological group Q.

- $\mathcal{CSGR}^s(G')$ is the subset of $\mathcal{CSGR}(G')$ consisting of the K_0 for which $k \in K_0 \mapsto \tilde{k}$ is a group isomorphism. Any sufficiently small $K_0 \in \mathcal{CSGR}(G')$ belongs to $\mathcal{CSGR}^s(G')$. We identify any $K_0 \in \mathcal{CSGR}^s(G')$ with its image under $k \mapsto \tilde{k}$ (an element of $\mathcal{CSGR}(\widetilde{G})$).

- For an ℓ-group Q, let $C(Q)$ be the space of continuous functions on Q, and $\mathcal{S}(Q)$ the space of Schwartz functions on Q.

- When F is p-adic, if Q' is a closed subgroup of Q and χ is a character of Q', we denote by $C(Q'\backslash Q, \chi)$ (resp. $C^{\mathrm{sm}}(Q'\backslash Q, \chi)$, $C_c^\infty(Q'\backslash Q, \chi)$) the spaces of continuous (resp. Q-smooth,[1] smooth and compactly supported modulo Q') complex-valued left (Q', χ)-equivariant functions on Q.

- For an ℓ-group Q, we write $\operatorname{Irr} Q$ for the set of equivalence classes of irreducible representations of Q. If Q is reductive, we also write $\operatorname{Irr}_{\mathrm{sqr}} Q$ and $\operatorname{Irr}_{\mathrm{temp}} Q$ for the subsets of irreducible unitary square-integrable (modulo center) and tempered representations, respectively. We write $\operatorname{Irr}_{\mathrm{gen}} \mathbb{M}$ and $\operatorname{Irr}_{\mathrm{meta}} \mathbb{M}$ for the subsets of irreducible generic representations of $\mathbb{M}$ and representations of metaplectic type (see below), respectively. For the set of irreducible generic representations of $\widetilde{G}$, we use the notation $\operatorname{Irr}_{\mathrm{gen},\psi_{\tilde{N}}} \widetilde{G}$ to emphasize the dependence on the character $\psi_{\tilde{N}}$.

- For $\pi \in \operatorname{Irr} Q$, let $\pi^\vee$ be the contragredient of π.

- For $\pi \in \operatorname{Irr}_{\mathrm{gen}} \mathbb{M}$, we denote by $\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)$ the (uniquely determined) Whittaker space of π with respect to the character $\psi_{N_{\mathbb{M}}}$. We use the notation $\mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}$, $\mathbb{W}^{\psi_{N_M}}$, $\mathbb{W}^{\psi_{N_M}^{-1}}$, $\mathbb{W}^{\psi_{\tilde{N}}}$, $\mathbb{W}^{\psi_{\tilde{N}}^{-1}}$ similarly.

- For $\pi \in \operatorname{Irr}_{\mathrm{gen}} M$, let $\operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ be the space of smooth left U-invariant functions $W : G \to \mathbb{C}$ such that for all $g \in G$, the function $m \mapsto \delta_P(m)^{-1/2} W(mg)$ on M belongs to $\mathbb{W}^{\psi_{N_M}}(\pi)$. Similarly define $\operatorname{Ind}(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi))$.

- If a group G_0 acts on a vector space W and H_0 is a subgroup of G_0, we denote by W^{H_0} the subspace of H_0-fixed points.

- We use the following bracket notation for iterated integrals: $\iint\big(\iint \dots\big)\dots$ implies that the inner integrals converge as a double integral and after evaluating them, the outer double integral is absolutely convergent.

2D. ***Measures.*** We take the self-dual Haar measure on F with respect to ψ. We use the following convention for Haar measures for algebraic subgroups of G. (We consider G', $\mathbb{M}$, $\mathbb{M}'$ as subgroups of G through the embeddings η, ϱ and $\eta \circ \varrho'$.)

[1]That is, right-invariant under an open subgroup of Q.

When F is a p-adic field, let $\mathcal{O}$ be its ring of integers. The Lie algebra $\mathfrak{M}$ of $\mathrm{GL}_{4\mathbf{n}}$ consists of the $4\mathbf{n} \times 4\mathbf{n}$ matrices X over F. Let $\mathfrak{M}_{\mathcal{O}}$ be the lattice of integral matrices in $\mathfrak{M}$. For any algebraic subgroup $\mathbf{Q}$ of $\mathrm{GL}_{4\mathbf{n}}$ defined over F (e.g., an algebraic subgroup of $\mathbf{G}$), let $\mathfrak{q} \subset \mathfrak{M}$ be the Lie algebra of $\mathbf{Q}$. The lattice $\mathfrak{q} \cap \mathfrak{M}_{\mathcal{O}}$ of $\mathfrak{q}$ gives rise to a gauge form of $\mathbf{Q}$ (determined up to multiplication by an element of $\mathcal{O}^*$) and we use it to define a Haar measure on Q by the recipe of [Kneser 1967].

When $F = \mathbb{R}$, the measures are fixed similarly, except that we use $\mathfrak{M}_{\mathbb{Z}}$ in place of $\mathfrak{M}_{\mathcal{O}}$. When $F = \mathbb{C}$ we only consider subgroups of G which are defined over $\mathbb{R}$ and take the gauge forms induced from the above recipe for $\mathbb{R}$.

2E. *Weil representation.* Let $\mathbb{V}$ be a symplectic space over F with a symplectic form $\langle \cdot , \cdot \rangle$. Let $\mathcal{H} = \mathcal{H}_{\mathbb{V}}$ be the Heisenberg group of $(\mathbb{V}, \langle \cdot , \cdot \rangle)$. Recall that $\mathcal{H}_{\mathbb{V}} = \mathbb{V} \oplus F$ with the product rule

$$(x, t) \cdot (y, z) = \bigl(x + y, t + z + \tfrac{1}{2}\langle x, y\rangle\bigr).$$

Fix a polarization $\mathbb{V} = \mathbb{V}_+ \oplus \mathbb{V}_-$. The group $\mathrm{Sp}(\mathbb{V})$ acts on the right on $\mathbb{V}$. We write a typical element of $\mathrm{Sp}(\mathbb{V})$ as $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where $A \in \mathrm{Hom}(\mathbb{V}_+, \mathbb{V}_+)$, $B \in \mathrm{Hom}(\mathbb{V}_+, \mathbb{V}_-)$, $C \in \mathrm{Hom}(\mathbb{V}_-, \mathbb{V}_+)$ and $D \in \mathrm{Hom}(\mathbb{V}_-, \mathbb{V}_-)$. Let $\widetilde{\mathrm{Sp}}(\mathbb{V})$ be the metaplectic two-fold cover of $\mathrm{Sp}(\mathbb{V})$ with respect to the Rao cocycle determined by the splitting. Consider the Weil representation ω_ψ of the group $\mathcal{H}_{\mathbb{V}} \rtimes \widetilde{\mathrm{Sp}}(\mathbb{V})$ on $\mathcal{S}(\mathbb{V}_+)$. Explicitly, for any $\Phi \in \mathcal{S}(\mathbb{V}_+)$ and $X \in \mathbb{V}_+$, the action of $\mathcal{H}_{\mathbb{V}}$ is given by

$$\omega_\psi(a, 0)\Phi(X) = \Phi(X + a), \qquad a \in \mathbb{V}_+, \tag{2-1a}$$

$$\omega_\psi(b, 0)\Phi(X) = \psi(\langle X, b\rangle)\Phi(X), \quad b \in \mathbb{V}_-, \tag{2-1b}$$

$$\omega_\psi(0, t)\Phi(X) = \psi(t)\Phi(X), \qquad t \in F, \tag{2-1c}$$

while the action of $\widetilde{\mathrm{Sp}}(\mathbb{V})$ is (partially) given by

$$\omega_\psi\bigl(\left(\begin{smallmatrix} g & 0 \\ 0 & g^* \end{smallmatrix}\right), \epsilon\bigr)\Phi(X) = \epsilon\gamma_\psi(\det g)|\det g|^{1/2}\Phi(Xg), \quad g \in \mathrm{GL}(\mathbb{V}_+), \tag{2-2a}$$

$$\omega_\psi\bigl(\left(\begin{smallmatrix} I & B \\ 0 & I \end{smallmatrix}\right), \epsilon\bigr)\Phi(X) = \epsilon\psi\bigl(\tfrac{1}{2}\langle X, XB\rangle\bigr)\Phi(X), \qquad B \in \mathrm{Hom}(\mathbb{V}_+, \mathbb{V}_-) \text{ self-dual}, \tag{2-2b}$$

where γ_ψ is Weil's factor.

We now take $\mathbb{V} = F^{2\mathbf{n}}$ with the standard symplectic form

$$\langle (x_1, \dots, x_{2\mathbf{n}}), (y_1, \dots, y_{2\mathbf{n}})\rangle = \sum_{i=1}^{\mathbf{n}} x_i y_{2\mathbf{n}+1-i} - \sum_{i=1}^{\mathbf{n}} y_i x_{2\mathbf{n}+1-i}$$

and the standard polarization

$$\mathbb{V}_+ = \{(x_1, \dots, x_{\mathbf{n}}, 0, \dots, 0)\}, \qquad \mathbb{V}_- = \{(0, \dots, 0, y_1, \dots, y_{\mathbf{n}})\}.$$

(We identify $\mathbb{V}_+$ and $\mathbb{V}_-$ with $F^{\mathbf{n}}$.) The corresponding Heisenberg group is isomorphic to the quotient $V/V^\sharp$ (with V, $V^\sharp$ as defined in Section 2A) via $v \mapsto v_{\mathcal{H}} := \big((v_{\mathbf{n},\mathbf{n}+j})_{j=1,\dots,2\mathbf{n}}, \tfrac{1}{2}v_{\mathbf{n},3\mathbf{n}+1}\big)$.

For $X = (x_1, \dots, x_{\mathbf{n}})$, $X' = (x'_1, \dots, x'_{\mathbf{n}}) \in F^{\mathbf{n}}$, define

$$\langle X, X'\rangle' = x_1 x'_{\mathbf{n}} + \cdots + x_{\mathbf{n}} x'_1.$$

For $\Phi \in \mathcal{S}(F^{\mathbf{n}})$, define the Fourier transform

$$\hat{\Phi}(X) = \int_{F^{\mathbf{n}}} \Phi(X')\psi(\langle X, X'\rangle')\, dX'.$$

Then, realized on $\mathcal{S}(F^{\mathbf{n}})$, the Weil representation satisfies

$$\omega_\psi(\widetilde{\varrho'(h)})\Phi(X) = |\det(h)|^{1/2}\beta_\psi(\varrho'(h))\Phi(Xh), \quad h \in \mathbb{M}', \tag{2-3a}$$

$$\omega_\psi\big(\widetilde{\left(\begin{smallmatrix} 1 & B \\ & 1 \end{smallmatrix}\right)}\big)\Phi(X) = \psi\big(\tfrac{1}{2}\langle X, XB\rangle'\big)\Phi(X), \quad B \in \mathfrak{s}_{\mathbf{n}}, \tag{2-3b}$$

$$\omega_\psi(\widetilde{w'_{U'}})\Phi(X) = \beta_\psi(w'_{U'})\hat{\Phi}(X). \tag{2-3c}$$

Here $\beta_\psi(g)$ for $g \in G'$ is a certain root of unity; moreover, $\beta_\psi(\varrho'(h)) = \gamma_\psi(\det h)$.

We extend ω_ψ to $V \rtimes \widetilde{G}$ by setting

$$\omega_\psi(v\tilde{g})\Phi = \psi(v_{1,2} + \cdots + v_{\mathbf{n}-1,\mathbf{n}})^{-1}\omega_\psi(v_{\mathcal{H}})(\omega_\psi(\tilde{g})\Phi), \quad v \in V,\ g \in G'. \tag{2-4}$$

Then for any $g \in G'$, $v \in V$ we have

$$\omega_\psi((\eta(g)v\eta(g)^{-1})\tilde{g})\Phi = \omega_\psi(\tilde{g})(\omega_\psi(v)\Phi). \tag{2-5}$$

2F. *Stable integral.* For the rest of the section, we assume F is p-adic.

Suppose that U_0 is a unipotent group over F with a fixed Haar measure du. Recall that the group generated by a relatively compact subset of U_0 is relatively compact. In particular, the set $\mathcal{CSGR}(U_0)$ is directed. Recall the following definition of stable integral from [Lapid and Mao 2015a].

Definition 2.1. Let f be a smooth function on U_0. We say that f has a *stable integral* over U_0 if there exists $U_1 \in \mathcal{CSGR}(U_0)$ such that for any $U_2 \in \mathcal{CSGR}(U_0)$ containing U_1, we have

$$\int_{U_2} f(u)\, du = \int_{U_1} f(u)\, du. \tag{2-6}$$

In this case, we write $\int_{U_0}^{\mathrm{st}} f(u)\, du$ for the common value of (2-6) and say that $\int_{U_0}^{\mathrm{st}} f(u)\, du$ stabilizes at U_1. In other words, $\int_{U_0}^{\mathrm{st}} f(u)\, du$ is the limit of the net

$$\left(\int_{U_1} f(u)\, du\right)_{U_1 \in \mathcal{CSGR}(U_0)}$$

with respect to the discrete topology of $\mathbb{C}$.

Given a family of functions $f_x \in C^{\mathrm{sm}}(U_0)$, we say that the integral $\int^{\mathrm{st}}_{U_0} f_x(u)\,du$ *stabilizes uniformly in* x if U_1 as above can be chosen independently of x. Similarly, if x ranges over a topological space X, then we say that $\int^{\mathrm{st}}_{U_0} f_x(u)\,du$ stabilizes locally uniformly in x if any $y \in X$ admits a neighborhood on which $\int^{\mathrm{st}}_{U_0} f_x(u)\,du$ stabilizes uniformly.

2G. ***A remark on convergence.*** We frequently make use of the following elementary remark.

Remark 2.2. Let $\mathbf{H}$ be any algebraic group over F and $\mathbf{H}'$ a closed subgroup. Assume that $\delta_H|_{H'} \equiv \delta_{H'}$. Suppose that $f \in C^{\mathrm{sm}}(H)$ and that the integral $\int_H f(h)\,dh$ converges absolutely. Then the same is true for $\int_{H'} f(h')\,dh'$.

3. Statement of main result

3A. ***Local Fourier–Jacobi transform.*** For any $f \in C(G)$ and $s \in \mathbb{C}$, we define $f_s(g) = f(g)\nu(g)^s$ for $g \in G$. Let $\pi \in \mathrm{Irr}_{\mathrm{gen}}\, M$ with Whittaker model $\mathbb{W}^{\psi_{N_M}}(\pi)$. Let $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ be the space of G-smooth left U-invariant functions $W : G \to \mathbb{C}$ such that for all $g \in G$, the function $\delta_P(m)^{-1/2} W(mg)$ on M belongs to $\mathbb{W}^{\psi_{N_M}}(\pi)$. For any $s \in \mathbb{C}$ we have a representation $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi), s)$ on the space $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ given by $(I(s, g)W)_s(x) = W_s(xg)$ for $x, g \in G$. It is equivalent to the induced representation of $\pi \otimes \nu^s$ from P to G. The family W_s, $s \in \mathbb{C}$ is a holomorphic section of this family of induced representations.

Let F be a p-adic field. As in [Ginzburg et al. 1998], for any $W \in C^{\mathrm{sm}}(N\backslash G, \psi_N)$ and $\Phi \in \mathcal{S}(F^{\mathbf{n}})$ define a genuine function on $\widetilde{G}$ by

$$A^{\psi}(W, \Phi, \tilde{g}) = \int_{V_\gamma\backslash V} W(\gamma v \eta(g))\omega_{\psi^{-1}}(v\tilde{g})\Phi(\xi_{\mathbf{n}})\,dv, \quad g \in G', \tag{3-1}$$

where the element $\gamma \in G$ and the groups V and V_γ are as defined in Section 2A. Its properties were studied in [Lapid and Mao 2017, §4]. In particular, the integrand is always compactly supported and A^{ψ} gives rise to a $V \rtimes \widetilde{G}$-intertwining map

$$A^{\psi} : C^{\mathrm{sm}}(N\backslash G, \psi_N) \otimes \mathcal{S}(F^{\mathbf{n}}) \to C^{\mathrm{sm}}(N'\backslash \widetilde{G}, \psi_{\widetilde{N}}), \tag{3-2}$$

where $V \rtimes \widetilde{G}$ acts via $V \rtimes \eta(G')$ by right translation on $C^{\mathrm{sm}}(N\backslash G, \psi_N)$, through $\omega_{\psi^{-1}}$ on $\mathcal{S}(F^{\mathbf{n}})$ and via the projection to $\widetilde{G}$ by right translation on $C^{\mathrm{sm}}(N'\backslash \widetilde{G}, \psi_{\widetilde{N}})$.

Let $V_+ \subset V$ be the image under ℓ_M of the space of $\mathbf{n} \times \mathbf{n}$ matrices whose rows are zero except possibly for the last one. For $c = \ell_M(x) \in V_+$, we denote by $\underline{c} \in F^{\mathbf{n}}$ the last row of x. Then

$$A^{\psi}(W(\cdot\, c), \Phi(\cdot + \underline{c}), \tilde{g}) = A^{\psi}(W, \Phi, \tilde{g}), \quad c \in V_+,\ g \in G'.$$

It follows that the function $A^\psi(W,\Phi,\cdot)$ factors through $W\otimes\Phi\mapsto\Phi*W$, where for any function $f\in C^\infty(G)$ we set

$$\Phi*f(g)=\int_{V_+}f(gc)\Phi(\underline{c})\,dc.$$

We denote by $A^\psi_\sharp$ the map

$$A^\psi_\sharp:C^{\mathrm{sm}}(N\backslash G,\psi_N)\to C^{\mathrm{sm}}(\widetilde{N}\backslash\tilde{G},\psi_{\widetilde{N}})$$

such that

$$A^\psi(W,\Phi,\cdot)=A^\psi_\sharp(\Phi*W,\cdot).$$

(Informally, $A^\psi_\sharp(W,\cdot)=A^\psi(W,\delta_0,\cdot)$, where δ_0 is the delta function at 0.) The map $A^\psi_\sharp$ is no longer $\widetilde{G}$ or V-equivariant since neither $\widetilde{G}$ nor V (acting through $\omega_{\psi^{-1}}$) stabilizes δ_0. However, $A^\psi_\sharp$ satisfies the following equivariance property. Let $V_-=V_U\rtimes V_M^\sharp$, where $V_U=V\cap U$ and $V_M^\sharp\subset V_M$ is defined in Section 2A. Thus, V_- is the preimage of $\mathbb{V}_-$ under the composition $V\to V/V^\sharp\simeq\mathcal{H}\to\mathbb{V}$ and we have $V=V_+\ltimes V_-$. Note that V_- is normalized by P'. Let ψ_{V_-} be the character on V_- given by

$$\psi_{V_-}(vu)=\psi_{N_M}(v)^{-1}\psi_U(u),\quad v\in V_M^\sharp,\ u\in V_U.$$

Lemma 3.1. *For any $v\in V_-$ and $p=mu\in P'$ where $m\in M'$ and $u\in U'$, we have*

$$A^\psi_\sharp(W(\cdot\,v\eta(p)),\tilde{g})=\nu'(m)^{1/2}\beta_{\psi^{-1}}(m)^{-1}\psi_{V_-}(v)A^\psi_\sharp(W,\tilde{g}\tilde{p}).$$

Proof. The statement simply reflects the fact that $V_-\rtimes\widetilde{P}$ acts on δ_0 by multiplication by the character

$$\widetilde{m'}\widetilde{u'}v\mapsto\nu'(m')^{-1/2}\beta_{\psi^{-1}}(m')\psi_{V_-}^{-1}(v),\quad v\in V_-,\ m'\in M',\ u'\in U'.$$

More rigorously, fix W, p and v and let C be a small neighborhood of 0 in $F^{\mathbf{n}}$. Suppose that $\Phi\in C_c^\infty(F^{\mathbf{n}})$ is supported in C and that $\int_{F^{\mathbf{n}}}\Phi(c)\,dc=1$. Then $A^\psi(W,\Phi,\tilde{g})=A^\psi_\sharp(W,\tilde{g})$ for all $g\in G'$ provided that C is sufficiently small. On the other hand, if $\Phi'=\omega_{\psi^{-1}}(v\tilde{p})\Phi$ then

$$A^\psi(W(\cdot\,v\eta(p)),\Phi',\tilde{g})=A^\psi(W,\Phi,\tilde{g}\tilde{p}).$$

By (2-1b), (2-1c), (2-3a), (2-3b) and (2-4), $\operatorname{supp}\Phi'\subset Cm^{-1}$ and, when C is sufficiently small,

$$\int_{F^{\mathbf{n}}}\Phi'(c)\,dc=\psi_{V_-}(v)^{-1}\beta_{\psi^{-1}}(m)\nu'(m)^{-1/2}.$$

Thus, if C is small enough then

$$\Phi'*W(\cdot\,vp)=\psi_{V_-}(v)^{-1}\beta_{\psi^{-1}}(m)\nu'(m)^{-1/2}W(\cdot\,vp).$$

Hence, $A^{\psi}(W(\cdot\, v\eta(p)), \Phi', \tilde g) = \psi_{V_-}(v)^{-1}\beta_{\psi^{-1}}(m)\nu'(m)^{-1/2}A^{\psi}_{\sharp}(W(\cdot\, v\eta(p)), \tilde g)$. The lemma follows. $\square$

For later reference we record the following result, which is implicit in the proof of [Lapid and Mao 2017, Lemma 4.5].

Lemma 3.2. *Suppose that F is p-adic. Then for any $K_0 \in \mathcal{CSGR}(G)$, there exists $\Omega \in \mathcal{CSGR}(V_U)$ such that for any $W \in C(N\backslash G, \psi_N)^{K_0}$, the support of $W(\gamma\,\cdot)|_{V_-}$ is contained in $V_\gamma\eta(w'_{U'})\Omega\eta(w'_{U'})^{-1}$.*

Remark 3.3. When F is an archimedean field, the above discussion still holds for $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ (or more generally if $W \in C^{\mathrm{sm}}(N\backslash G, \psi_N)$ is of moderate growth). From [Lapid and Mao 2017, Lemma 4.9], the integral (3-1) converges. Moreover, $A^{\psi}(W, \Phi, \cdot)$ factors through $W \otimes \Phi \mapsto \Phi * W$, and the induced linear form $A^{\psi}_{\sharp}$ satisfies the equivariance property in Lemma 3.1 (by an approximate identity argument).

3B. ***Explicit local descent.*** Let F be a local field of characteristic 0. Define the intertwining operator

$$M(\pi, s) = M(s) : \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi), s) \to \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi^\vee), -s)$$

by (the analytic continuation of)

$$M(s)W(g) = \nu(g)^s \int_U W_s(\varrho(\mathfrak{t})w_U ug)\, du, \tag{3-3}$$

where $\mathfrak{t} = \mathrm{diag}(1, -1, \dots, 1, -1)$ is introduced in order to preserve the character ψ_{N_M}. By abuse of notation we also denote by $M(\pi, s)$ the intertwining operator $\mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi), s\big) \to \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee), -s\big)$ defined in the same way.

For simplicity, we define $M^*_s W := (M(s)W)_{-s}$, so that

$$M^*_s W = \int_U W_s(\varrho(\mathfrak{t})w_U u\,\cdot)\, du$$

for $\mathrm{Re}\, s \gg_\pi 1$. Set $M^* W := M^*_{1/2} W$.

Recall that $H_{\mathbb{M}}$ is the centralizer of $E = \mathrm{diag}(1, -1, \dots, 1, -1)(= \mathfrak{t})$, isomorphic to $\mathrm{GL}_{\mathbf{n}} \times \mathrm{GL}_{\mathbf{n}}$. The involution $w_0^{\mathbb{M}}$ lies in the normalizer of $H_{\mathbb{M}}$. We consider the class $\mathrm{Irr}_{\mathrm{meta}}\, \mathbb{M}$ of irreducible representations of $\mathbb{M}$ which admit a continuous nonzero $H_{\mathbb{M}}$-invariant linear form ℓ on the space of π. It is known that any such π is self-dual and ℓ is unique up to a scalar [Jacquet and Rallis 1996; Aizenbud and Gourevitch 2009]. Thus, $\ell \circ \pi(w_0^{\mathbb{M}}) = \epsilon_\pi \ell$, where $\epsilon_\pi \in \{\pm 1\}$ does not depend on the choice of ℓ. By [Lapid and Mao 2017, Theorem 3.2], when F is p-adic, for any $\pi \in \mathrm{Irr}_{\mathrm{meta,gen}}\, \mathbb{M}$ we have

$$\epsilon_\pi = \epsilon\big(\tfrac{1}{2}, \pi, \psi\big), \tag{3-4}$$

where $\epsilon(s, \pi, \psi)$ is the standard ϵ-factor attached to π.

Let $\pi \in \mathrm{Irr}_{\mathrm{gen,meta}}\,\mathbb{M}$, considered also as a representation of M via ϱ. By [Lapid and Mao 2017, Proposition 4.1], $M(s)$ is holomorphic at $s = \frac{1}{2}$. Denote by $\mathcal{D}_\psi(\pi)$ the space of Whittaker functions on $\widetilde{G}$ generated by $A^\psi(M^*W, \Phi, \cdot)$, $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, $\Phi \in \mathcal{S}(F^{\mathbf{n}})$, i.e.,

$$\mathcal{D}_\psi(\pi) = \{A^\psi_\sharp(M^*W, \cdot) : W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))\}.$$

This defines an explicit descent map $\pi \mapsto \mathcal{D}_\psi(\pi)$ on $\mathrm{Irr}_{\mathrm{gen,meta}}\,\mathbb{M}$. By [Ginzburg et al. 1999, theorem in §1.3], $\mathcal{D}_\psi(\pi) \neq 0$. It can be shown that $\mathcal{D}_\psi(\pi)$ is admissible but we do not do it here since we do not use this fact directly.

3C. Good *representations.* Consider $\pi \in \mathrm{Irr}_{\mathrm{gen}}\,M$ and $\tilde{\sigma} \in \mathrm{Irr}_{\mathrm{gen},\psi_{\widetilde{N}}^{-1}}\,\widetilde{G}$ with Whittaker model $\mathbb{W}^{\psi_{\tilde{N}}^{-1}}(\tilde{\sigma})$. Following [Ginzburg et al. 1998], for any

$$\widetilde{W} \in \mathbb{W}^{\psi_{\tilde{N}}^{-1}}(\tilde{\sigma}), \qquad W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi)),$$

define the local Shimura type integral

$$\tilde{J}(\widetilde{W}, W, s) := \int_{N'\backslash G'} \widetilde{W}(\tilde{g}) A^\psi_\sharp(W_s, \tilde{g})\, dg. \tag{3-5}$$

By [Ginzburg et al. 1998, §6.3; 1999], $\tilde{J}$ converges for $\mathrm{Re}\, s \gg_{\pi,\tilde{\sigma}} 1$ and admits a meromorphic continuation in s. Moreover, for any $s \in \mathbb{C}$ we can choose $\widetilde{W}$ and W such that $\tilde{J}(\widetilde{W}, W, s) \neq 0$.

Let $\pi \in \mathrm{Irr}_{\mathrm{gen,meta}}\,\mathbb{M}$. We say that π is *good* if the following conditions are satisfied for all ψ:

(1) $\mathcal{D}_\psi(\pi)$ is irreducible.

(2) $\tilde{J}(\widetilde{W}, W, s)$ is holomorphic at $s = \frac{1}{2}$ for $\widetilde{W} \in \mathcal{D}_{\psi^{-1}}(\pi)$, $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$. (We do not assume that the integral defining $\tilde{J}$ converges at $s = \frac{1}{2}$.)

(3) There is a nondegenerate $\widetilde{G}$-invariant pairing $[\cdot,\cdot]$ on $\mathcal{D}_{\psi^{-1}}(\pi) \times \mathcal{D}_\psi(\pi)$ such that

$$\tilde{J}\big(\widetilde{W}, W, \tfrac{1}{2}\big) = [\widetilde{W}, A^\psi_\sharp(M^*W, \cdot)]$$

for any $\widetilde{W} \in \mathcal{D}_{\psi^{-1}}(\pi)$, $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$. In particular, $\mathcal{D}_\psi(\pi)^\vee \simeq \mathcal{D}_{\psi^{-1}}(\pi)$.

This property was introduced and discussed in [Lapid and Mao 2017, §5]. In particular, if π is good and $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ then there is a constant c_π such that for any $\widetilde{W} \in \mathbb{W}^{\psi_{\tilde{N}}^{-1}}(\tilde{\pi}^\vee)$, $\widetilde{W}^\vee \in \mathbb{W}^{\psi_{\tilde{N}}}(\tilde{\pi})$, we have

$$\int_{N'}^{\mathrm{st}} [\tilde{\pi}(\tilde{n})\widetilde{W}, \widetilde{W}^\vee]\psi_{\widetilde{N}}(n)\, dn = c_\pi \widetilde{W}(e)\widetilde{W}^\vee(e). \tag{3-6}$$

More explicitly, for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ and $W^\wedge \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi))$ we have

$$\int_{N'}^{\mathrm{st}} \tilde{J}\big(A_{\sharp}^{\psi^{-1}}(M^* W^\wedge, \cdot\,\tilde{n}), W, \tfrac{1}{2}\big)\psi_{\widetilde{N}}(n)\,dn = c_\pi A_{\sharp}^{\psi^{-1}}(M^* W^\wedge, e) A_{\sharp}^{\psi}(M^* W, e). \quad (3\text{-}7)$$

Let F be a p-adic field. In the rest of the paper, we prove the following statement:

Theorem 3.4. *For any unitarizable $\pi \in \mathrm{Irr}_{\mathrm{gen,meta}}\,\mathbb{M}$ which is good we have $c_\pi = \epsilon_\pi$.*

Remark 3.5. Of course, we expect Theorem 3.4 to hold in the archimedean case as well. However, we do not deal with the archimedean case in this paper.

The following special case is of particular importance (see Remark 3.9 below).

Corollary 3.6. *Suppose $\pi = \tau_1 \times \cdots \times \tau_l$, where $\tau_j \in \mathrm{Irr}_{\mathrm{sqr,meta}}\,\mathrm{GL}_{2m_j}$, $j = 1, \dots, l$, are distinct and $\mathbf{n} = m_1 + \cdots + m_l$. Then π is good and $\tilde{\pi} := \mathcal{D}_{\psi^{-1}}(\pi)$ is square-integrable. Moreover,*

$$\int_{N'} \tilde{J}\big(\tilde{\pi}(\tilde{n})\widetilde{W}, W, \tfrac{1}{2}\big)\psi_{\widetilde{N}}(n)\,dn = \epsilon_\pi \widetilde{W}(e) A_{\sharp}^{\psi}(M^* W, e)$$

for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ and $\widetilde{W} \in \mathbb{W}^{\psi_{\widetilde{N}}^{-1}}(\tilde{\pi})$.

3D. ***Relation to global statement.*** Suppose now that F is a number field and $\mathbb{A}$ is its ring of adeles. We say that an irreducible cuspidal representation π of $\mathbb{M}$ is of *metaplectic type* if

$$\int_{H_{\mathbb{M}}(F)\backslash H_{\mathbb{M}}(\mathbb{A})\cap \mathbb{M}(\mathbb{A})^1} \varphi(h)\,dh \neq 0$$

for some φ in the space of π. (Here, $\mathbb{M}(\mathbb{A})^1 = \{m \in \mathbb{M}(\mathbb{A}) : |\det m| = 1\}$.) Equivalently, $L^S\big(\frac{1}{2}, \pi\big)\,\mathrm{res}_{s=1} L^S(s, \pi, \wedge^2) \neq 0$ [Friedberg and Jacquet 1993].[2] In particular, π is self-dual and admits a trivial central character. We write $\mathrm{Cusp}_{\mathrm{meta}}\,\mathbb{M}$ for the set of irreducible cuspidal representations of metaplectic type.

Consider the set $\mathrm{MCusp}\,\mathbb{M}$ of automorphic representations π of $\mathbb{M}(\mathbb{A})$ which are realized on Eisenstein series induced from $\pi_1 \otimes \cdots \otimes \pi_k$, where $\pi_i \in \mathrm{Cusp}_{\mathrm{meta}}\,\mathrm{GL}_{2n_i}$, $i = 1, \dots, k$, are distinct and $\mathbf{n} = n_1 + \cdots + n_k$. The representation π is irreducible: it is equivalent to the parabolic induction $\pi_1 \times \cdots \times \pi_k$. Moreover, π determines $\pi_1, \dots, \pi_k$ uniquely up to permutation [Jacquet and Shalika 1981a; 1981b].

Recall that Conjecture 1.1 is pertaining to the representations $\pi \in \mathrm{MCusp}\,\mathbb{M}$. The following fact is crucial for us.

Proposition 3.7 [Lapid and Mao 2017, Theorem 6.2]. *If $\pi \in \mathrm{MCusp}\,\mathbb{M}$, then its local components π_v are good.*

Thus, Theorem 3.4 implies Theorem 1.3, our main result.

[2] We can replace the partial L-function by the completed one since the local factors are holomorphic and nonzero.

3E. ***Relation to Bessel functions.*** We record here a *purely formal* argument that relates Theorem 3.4 to an identity of Bessel functions, which are defined in [Lapid and Mao 2013]. We do not worry about convergence issues in this subsection.

Using the functional equation (see [Kaplan 2015], noting that the central character of π is trivial)

$$\tilde{J}(\widetilde{W}, M(s)W, -s) = |2|^{2\mathbf{n}s} \frac{\gamma\big(\tilde{\pi} \otimes \pi, s+\frac{1}{2}, \psi\big)}{\gamma(\pi, s, \psi)\gamma(\pi, \wedge^2, 2s, \psi)} \tilde{J}(\widetilde{W}, W, s), \tag{3-8}$$

(3-6) becomes

$$\int_{N'} \tilde{J}\big(\tilde{\pi}(\tilde{n})\tilde{W}, M\big(\tfrac{1}{2}\big)W, -\tfrac{1}{2}\big)\psi_{\widetilde{N}}(n)\, dn \\ = |2|^{\mathbf{n}} \frac{\gamma\big(\tilde{\pi} \otimes \pi, s+\frac{1}{2}, \psi\big)}{\gamma(\pi, s, \psi)\gamma(\pi, \wedge^2, 2s, \psi)}\bigg|_{s=1/2} c_\pi \widetilde{W}(e) A_\sharp^\psi(M^*W, e).$$

Explicitly, the left-hand side is

$$\int_{N'} \int_{N'\backslash G'} \widetilde{W}(\tilde{g}\tilde{n}) A_\sharp^\psi(M^*W, \tilde{g})\, dg\, \psi_{\widetilde{N}}(n)\, dn \\ = \int_{N'} \int_{N'\backslash G'} \widetilde{W}(\tilde{g}\tilde{n}) \widetilde{W}'(\tilde{g})\, dg\, \psi_{\widetilde{N}}(n)\, dn,$$

where $\widetilde{W}' = A_\sharp^\psi(M^*W, \cdot)$. Using Bruhat decomposition, this is

$$\int_{T'} \int_{N'} \int_{N'} \delta_{B'}(t) \widetilde{W}\big(\widetilde{w_0'} \tilde{t} \widetilde{n'} \tilde{n}\big) \widetilde{W}'\big(\widetilde{w_0'} \tilde{t} \widetilde{n'}\big) \psi_{\widetilde{N}}(n)\, dn'\, dn\, dt \\ = \int_{T'} \int_{N'} \int_{N'} \delta_{B'}(t) \widetilde{W}\big(\widetilde{w_0'} \tilde{t} \tilde{n}\big) \widetilde{W}'\big(\widetilde{w_0'} \tilde{t} \widetilde{n'}\big) \psi_{\widetilde{N}}(n) \psi_{\widetilde{N}}^{-1}(n')\, dn'\, dn\, dt.$$

By definition of Bessel functions $\mathbb{B}_{\tilde{\pi}}^{\psi_{\widetilde{N}}^{-1}}$ (see (5-1)), this is

$$\widetilde{W}(e)\widetilde{W}'(e) \int_{T'} \mathbb{B}_{\tilde{\pi}}^{\psi_{\widetilde{N}}^{-1}}\big(\widetilde{w_0'} \tilde{t}\big) \mathbb{B}_{\tilde{\pi}^\vee}^{\psi_{\widetilde{N}}}\big(\widetilde{w_0'} \tilde{t}\big) \delta_{B'}(t)\, dt.$$

Here $\tilde{\pi}^\vee = \mathcal{D}_\psi(\pi)$. Thus, on a formal level, Theorem 3.4 becomes the following inner product identity:

$$\int_{T'} \mathbb{B}_{\tilde{\pi}}^{\psi_{\widetilde{N}}^{-1}}\big(\widetilde{w_0'} \tilde{t}\big) \mathbb{B}_{\tilde{\pi}^\vee}^{\psi_{\widetilde{N}}}\big(\widetilde{w_0'} \tilde{t}\big) \delta_{B'}(t)\, dt = |2|^{\mathbf{n}} \frac{\gamma\big(\tilde{\pi} \otimes \pi, s+\frac{1}{2}, \psi\big)}{\gamma(\pi, \wedge^2, 2s, \psi)}\bigg|_{s=1/2}. \tag{3-9}$$

(It is easy to determine the order of the poles in the numerator and denominator on the right-hand side in terms of the data of Theorem 3.8 below.) Of course, the above manipulations are purely formal as the integrals are not absolutely convergent.

3F. ***A reduction of the main theorem.*** For the rest of the paper, let F be a p-adic field. The proof of Theorem 3.4 is essentially local under the additional assumption that $\pi \in \mathrm{Irr}_{\mathrm{temp}}\, \mathbb{M}$ and $\tilde{\pi} := \mathcal{D}_{\psi^{-1}}(\pi) \in \mathrm{Irr}_{\mathrm{temp}}\, \widetilde{G}$. We first show that it indeed suffices to prove Theorem 3.4 under these additional assumptions. This uses a global argument as well as the following classification result due to Matringe. We denote by $\times$ parabolic induction for GL_m.

Theorem 3.8 [Matringe 2015]. *The set* $\mathrm{Irr}_{\mathrm{gen,meta}}\, \mathbb{M}$ *consists of the irreducible representations of the form*

$$\pi = \sigma_1 \times \sigma_1^\vee \times \cdots \times \sigma_k \times \sigma_k^\vee \times \tau_1 \times \cdots \times \tau_l,$$

with $\sigma_1, \ldots, \sigma_k$ *essentially square-integrable,*[3] $\tau_1, \ldots, \tau_l$ *square-integrable and* $L(0, \tau_i, \wedge^2) = \infty$ *for all* i.

(We expect the same result to hold in the archimedean case as well.)

Denote by ω_σ the central character of $\sigma \in \mathrm{Irr}\, \mathrm{GL}_m$. We also write $\sigma[s]$ for the twist of σ by $|\det \cdot|^s$.

Let $\tau_j \in \mathrm{Irr}_{\mathrm{sqr,meta}}\, \mathrm{GL}_{2m_j}$, $j = 1, \ldots, l$, be distinct, and let $\delta_i \in \mathrm{Irr}_{\mathrm{sqr}}\, \mathrm{GL}_{n_i}$, $i = 1, \ldots, k$ with $\mathbf{n} = n_1 + \cdots + n_k + m_1 + \cdots + m_l$. (Possibly $k = 0$ or $l = 0$.) For $\underline{s} = (s_1, \ldots, s_k) \in \mathbb{C}^k$ we consider the representation

$$\pi(\underline{s}) = \delta_1[s_1] \times \delta_1^\vee[-s_1] \times \cdots \times \delta_k[s_k] \times \delta_k^\vee[-s_k] \times \tau_1 \times \cdots \times \tau_l.$$

Suppose that our given p-adic field is the completion at a place v of a number field. We claim that for a dense set of $\underline{s} \in \mathrm{i}\mathbb{R}^k$, $\pi(\underline{s})$ is the local component at v of an element of $\mathrm{MCusp}\, \mathbb{M}$. This follows from [Ichino et al. 2017, Appendix A].[4] Indeed, let $m = m_1 + \cdots + m_l$, let $\rho \in \mathrm{Irr}_{\mathrm{sqr,gen}}\, \mathrm{SO}(2m+1)$ be the representation corresponding to $\tau_1 \times \cdots \times \tau_l$ under Jiang and Soudry [2004] and let $\tilde{\rho} \in \mathrm{Irr}_{\mathrm{sqr,gen}}\, \widetilde{\mathrm{Sp}}_m$ be the theta lift of ρ. Let $\sigma(\underline{s}) = \delta_1[s_1] \times \cdots \times \delta_k[s_k] \rtimes \rho$ (parabolic induction) and $\tilde{\sigma}(\underline{s}) = \delta_1[s_1] \times \cdots \times \delta_k[s_k] \rtimes \tilde{\rho}$. Then for $\underline{s}$ in a dense open subset of $\mathrm{i}\mathbb{R}^k$ we have $\sigma(\underline{s}) \in \mathrm{Irr}\, \mathrm{SO}(2\mathbf{n}+1)$ and $\tilde{\sigma}(\underline{s}) \in \mathrm{Irr}\, \widetilde{G}$ and moreover, by [Gan and Savin 2012], $\tilde{\sigma}(\underline{s})$ is the theta lift of $\sigma(\underline{s})$. By [Ichino et al. 2017, Corollary A.8], for a dense set of $\underline{s} \in \mathrm{i}\mathbb{R}^k$, $\tilde{\sigma}(\underline{s})$ is the local component at v of a generic cuspidal automorphic representation of $\widetilde{\mathrm{Sp}}_{\mathbf{n}}(\mathbb{A})$ whose theta lift to $\mathrm{SO}(2\mathbf{n}+1)$ is cuspidal. The Cogdell–Kim–Piatetski-Shapiro–Shahidi lift [Cogdell et al. 2004] of the latter to $\mathbb{M}$ is the required representation in $\mathrm{MCusp}\, \mathbb{M}$.

It follows from Proposition 3.7 that for a dense set of $\underline{s} \in \mathrm{i}\mathbb{R}^k$, $\pi(\underline{s}) \in \mathrm{Irr}_{\mathrm{temp,meta}}\, \mathbb{M}$ is good. Moreover, by [Ichino et al. 2017, Proposition 4.6], $\mathcal{D}_{\psi^{-1}}(\pi(\underline{s})) = \tilde{\sigma}(\underline{s})$, and in particular, it is tempered.

[3] That is, a twist of a square-integrable representation by a quasicharacter.

[4] We remark that the appendices and §4 of [Ichino et al. 2017], and in particular the proof of [Ichino et al. 2017, Theorem 3.1], are independent of the results of the current paper.

Suppose that $\underline{s}$ is in the domain

$$\mathfrak{D} = \left\{(s_1, \dots, s_k) \in \mathbb{C}^k : -\tfrac{1}{2} < \operatorname{Re} s_i < \tfrac{1}{2} \text{ for all } i\right\}.$$

We recall that by [Lapid and Mao 2017, Lemma 4.12], the integral defining $\tilde{J}\big(A_\sharp^{\psi^{-1}}(M^* W^\wedge, \cdot), W, s\big)$ converges and is holomorphic at $s = \frac{1}{2}$ for any

$$W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi(\underline{s}))), \qquad W^\wedge \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi(\underline{s}))\big). \tag{3-10}$$

Moreover, by the properties of A^ψ and $\widetilde{J}$ (see [Lapid and Mao 2017, (4.4) and (4.13)]), the function $g \mapsto \tilde{J}\big(A_\sharp^{\psi^{-1}}(M^* W^\wedge, \cdot\, \tilde{g}), W, \frac{1}{2}\big)$ is bi-K_0-invariant, provided that $K_0 \in \mathcal{CSGR}^s(G')$ is such that $I\big(\frac{1}{2}, \eta(k)\big)W = W$ and $I\big(\frac{1}{2}, \eta(k)\big)W^\wedge = W^\wedge$ for all $k \in K_0$. It follows from [Lapid and Mao 2015a, Proposition 2.11] (applied to $\widetilde{G}$) that the stable integral on the left-hand side of (3-7) can be written as an integral over a compact open subgroup of N' depending only on K_0. Thus, taking the Whittaker functions in (3-10) to be defined through Jacquet integrals, both sides of (3-7) for $\pi(\underline{s})$ are holomorphic functions of $\underline{s} \in \mathfrak{D}$. Note that by [Lapid and Mao 2017, Lemma 3.6], we have that $\epsilon_{\pi(\underline{s})} = \omega_{\delta_1}(-1) \cdots \omega_{\delta_k}(-1)\epsilon_{\tau_1} \cdots \epsilon_{\tau_l}$ is independent of $\underline{s}$. Thus, in order to prove (3-7) for $\underline{s} \in \mathfrak{D}$, it is enough to show it for a dense set of $\underline{s} \in \mathrm{i}\mathbb{R}^k$. Since every unitarizable $\pi \in \operatorname{Irr}_{\mathrm{meta}} \mathbb{M}$ is of the form $\pi(\underline{s})$ for some $\tau_1, \dots, \tau_l, \delta_1, \dots, \delta_k$ as above and $\underline{s} \in \mathfrak{D}$, the reduction step follows.

Remark 3.9. In the case $k = 0$, we can globalize π itself. Thus, by Proposition 3.7, π is good; $\mathcal{D}_{\psi^{-1}}(\pi)$ is irreducible and square integrable (see [Ichino et al. 2017, Theorem 3.1]).[4] This yields Corollary 3.6 from Theorem 3.4.

In the remainder of the paper we prove Theorem 3.4, i.e., the main identity

$$\int_{N'}^{\mathrm{st}} \left(\int_{N' \backslash G'} A_\sharp^{\psi}(W_s, \tilde{g}) A_\sharp^{\psi^{-1}}(M^* W^\wedge, \tilde{g}\tilde{u})\, dg \right) \psi_{\widetilde{N}}(\tilde{u})\, du \Big|_{s=1/2}$$
$$= \epsilon_\pi A_\sharp^{\psi^{-1}}(M^* W^\wedge, e) A_\sharp^{\psi}(M^* W, e) \tag{MI}$$

under the assumptions that $\pi \in \operatorname{Irr}_{\mathrm{temp,meta}} \mathbb{M}$ is good and $\tilde{\pi} := \mathcal{D}_{\psi^{-1}}(\pi)$ is tempered.

Remark 3.10. As already pointed out before (and used in the reduction step), by [Lapid and Mao 2017, Lemma 4.12] the inner integral on the left-hand side of (MI) converges (and is analytic) for $\operatorname{Re} s \geq \frac{1}{2}$ (for any unitary π). We will not need to use this fact explicitly any further.

4. An informal sketch

Before embarking on the rigorous proof of the relation (MI), let us give a brief sketch of the argument. For convenience we momentarily ignore all convergence issues (which are of course at the heart of the argument, and will be discussed below). Thus, the discussion of this section is *purely formal*.

Define (assuming convergent)

$$B(W, W^\vee) = \int_{N'} \left(\int_{N' \backslash G'} A_\sharp^{\psi}(W, \tilde{g}) A_\sharp^{\psi^{-1}}(W^\vee, \tilde{g}\tilde{u})\, dg \right) \psi_{\widetilde{N}}(\tilde{u})\, du,$$

so that the left-hand side of (MI) is $B(W_{1/2}, M^* W)$.

We use the following formal identity (which follows from the Bruhat decomposition for G'): for (suitable) $\widetilde{W} \in C^{\mathrm{sm}}(\widetilde{N} \backslash \widetilde{G}, \psi_{\widetilde{N}})$ and $\widetilde{W}^\vee \in C^{\mathrm{sm}}(\widetilde{N} \backslash \widetilde{G}, \psi_{\widetilde{N}}^{-1})$,

$$\begin{aligned} &\int_{N'} \left(\int_{N' \backslash G'} \widetilde{W}(\tilde{g}) \widetilde{W}^\vee(\tilde{g}\tilde{u})\, dg \right) \psi_{\widetilde{N}}(\tilde{u})\, du \\ &\quad = \int_{T'} \delta_{B'}(t) \int_{N'} \widetilde{W}(\widetilde{w_0' t n_1}) \psi_{\widetilde{N}}(\tilde{n}_1)^{-1}\, dn_1 \int_{N'} \widetilde{W}^\vee(\widetilde{w_0' t n_2}) \psi_{\widetilde{N}}(\tilde{n}_2)\, dn_2\, dt. \end{aligned} \tag{4-1}$$

(In the ensuing discussion it would be more natural to replace the right-hand side by

$$\begin{aligned} \int_{N'_{M'}} \int_{N'_{M'} \backslash M'} \delta_{P'}(m) \int_{U'} \widetilde{W}(\widetilde{w'_{U'}} \tilde{m} \tilde{n} \widetilde{u_1}) \psi_{\widetilde{N}}(\widetilde{u_1})^{-1}\, du_1 & \\ \int_{U'} \widetilde{W}^\vee(\widetilde{w'_{U'}} \tilde{m} \widetilde{u_2}) \psi_{\widetilde{N}}(\widetilde{u_2})\, du_2\, dm\, \psi_{N'_{M'}}(n)\, dn. & \end{aligned} \tag{4-2}$$

However, it is analytically advantageous to work with the former expression.)

Applying this to $\widetilde{W} = A_\sharp^{\psi}(W, \cdot)$ and $\widetilde{W}^\vee = A_\sharp^{\psi^{-1}}(W^\vee, \cdot)$ we get

$$B(W, W^\vee) = \int_{T'} Y^{\psi}(W, t) Y^{\psi^{-1}}(W^\vee, t) \delta_{B'}(t)\, dt, \tag{4-3}$$

where

$$Y^{\psi}(W, t) := \int_{N'} A_\sharp^{\psi}(W, \widetilde{w_0' t n}) \psi_{\widetilde{N}}(\tilde{n})^{-1}\, dn.$$

We note the following equivariance property of $Y^{\psi}(W, t)$. Let $N^\sharp = V_- \rtimes \eta(N')$, which is the stabilizer of ψ_U in N. Let $\psi_{N^\sharp}$ be the character

$$\psi_{N^\sharp}(\eta(n)v) = \psi_{N'}(n)\psi_{V_-}(v), \quad n \in N',\ v \in V_-$$

on $N^\sharp$. Then by Lemma 3.1, $Y^{\psi}(W, t)$ is $(N^\sharp, \psi_{N^\sharp})$-equivariant in W.

Next, we consider $A_\sharp^{\psi}(W, \widetilde{w_0' t n})$. By Lemma 3.1 it is enough to determine $A_\sharp^{\psi}(W, \widetilde{w'_{U'}})$. We have (up to an immaterial root of unity, which we suppress)

$$A_\sharp^{\psi}(W, \widetilde{w'_{U'}}) = * \int_{V_M^\sharp \backslash V_-} W(w_U v) \psi_{V_-}(v)^{-1}\, dv = * \int_{V_U} W(w_U v) \psi_U(v)^{-1}\, dv.$$

(See Remark 6.5 for a more precise statement.) We also mention that

$$A_\sharp^{\psi}(W, e) = \int_{\epsilon_1^{-1}(V_\gamma \rtimes \eta(N'))\epsilon_1 \backslash N^\sharp} W(\gamma \epsilon_1 n) \psi_{N^\sharp}(n)^{-1}\, dn$$

for a suitable unipotent element ϵ_1 (this will be used for the right-hand side of (MI); see Lemma 6.1). This expression is clearly $(N^\sharp, \psi_{N^\sharp})$-equivariant in W.

Remark 4.1. By [Ginzburg et al. 1999, theorem in §4.4; Lapid and Mao 2015b, Lemma 6.1], for $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}}\,\mathbb{M}$ the space of $(N^\sharp, \psi_{N^\sharp})$-equivariant linear forms on the Langlands quotient of $\mathrm{Ind}\big(\mathbb{W}^{\psi_{N_\mathbb{M}}}(\pi), \frac{1}{2}\big)$ is one-dimensional. This implies that $Y^\psi(M^*W, t)$ is proportional to $A^\psi_\sharp(M^*W, e)$. The constant of proportionality is a complicated function of t which is probably related to the Bessel function. At best, the putative equality becomes something like (3-9), which seems difficult to approach directly. Instead, we take a different approach.

It follows from the formula for $A^\psi_\sharp(W, \widetilde{w'_{U'}})$ above that

$$Y^\psi(W,t) = *\nu'(t)^{\mathbf{n}-1/2} \int_{N'_{M'}} \int_U W(w_U \eta(w_0^{M'} tn)v)\psi_{N'_{M'}}(n)^{-1}\psi_U(u)^{-1}\,du\,dn. \tag{4-4}$$

Substituting this in (4-3) and using the Bruhat decomposition for $\mathbb{M}'$, we obtain

$$\begin{aligned} B(W, W^\vee) = \int_U \int_U \int_{N'_{\mathbb{M}'}} \int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} & W(\eta_M(gn)w_U u_1)W^\vee(\eta_M(g)w_U u_2) \\ & \times \delta_P(\eta_M(g))^{-1}|\det g|^{1-\mathbf{n}}\,dg\,\psi_{N'_{\mathbb{M}'}}(n)\,dn\,\psi_U(u_1)^{-1}\psi_U(u_2)\,du_1\,du_2. \end{aligned} \tag{4-5}$$

For the inner integral, we will use the following functional equation proved in [Lapid and Mao 2014]: for any $W \in \mathbb{W}^{\psi_{N_\mathbb{M}}}(\pi)$ and $W^\vee \in \mathbb{W}^{\psi^{-1}_{N_\mathbb{M}}}(\pi^\vee)$, we have

$$\begin{aligned} \int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} & W(\eta_\mathbb{M}(g))W^\vee(\eta_\mathbb{M}(g))|\det g|^{1-\mathbf{n}}\,dg \\ &= \int_J \int_J \int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(w_{2\mathbf{n},\mathbf{n}}\eta_\mathbb{M}(g)\ell_\mathbb{M}(X))W^\vee(w_{2\mathbf{n},\mathbf{n}}\eta_\mathbb{M}(g)\ell_\mathbb{M}(Y)) \\ &\qquad\qquad \times |\det g|^{\mathbf{n}-1}\,dg\,dX\,dY, \end{aligned}$$

where J is the subspace of $\mathrm{Mat}_\mathbf{n}$ consisting of the matrices whose first column is zero. From this it is easy to infer that

$$\begin{aligned} \int_{N'_{\mathbb{M}'}} & \left(\int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(\eta_\mathbb{M}(gn))W^\vee(\eta_\mathbb{M}(g))|\det g|^{1-\mathbf{n}}\,dg\right)\psi_{N'_{\mathbb{M}'}}(n)\,dn \\ &= \int_{\eta^\vee_\mathbb{M}(N'_{\mathbb{M}'})\backslash N^\flat_\mathbb{M}} \int_{\eta^\vee_\mathbb{M}(N'_{\mathbb{M}'})\backslash N^\flat_\mathbb{M}} \int_{T'_{\mathbb{M}'}} W(\eta^\vee_\mathbb{M}(t)\boldsymbol{\epsilon}_3 r_1)W^\vee(\eta^\vee_\mathbb{M}(t)\boldsymbol{\epsilon}_3 r_2) \\ &\qquad\qquad \times |\det t|^{\mathbf{n}-1}\delta_{B'_{\mathbb{M}'}}(t)^{-1}\psi_{N^\flat_\mathbb{M}}(r_1^{-1}r_2)\,dt\,dr_2\,dr_1, \end{aligned} \tag{4-6}$$

where $N^\flat_\mathbb{M} = (N^\sharp_\mathbb{M})^*$, $\psi_{N^\flat_\mathbb{M}}$ is the character on $N^\flat_\mathbb{M}$ with

$$\psi_{N^\flat_\mathbb{M}}(m) = \psi_{N^\sharp_\mathbb{M}}(m^*)$$

and $\boldsymbol{\epsilon}_3 \in \mathbb{M}$ is a suitable element to be introduced in Section 7B. Substituting this in (4-5), we obtain

$$B(W, W^\vee) = \int_{\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'})} E^\psi(W, t) E^{\psi^{-1}}(W^\vee, t)\, \frac{dt}{|\det t|},$$

where

$$E^\psi(W, t) := \delta_B^{-1/2}(\varrho(t)) \int_{\eta_M(N'_{\mathbb{M}'})\backslash N^\sharp} W(\varrho(t\epsilon_4\epsilon_3) w_U v)\psi_{N^\sharp}(v)^{-1}\, dv, \quad t \in \eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}).$$

Here, $\epsilon_4 \in \mathbb{M}$ is an auxiliary unipotent element chosen so that $\varrho(\epsilon_4\epsilon_3)w_U$ conjugates ψ_U to a character which is supported on a single (short negative) root. The sought-after identity (MI) is now a consequence of the following two results concerning $(N^\sharp, \psi_{N^\sharp})$-equivariant linear forms in W, which are proved in [Lapid and Mao 2015b].

(1) $E^\psi(M^* W, t)$ is the constant function $\epsilon_\pi^{\mathbf{n}} A_e^\psi(M^* W)$.

(2) $\displaystyle\int_{\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'})} E^\psi(W_{1/2}, t)\, \frac{dt}{|\det t|} = \epsilon_\pi^{\mathbf{n}+1} A_e^\psi(M^* W).$

It is here that we use in an essential way that $\pi \in \mathrm{Irr}_{\mathrm{meta}}\, \mathbb{M}$ (and where the factor ϵ_π shows up).

This concludes the sketch of the argument. We finish with a few technical remarks about the rigorous justification of the above argument, and also give a few pointers to the upcoming sections. Hopefully, this will shed some light on (if not motivate) the contents of the remaining sections.

(1) To justify (4-1), with $\int_{N'}$ replaced by $\int_{N'}^{\mathrm{st}}$ on both sides, we are forced to assume that $\widetilde{W}(\widetilde{w'_0 tn})$ is compactly supported on $T' \times N'$ (see Lemma 5.4). This entails imposing a certain support condition on the section W (see Definition 5.5 and Lemma 5.6). Fortunately (and this is a key point in our argument), it is enough to prove (MI) for a single pair $(W, W^\wedge)$ for which the right-hand side is nonzero. For the special sections as above, the nonvanishing is guaranteed by that of the Bessel function of $\tilde{\pi}$, which is proved in the Appendix, following ideas of Ichino and Zhang [2014].

(2) The fact that Y^ψ is well-defined (with $\int_{N'}^{\mathrm{st}}$ instead of $\int_{N'}$) relies on results and techniques of Baruch [2005] (modified to the case at hand in [Lapid and Mao 2013]) on the stability of Bessel functions (Corollary 5.3). This is the reason why we use the Bruhat decomposition with respect to B' (i.e., (4-1)) rather than P' (i.e., (4-2)).

(3) Since we cannot control the support of $M^* W^\wedge$, it is necessary to justify (4-4) for general sections. To that end, we introduce a complex parameter s and consider the family

$$B(W, W^\wedge, s) = B(W_s, W^\wedge_{-s}).$$

The equality (4-4) is then justified for W_s with $\mathrm{Re}\, s \gg 1$. (See Lemma 6.6.) This yields the analogue of (4-5) for $B(W, W^\wedge, s)$, provided that $-\operatorname{Re} s \gg 1$ and W is special (Lemma 6.11).

(4) A similar issue arises with (4-6): we can only justify it if at least one of the Whittaker functions satisfies an additional support constraint (which is in some sense dual to the support condition considered in the first remark). (See Definition 7.5 and Proposition 7.6.) Thus, we take W to be special of the first kind and $W^\wedge$ to be special of the second kind (see Definition 7.8).

(5) Ultimately, we need to consider $B(W, M(s)W^\wedge, s)$ at $s = \frac{1}{2}$. In order to exploit the assumptions on W and $W^\wedge$, we need to know that $M(s)$ is self-adjoint with respect to the bilinear form B. Such a relation was proved in [Lapid and Mao 2014, Appendix B] for a variant of B (where the order of integration is slightly different). (See Corollary 7.2.) For $-\operatorname{Re} s \gg 1$ and W special of the first kind, this agrees with the original B.

(6) The upshot is an identity

$$B(W, M(s)W^\wedge, s) = \int_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})} E^\psi(M_s^* W, t) E^{\psi^{-1}}(W_s^\wedge, t) \frac{dt}{|\det t|}$$

for special sections W, $W^\vee$ and for $-\operatorname{Re} s \gg 1$ (see Proposition 7.11). To finish the argument, we need to show that the right-hand side has analytic continuation at s and compute its value at $s = \frac{1}{2}$. These steps are carried out in [Lapid and Mao 2015b]. See Section 8, where it is also explained why the restriction on $W^\wedge$ is harmless from the point of view of the relation (MI).

5. A bilinear form

We begin the analysis of the main identity. A key ingredient is the stability of the integral defining a Bessel function, which was proved in [Lapid and Mao 2013] following ideas of Baruch [2005].

5A. We recall the main result of [Lapid and Mao 2013] for the group $\widetilde{G}$.

Theorem 5.1 [Lapid and Mao 2013]. *Let $K_0 \in \mathcal{CSGR}^s(G')$. Then the integral*

$$\int_{N'}^{\mathrm{st}} \widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'} tn})\psi_{\widetilde{N}}(n)^{-1}\, dn$$

stabilizes uniformly for $\widetilde{W} \in C(\widetilde{N}\backslash\widetilde{G}, \psi_{\widetilde{N}})^{K_0}$ and locally uniformly in $t \in T'$. In particular, if $\widetilde{W} \in \mathbb{W}^{\psi_{\tilde{N}}}(\tilde{\pi})$ with $\tilde{\pi} \in \operatorname{Irr}_{\mathrm{gen},\psi_{\widetilde{N}}} \widetilde{G}$ then

$$\int_{N'}^{\mathrm{st}} \widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'} tn})\psi_{\widetilde{N}}(n)^{-1}\, dn = \mathbb{B}_{\tilde{\pi}}^{\psi_{\tilde{N}}}(\widetilde{w'_{U'}}\widetilde{w_0^{M'} t})\tilde{W}(e), \tag{5-1}$$

where $\mathbb{B}_{\tilde{\pi}}^{\psi_{\tilde{N}}}$ is the Bessel function of $\tilde{\pi}$ (see (A-1)).

Remark 5.2. Note that $w_0' = w_{U'}' w_0^{M'}$. Of course, one can state the theorem above (equivalently) for

$$\int_{N'}^{\mathrm{st}} \widetilde{W}(\widetilde{w_0' t n}) \psi_{\widetilde{N}}(n)^{-1}\, dn.$$

The original formulation will be slightly more convenient for computation.

We apply this to the function $\widetilde{W}(g) = A_\sharp^\psi(W, \tilde{g})$.

Corollary 5.3. *Let $K_0 \in \mathcal{CSGR}(G)$. Then the integral*

$$Y^\psi(W, t) := \int_{N'}^{\mathrm{st}} A_\sharp^\psi(W, \widetilde{w_{U'}'} \widetilde{w_0^{M'} t n}) \psi_{\widetilde{N}}(n)^{-1}\, dn, \quad t \in T',$$

stabilizes uniformly for $W \in C(N\backslash G, \psi_N)^{K_0}$ and locally uniformly in $t \in T'$. In particular, $Y^\psi(W_s, t)$ is entire in $s \in \mathbb{C}$, and if $\pi \in \mathrm{Irr}_{\mathrm{gen}}\, \mathbb{M}$ and $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ then $Y^\psi(M_s^* W, t)$ is meromorphic in s. Both $Y^\psi(W_s, t)$ and $Y^\psi(M_s^* W, t)$ are locally constant in t, uniformly in $s \in \mathbb{C}$.

Finally, if we assume that $\pi \in \mathrm{Irr}_{\mathrm{meta,gen}}\, \mathbb{M}$ and that $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ is irreducible then for any $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)$ we have

$$Y^{\psi^{-1}}(M^* W^\wedge, t) = \mathbb{B}_{\tilde{\pi}}^{\psi_{\widetilde{N}}^{-1}}(\widetilde{w_{U'}'} \widetilde{w_0^{M'} t}) A_\sharp^{\psi^{-1}}(M^* W^\wedge, e).$$

Another useful consequence of Theorem 5.1 is the following.

Lemma 5.4. *Let $\widetilde{W} \in C^{\mathrm{sm}}(\widetilde{N}\backslash\widetilde{G}, \psi_{\widetilde{N}})$ and $\widetilde{W}^\vee \in C^{\mathrm{sm}}(\widetilde{N}\backslash\widetilde{G}, \psi_{\widetilde{N}}^{-1})$. Assume that the function*

$$(t, n) \in T' \times N' \mapsto \widetilde{W}(\widetilde{w_0' t n})$$

is compactly supported. (*In particular, $\widetilde{W} \in C_c^{\mathrm{sm}}(\widetilde{N}\backslash\widetilde{G}, \psi_{\widetilde{N}})$.*) *Then the iterated integral*

$$\int_{N'}^{\mathrm{st}} \left(\int_{N'\backslash G'} \widetilde{W}(\tilde{g}) \widetilde{W}^\vee(\tilde{g}\tilde{u})\, dg \right) \psi_{\widetilde{N}}(\tilde{u})\, du$$

is well defined and is equal to

$$\int_{T'} \int_{N'} \delta_{B'}(t) \widetilde{W}(\widetilde{w_{U'}'} \widetilde{w_0^{M'} t n}) \psi_{\widetilde{N}}(\tilde{n}^{-1}) \left(\int_{N'}^{\mathrm{st}} \widetilde{W}^\vee(\widetilde{w_{U'}'} \widetilde{w_0^{M'} t u}) \psi_{\widetilde{N}}(\tilde{u})\, du \right) dn\, dt.$$

Proof. Using the Bruhat decomposition we can write the left-hand side as

$$\int_{N'}^{\mathrm{st}} \left(\int_{T'} \int_{N'} \widetilde{W}(\widetilde{w_{U'}'} \widetilde{w_0^{M'} t n}) \widetilde{W}^\vee(\widetilde{w_{U'}'} \widetilde{w_0^{M'} t n \tilde{u}}) \delta_{B'}(t) \psi_{\widetilde{N}}(\tilde{u})\, dn\, dt \right) du.$$

Note that by the assumption on $\widetilde{W}$, the integrand is compactly supported in t, n. Consider

$$\int_{\Omega}\left(\int_{T'}\int_{N'}\widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn})\widetilde{W}^{\vee}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn\tilde{u}})\delta_{B'}(t)\psi_{\widetilde{N}}(\tilde{u})\,dn\,dt\right)du$$
$$=\int_{T'}\int_{N'}\int_{\Omega}\delta_{B'}(t)\widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn})\widetilde{W}^{\vee}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn\tilde{u}})\psi_{\widetilde{N}}(\tilde{u})\,du\,dn\,dt$$

for $\Omega\in\mathcal{CSGR}(N')$. By Theorem 5.1 this is equal to

$$\int_{T'}\int_{N'}\delta_{B'}(t)\widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn})\left(\int_{N'}^{\mathrm{st}}\widetilde{W}^{\vee}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn\tilde{u}})\psi_{\widetilde{N}}(\tilde{u})\,du\right)dn\,dt,$$

provided that Ω is sufficiently large (independently of t and n, since they can be confined to compact sets by the assumption on $\widetilde{W}$). Making a change of variable $u\mapsto n^{-1}u$, we get

$$\int_{T'}\int_{N'}\delta_{B'}(t)\widetilde{W}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn})\psi_{\widetilde{N}}(\tilde{n}^{-1})\left(\int_{N'}^{\mathrm{st}}\widetilde{W}^{\vee}(\widetilde{w'_{U'}}\widetilde{w_0^{M'}tu})\psi_{\widetilde{N}}(\tilde{u})\,du\right)dn\,dt,$$

as required. $\square$

5B. In order to apply Lemma 5.4 for $A_{\sharp}^{\psi}(W_s,\cdot)$, we make a special choice of W. Consider the P-invariant subspace $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}$ of $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ consisting of functions supported in the big cell $Pw_UP=Pw_UU$. Any element of $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}$ is a linear combination of functions of the form

$$W(u'mw_Uu)=\delta_P(m)^{1/2}W^M(m)\phi(u),\quad m\in M,\ u,u'\in U,\tag{5-2}$$

with $W^M\in\mathbb{W}^{\psi_{N_M}}(\pi)$ and $\phi\in C_c^{\infty}(U)$. Let $\eta_{\mathbb{M}}$ be the embedding $\eta_{\mathbb{M}}(g)=\begin{pmatrix}g&\\&I_{\mathbf{n}}\end{pmatrix}$ of $\mathbb{M}'$ into $\mathbb{M}$. Also let $\eta_M=\varrho\circ\eta_{\mathbb{M}}$, so that

$$\eta_M(g)=\begin{pmatrix}g&&\\&I_{2\mathbf{n}}&\\&&g^*\end{pmatrix}.$$

Definition 5.5. Let $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_{\sharp}^{\circ}$ be the linear subspace of $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}$ generated by those W as in (5-2) that satisfy the additional property that the function $(t,n)\mapsto W^M(\eta_M(tw_0^{\mathbb{M}'}n))$ is compactly supported on $T'_{\mathbb{M}'}\times N'_{\mathbb{M}'}$, or equivalently, that the function $W^M\circ\eta_M$ on $\mathbb{M}'$ is supported in the big cell $B'_{\mathbb{M}'}w_0^{\mathbb{M}'}N'_{\mathbb{M}'}$ and its support is compact modulo $N'_{\mathbb{M}'}$.

This space is nonzero; see the proof of Lemma 6.13 below. It is also invariant under $\eta(T')\ltimes N$.

Lemma 5.6. *For any $W\in\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_{\sharp}^{\circ}$, the function $A_{\sharp}^{\psi}(W_s,\widetilde{w'_{U'}}\widetilde{w_0^{M'}tn})$ is compactly supported in $t\in T'$ and $n\in N'$ uniformly in $s\in\mathbb{C}$.*

Proof. Since $\Phi * W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ whenever $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ (for any $\Phi \in C_c^\infty(F^{\mathbf{n}})$), it is enough to show (in view of the definition of $A^\psi_\sharp$, upon replacing the domain of integration in (3-1) by $\eta(w'_{U'})V_U\eta(w'_{U'})^{-1}$), that for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, the function $(v,t,n) \mapsto W(\gamma\eta(w'_{U'})v\eta(w_0^{M'}tn))$ is compactly supported in $V_U \times T' \times N'$. Write $t = \varrho'(t')$ and $n \in N'$ as $n = \varrho'(n'_1)n_2$ with $n_1 \in N'_{\mathbb{M}'}$ and $n_2 \in U'$. Also let $m' = w_0^{\mathbb{M}'}t'n'_1 \in \mathbb{M}'$ and $m = \eta(\varrho'(m')) = \eta(w_0^{M'}t\varrho'(n'_1))$. Then

$$W(\gamma\eta(w'_{U'})v\eta(w_0^{M'}tn)) = W(w_U v m\eta(n_2)) = W(\eta_M((m')^*)w_U m^{-1}vm\eta(n_2)).$$

It follows immediately from the definition of the space $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ that this function is compactly supported in $(v,n_1,n_2,t') \in V_U \times N'_{\mathbb{M}'} \times U' \times T'_{\mathbb{M}'}$. □

5C. For $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ and $W^\vee \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee)\big)$, we define

$$B(W,W^\vee,s) := \int^{\mathrm{st}}_{N'}\bigg(\int_{N'\backslash G'} A^\psi_\sharp(W_s,\tilde{g})A^{\psi^{-1}}_\sharp(W^\vee_{-s},\tilde{g}\tilde{u})\,dg\bigg)\psi_{\widetilde{N}}(\tilde{u})\,du.$$

By Lemma 5.6 and the proof of Lemma 5.4, $B(W,W^\vee,s)$ is an entire function of s and we have

$$B(W,W^\vee,s) = \int_{T'} Y^\psi(W_s,t)Y^{\psi^{-1}}(W^\vee_{-s},t)\delta_{B'}(t)\,dt \tag{5-3}$$

for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ and $W^\vee \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$.

Assume that $\pi \in \mathrm{Irr}_{\mathrm{gen,meta}}\,\mathbb{M}$ and $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ is irreducible. Then for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ and $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$,

$$\text{the left-hand side of the main identity (MI) is } B\big(W, M\big(\tfrac{1}{2}\big)W^\wedge, \tfrac{1}{2}\big). \tag{5-4}$$

We obtain a more explicit form of $B(W,W^\vee,s)$ for $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$ in the next section.

6. Further analysis

6A. *The function $A^\psi_\sharp(W_s,\tilde{g})$.* At this point it is necessary to investigate $A^\psi_\sharp(W_s,\cdot)$ on the big cell.

We start with $A^\psi_\sharp(W,e)$, which is easier and also useful for the right-hand side of (6-12).

Fix an element $\epsilon_1 \in V$ of the form $\ell_M(X)$, where $X \in \mathrm{Mat}_{\mathbf{n}}$ and the last row of X is $-\xi_{\mathbf{n}}$. For any $W \in C^{\mathrm{sm}}(N\backslash G,\psi_N)$, define

$$A^\psi_e(W) := \int_{V_\gamma\backslash V_-} W(\gamma v\epsilon_1)\psi_{V_-}(\epsilon_1^{-1}v\epsilon_1)^{-1}\,dv. \tag{6-1}$$

Since $\psi_{V_-}(\boldsymbol{\epsilon}_1^{-1}v\boldsymbol{\epsilon}_1)=\psi_{V_-}(v)\psi(v_{\mathbf{n},2\mathbf{n}+1})$, this expression is clearly independent of the choice of $\boldsymbol{\epsilon}_1$ as above. Moreover, we have

$$\begin{aligned} A_e^\psi(W) &= \int_{(V_\gamma\rtimes\eta(N'))\backslash N^\sharp} W(\gamma n\boldsymbol{\epsilon}_1)\psi_{N^\sharp}(\boldsymbol{\epsilon}_1^{-1}n\boldsymbol{\epsilon}_1)^{-1}\,dn \\ &= \int_{\boldsymbol{\epsilon}_1^{-1}(V_\gamma\rtimes\eta(N'))\boldsymbol{\epsilon}_1\backslash N^\sharp} W(\gamma\boldsymbol{\epsilon}_1 n)\psi_{N^\sharp}(n)^{-1}\,dn. \end{aligned} \tag{6-2}$$

The integrand in the first equation is invariant under $\eta(N')$ since the character $\psi_{N^\sharp}(\boldsymbol{\epsilon}_1^{-1}\eta(u)\boldsymbol{\epsilon}_1)$ is trivial on U' and agrees with $\psi_{N'_{M'}}$ on $N'_{M'}$.

Lemma 6.1. *For any $W\in C^{\mathrm{sm}}(N\backslash G,\psi_N)$, the integrand in (6-1) is compactly supported on $V_\gamma\backslash V_-$ and we have $A_\sharp^\psi(W,e)=A_e^\psi(W)$.*

Remark 6.2. Note that by Lemma 3.1, the linear form $W\mapsto A_\sharp^\psi(W,e)$ is $(N^\sharp,\psi_{N^\sharp})$-equivariant. By (6-2) the same is true for $A_e^\psi(W)$.

Proof. The support condition follows from Lemma 3.2. By (3-1) we have

$$A^\psi(W,\Phi,e)=\int_{V_\gamma\backslash V} W(\gamma v)\omega_{\psi^{-1}}(v)\Phi(\xi_{\mathbf{n}})\,dv.$$

We can write this as

$$\int_{V_+}\int_{V_\gamma\backslash V_-} W(\gamma vc)\omega_{\psi^{-1}}(vc)\Phi(\xi_{\mathbf{n}})\,dv\,dc,$$

which by (2-1a)–(2-1c) is equal to

$$\int_{V_+}\int_{V_\gamma\backslash V_-} W(\gamma vc)\psi_{V_-}(v)^{-1}\psi(v_{\mathbf{n},2\mathbf{n}+1})^{-1}\Phi(\underline{c}+\xi_{\mathbf{n}})\,dv\,dc.$$

Changing variables in c and v, we can rewrite this as

$$\begin{aligned} &\int_{V_+}\int_{V_\gamma\backslash V_-} W(\gamma v\boldsymbol{\epsilon}_1 c)\Phi(\underline{c})\psi_{V_-}(v)^{-1}\psi(v_{\mathbf{n},2\mathbf{n}+1})^{-1}\,dv\,dc \\ &\qquad\qquad = \int_{V_\gamma\backslash V_-}(\Phi * W)(\gamma v\boldsymbol{\epsilon}_1)\psi_{V_-}(v)^{-1}\psi(v_{\mathbf{n},2\mathbf{n}+1})^{-1}\,dv\,dc. \end{aligned}$$

This is $A_e^\psi(\Phi * W)$, as $\psi_{V_-}(\boldsymbol{\epsilon}_1^{-1}v\boldsymbol{\epsilon}_1)=\psi_{V_-}(v)\psi(v_{\mathbf{n},2\mathbf{n}+1})$. □

We now turn to $A_\sharp^\psi(W_s,\cdot)$ on the big cell. By Lemma 3.1, it is enough to consider the element $w'_{U'}$.

Lemma 6.3. *Let $\pi\in\mathrm{Irr}_{\mathrm{gen}}\,M$. Then for $\mathrm{Re}\,s\gg_\pi 1$ and any $W\in\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, we have*

$$A_\sharp^\psi(W_s,\widetilde{w'_{U'}})=\beta_{\psi^{-1}}(w'_{U'})\int_{V_U}W_s(w_Uv)\psi_U(v)^{-1}\,dv$$
$$=\beta_{\psi^{-1}}(w'_{U'})\int_{V_M^\sharp\backslash V_-}W_s(w_Uv)\psi_{V_-}(v)^{-1}\,dv. \tag{6-3}$$

Remark 6.4. It is easy to see that both sides of (6-3) are $\bigl(V_-\rtimes\eta(N'_{M'}),\,\psi_{V_-}\psi_{N'_{M'}}^{-1}\bigr)$-equivariant in W.

Proof. From (3-1),

$$A^\psi(W_s,\Phi,\widetilde{w'_{U'}})=\int_{V_\gamma\backslash V}W_s(\gamma v\eta(w'_{U'}))\omega_{\psi^{-1}}(v\widetilde{w'_{U'}})\Phi(\xi_{\mathbf{n}})\,dv.$$

Make a change of variable $v\mapsto\eta(w'_{U'})v\eta(w'_{U'})^{-1}$. Note that

$$V_\gamma=\eta(w'_{U'})V_M\eta(w'_{U'})^{-1},$$

$\eta(w'_{U'})$ normalizes V, and $V=V_M\ltimes V_U$. By (2-5) we infer that

$$A^\psi(W_s,\Phi,\widetilde{w'_{U'}})=\int_{V_U}W_s(w_Uv)\omega_{\psi^{-1}}(\widetilde{w'_{U'}}v)\Phi(\xi_{\mathbf{n}})\,dv,$$

and using (2-3c) we get

$$\beta_{\psi^{-1}}(w'_{U'})\int_{V_U}\left(\int_{F^{\mathbf{n}}}W_s(w_Uv)\omega_{\psi^{-1}}(v)\Phi(Y)\psi(Y_1)^{-1}\,dY\right)dv.$$

We claim that the double integral converges absolutely when $\operatorname{Re}s$ is large enough. Note that $V_U=\bigl\{\ell\bigl(\bigl(\begin{smallmatrix}x&y\\&\check{x}\end{smallmatrix}\bigr)\bigr):x\in\mathrm{Mat}_{\mathbf{n}},\,y\in\mathfrak{s}_{\mathbf{n}}\bigr\}$ and hence $|\omega_{\psi^{-1}}(v)\Phi(Y)|=|\Phi(Y)|$ for $v\in V_U$. Thus,

$$\int_{V_U}\int_{F^{\mathbf{n}}}|W_s(w_Uv)\omega_{\psi^{-1}}(v)\Phi(Y)\psi(Y_1)^{-1}|\,dY\,dv=\widehat{|\Phi|}(0)\int_{V_U}|W_s(w_Uv)|\,dv.$$

The integration on the right-hand side is absolutely convergent when $\operatorname{Re}s\gg_\pi 1$. This follows from the convergence of $\int_U|W_s(w_Uu)|\,du$ and Remark 2.2.

For $c\in V_+$ such that $\underline{c}=Y\in F^{\mathbf{n}}$, we have $W_s(w_Ucg)=\psi(Y_1)^{-1}W_s(w_Ug)$ by the equivariance of W, and $\omega_{\psi^{-1}}(c)\Phi(0)=\Phi(Y)$ by (2-1a). Thus,

$$\beta_{\psi^{-1}}(w'_{U'})^{-1}A^\psi(W_s,\Phi,\widetilde{w'_{U'}})=\int_{V_U}\int_{V_+}W_s(w_Ucv)\omega_{\psi^{-1}}(cv)\Phi(0)\,dc\,dv.$$

Since V_+ normalizes V_U and the double integral converges we can write the last integral as

$$\int_{V_U}\int_{V_+} W_s(w_U vc)\omega_{\psi^{-1}}(vc)\Phi(0)\,dc\,dv$$
$$= \int_{V_U}\int_{V_+} W_s(w_U vc)\psi_{V_-}(v)^{-1}\Phi(\underline{c})\,dc\,dv = \int_{V_U} (\Phi * W_s)(w_U v)\psi_{V_-}(v)^{-1}\,dv.$$

This gives the first equality in (6-3), and the second one follows from it since the integrand is left $V_M^\sharp$-invariant. □

Remark 6.5. One can show that for any $W \in C^{\mathrm{sm}}(N\backslash G, \psi_N)$, we have

$$A_\sharp^\psi(W, \widetilde{w'_{U'}}) = \beta_{\psi^{-1}}(w'_{U'}) \int_{V''\backslash V_-} \left(\int_{V_M^\sharp\backslash V''} W(w_U vx)\psi_{V_-}(vx)^{-1}\,dv \right) dx, \quad (6\text{-}4)$$

where on the right-hand side V'' is the preimage of the center of the Heisenberg group under $v \mapsto v_{\mathcal{H}}$ and in both the inner and outer integral the integrand is compactly supported, thus convergent. We do not give the details since we are not going to use this result.

6B. We can now compute $Y^\psi(W_s, t)$ for $\operatorname{Re} s \gg 1$.

Lemma 6.6. *Let $\pi \in \operatorname{Irr}_{\mathrm{gen}} M$. For $\operatorname{Re} s \gg_\pi 1$ and any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, $t \in T'$, we have the identity*

$$Y^\psi(W_s, t) = \nu'(t)^{\mathbf{n}-1/2}\beta_{\psi^{-1}}(w'_{U'})\beta_{\psi^{-1}}(w_0^{M'}t)$$
$$\times \int_{V_M^\sharp\backslash N^\sharp} W_s(w_U \eta(w_0^{M'}t)v)\psi_{N^\sharp}(v)^{-1}\,dv, \quad (6\text{-}5)$$

where the right-hand side is absolutely convergent.

Remark 6.7. Clearly, both sides of (6-5) are $(N^\sharp, \psi_{N^\sharp})$-equivariant in W.

We first need the following convergence result.

Lemma 6.8. *Let $\pi \in \operatorname{Irr}_{\mathrm{gen}} M$. Then for $\operatorname{Re} s \gg_\pi 1$, we have*

$$\int_{N_{\mathbb{M}}^t\cap\mathcal{P}}\int_U |W_s(\varrho(n)w_U ug)|\,du\,dn < \infty, \quad (6\text{-}6a)$$
$$\int_{N_{\mathbb{M}}^t\cap\mathcal{P}^*}\int_U |W_s(\varrho(n)w_U ug)|\,du\,dn < \infty, \quad (6\text{-}6b)$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, $g \in G$.

Proof. We may assume without loss of generality that $g = e$. Denote by $\varrho(a(x))$ the torus part in the Iwasawa decomposition of $x \in G$. Let $\Delta_i(x)$ be the set of nonzero $i \times i$ minors in the last i rows of x (with the convention that $\Delta_0(x) = \{1\}$). Then $\Delta_i(\varrho(a(x)))$ is (obviously) a singleton and its absolute value is $\max_{\delta\in\Delta_i(x)}|\delta|$.

For any $n \in N^t_{\mathbb{M}}$ and $u \in U$ consider

$$a(\varrho(n)w_U u) = \operatorname{diag}(a_1, \ldots, a_{2\mathbf{n}}).$$

Notice that

$$\prod_{i=1}^{2\mathbf{n}} |a_i| = |\det(a(\varrho(n)w_U u))| = |\det(a(w_U u))| \le 1 \tag{6-7}$$

and for all $1 \le j \le i \le 2\mathbf{n}$ we have $n_{i,j} \in \{0\} \cup \Delta_{i-1}(\varrho(n)w_U u)$ (since $n_{i,j}$ is an entry in the adjugate of n^*). Hence

$$|n_{i,j}| \le \prod_{k=1}^{i-1} |a_k|^{-1}. \tag{6-8}$$

We show below that for all $n \in N^t_{\mathbb{M}} \cap (\mathcal{P} \cup \mathcal{P}^*)$ and $u \in U$ such that $W(\varrho(n)w_U u) \neq 0$, we have

$$|\det(a(w_U u))| \ll_W |a_i| \ll_W |\det(a(w_U u))|^{2-2\mathbf{n}}, \quad i = 1, \ldots, 2\mathbf{n}. \tag{6-9}$$

Assuming this, we first show the lemma. It follows from (6-7) and (6-9) that there exists $\lambda \in \mathbb{R}$, depending only on π such that

$$|W_s(\varrho(n)w_U u)| \ll_W |\det(a(w_U u))|^{\operatorname{Re} s - \lambda} \tag{6-10}$$

for any $u \in U$ and $n \in N^t_{\mathbb{M}} \cap (\mathcal{P} \cup \mathcal{P}^*)$. Together with (6-8) and (6-9) we obtain $|n_{i,j}| \ll_W |\det(a(w_U u))|^{1-i}$ whenever $W(\varrho(n)w_U u) \neq 0$. Thus,

$$\int_{N^t_{\mathbb{M}} \cap \mathcal{P}} \int_U |W_s(\varrho(n)w_U u)| \, du \, dn \ll_W \int_U |\det(a(w_U u))|^{\operatorname{Re} s - \lambda - \sum_{i=1}^{2\mathbf{n}-1} i^2} \, du$$

(and similarly for $\int_{N^t_{\mathbb{M}} \cap \mathcal{P}^*}$), where the last integral is finite when $\operatorname{Re} s \gg_\pi 1$ by a standard result on intertwining operators.

We are left to prove (6-9). First, consider the case where $n \in N^t_{\mathbb{M}} \cap \mathcal{P}$. Then the $2\mathbf{n}$-th row of $\varrho(n)w_U u$ is $\xi_{4\mathbf{n}}$. Thus, $\Delta_{2\mathbf{n}+1}(\varrho(n)w_U u) \subset \Delta_{2\mathbf{n}}(\varrho(n)w_U u)$, and hence

$$|a_{2\mathbf{n}}| \prod_{i=1}^{2\mathbf{n}} |a_i|^{-1} \le \prod_{i=1}^{2\mathbf{n}} |a_i|^{-1},$$

or equivalently $|a_{2\mathbf{n}}| \le 1$. Thus, if $W(\varrho(n)w_U u) \neq 0$ then $|a_1| \ll_W \cdots \ll_W |a_{2\mathbf{n}}| \le 1$. The relation (6-9) follows (since $1 \le |\det(a(w_U u))|^{2-2\mathbf{n}}$).

Next, consider the case $n \in N^t_{\mathbb{M}} \cap \mathcal{P}^*$. The last row of $\varrho(n)w_U u$ is the same as the last row of $w_U u$. Thus, $|a_1|^{-1}$ is bounded above by the norm of the entries in u, which in turn is bounded above by $|\det(a(w_U u))|^{-1}$ (since each nonzero entry of u belongs to $\Delta_{2\mathbf{n}}(w_U u)$). Again, if $W(\varrho(n)w_U u) \neq 0$ then $|a_1|_W \ll \cdots \ll_W |a_{2\mathbf{n}}|$.

From (6-7) we conclude that $|\det(a(\varrho(n)w_U u))| \ll |a_1| \ll_W |a_i|$ for all i. On the other hand,

$$\begin{aligned}|\det(a(w_U u))|^{2\mathbf{n}-2} &= |a_1|^{2\mathbf{n}-2}|a_2\cdots a_{2\mathbf{n}-1}|^{2\mathbf{n}-2}|a_{2\mathbf{n}}|^{2\mathbf{n}-2}\\ &\ll_W |a_2\cdots a_{2\mathbf{n}-1}|^{2\mathbf{n}-1}|a_{2\mathbf{n}}|^{2\mathbf{n}-2}\\ &= |a_2\cdots a_{2\mathbf{n}}|^{2\mathbf{n}-1}|a_{2\mathbf{n}}|^{-1} \le |a_{2\mathbf{n}}|^{-1},\end{aligned}$$

and the relation (6-9) follows. □

Remark 6.9. From the proof it is clear that we can take a uniform region of convergence $\operatorname{Re} s \gg 1$ for all unitarizable π.

Lemma 6.6 is an immediate consequence of Lemmas 6.3, 3.1 and 6.8. Indeed, the right-hand side of (6-5) can be viewed as a "partial integration" of the double integral (6-6a) for $g = \eta(w_0^{M'}t)$. Thus, by Remark 2.2, it is absolutely convergent for $\operatorname{Re} s \gg_\pi 1$. Meanwhile, by Lemma 3.1,

$$\begin{aligned}Y^\psi(W_s,t) &= \int_{N'} A_\sharp^\psi(W_s, \widetilde{w'_{U'}}\,\widetilde{w_0^{M'}tn})\psi_{\widetilde{N}}(n)^{-1}\,dn\\ &= \nu'(t)^{-1/2}\beta_{\psi^{-1}}(w_0^{M'}t)\int_{N'} A_\sharp^\psi(W_s(\cdot\,\eta(w_0^{M'}tn)), \widetilde{w'_{U'}})\psi_{\widetilde{N}}(n)^{-1}\,dn,\end{aligned}$$

assuming the integral converges. From (6-3), this is

$$\nu'(t)^{-1/2}\beta_{\psi^{-1}}(w_0^{M'}t)\beta_{\psi^{-1}}(w'_{U'}) \times \int_{N'}\int_{V_U} W_s(w_U v\eta(w_0^{M'}tn))\psi_U(v)^{-1}\psi_{\widetilde{N}}(n)^{-1}\,dv\,dn.$$

Conjugating v over $\eta(w_0^{M'}t)$ and combining the two integrals (which converge by the above), we get Lemma 6.6. (We note that the integrand on the right-hand side of (6-5) is indeed left $V_M^\sharp$-invariant.)

6C. We now go back to the bilinear form B defined in Section 5C. Recall the definition of the space $\operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_\sharp^\circ$ in Section 5B.

Lemma 6.10. *For $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_\sharp^\circ$, the integrand on the right-hand side of (6-5) is compactly supported in t, v uniformly in s (i.e., the support in (t,v) is contained in a compact set which is independent of s). In particular, the identity (6-5) holds for all $s \in \mathbb{C}$.*

Proof. Suppose that $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_\sharp^\circ$ is of the form (5-2). Then for $n \in N'_{\mathbb{M}'}$, $u \in U$ and $t \in T'_{\mathbb{M}'}$, we have $\eta(\varrho'(w_0^{\mathbb{M}'}tn)) \in M$ and

$$W(w_U\eta(\varrho'(w_0^{\mathbb{M}'}tn))u) = \delta_P(\eta(t))^{-1/2}W^M(\eta_M((w_0^{\mathbb{M}'}tn)^*))\phi(u).$$

The lemma follows from the definition of $\operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_\sharp^\circ$. □

Lemma 6.11. *Let $\pi \in \operatorname{Irr}_{\mathrm{gen}} M$. Then for $-\operatorname{Re} s \gg 1$, we have*

$$B(W, W^\vee, s) = \int_U \int_{V_M^\sharp \backslash N^\sharp} \int_{N'_{\mathbb{M}'} \backslash \mathbb{M}'} W_s(\eta_M(g) w_U v) W^\vee_{-s}(\eta_M(g) w_U u) \\ \delta_P(\eta_M(g))^{-1} |\det g|^{1-\mathbf{n}} \psi_{N^\sharp}(v)^{-1} \psi_U(u)\, dg\, dv\, du$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee)\big)$, with the integral being absolutely convergent.

Proof. Suppose that $-\operatorname{Re} s \gg 1$ and $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$. Then by (5-3), Lemmas 6.6 and 6.10, and the fact that $\delta_{B'}(t) = \delta_{B'_{M'}}(t) \nu'(t)^{\mathbf{n}+1}$, $B(W, W^\vee, s)$ equals

$$\int_{T'_{\mathbb{M}'}} \int_{\eta(N'_{M'}) \ltimes U} \int_{\eta(N'_{M'}) \ltimes U} W_s(\eta_M((w_0^{\mathbb{M}'} t)^*) w_U v_1) W^\vee_{-s}(\eta_M((w_0^{\mathbb{M}'} t)^*) w_U v_2) \\ \times |\det t|^{3\mathbf{n}} \delta_{B'_{\mathbb{M}'}}(t) \psi_{N^\sharp}(v_1)^{-1} \psi_{N^\sharp}(v_2)\, dv_1\, dv_2\, dt,$$

where the integral is absolutely convergent. Here we use $\eta(N'_{M'}) \ltimes U$ as a section of $V_M^\sharp \backslash N^\sharp$.

Making a change of variable $v_1 \mapsto v_2 v_1$ and writing $v_2 = \eta(\varrho'(n))u$ with $n \in N'_{\mathbb{M}'}$ and $u \in U$, we get that $B(W, W^\vee, s)$ is equal to

$$\int_{T'_{\mathbb{M}'}} \int_{N'_{\mathbb{M}'}} \int_U \int_{\eta(N'_{M'}) \ltimes U} W_s(\eta_M((w_0^{\mathbb{M}'} tn)^*) w_U u v_1) W^\vee_{-s}(\eta_M((w_0^{\mathbb{M}'} tn)^*) w_U u) \\ \times |\det t|^{3\mathbf{n}} \delta_{B'_{\mathbb{M}'}}(t) \psi_{N^\sharp}(v_1)^{-1}\, dv_1\, du\, dn\, dt.$$

Make a further change of variable $v_1 \mapsto u^{-1} v_1$. It remains to use the Bruhat decomposition for $\mathbb{M}'$ and to note that $\delta_P(\eta_M(g)) = |\det g|^{2\mathbf{n}+1}$ for any $g \in \mathbb{M}'$ and that as a function of v, the integrand in the statement of the lemma is left $V_M^\sharp$-invariant. □

Define when convergent

$$\{W, W^\vee\} := \int_{N'_{\mathbb{M}'} \backslash \mathbb{M}'} W(\eta_M(g)) W^\vee(\eta_M(g)) \delta_P(\eta_M(g))^{-1} |\det g|^{1-\mathbf{n}}\, dg.$$

Then we get, for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee)\big)$ and when $-\operatorname{Re} s \gg 1$,

$$B(W, W^\vee, s) \\ = \int_U \int_{V_M^\sharp \backslash N^\sharp} \{W_s(\cdot\, w_U v), W^\vee_{-s}(\cdot\, w_U u)\} \psi_{N^\sharp}(v)^{-1} \psi_U(u)\, dv\, du. \quad (6\text{-}11)$$

Remark 6.12. If π is unitarizable then the integral defining $\{W, W^\vee\}$ is absolutely convergent for any $(W, W^\vee) \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi)) \times \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee)\big)$ by [Lapid and Mao 2014, Lemma 1.2].

6D. We need a nonvanishing result which follows from Theorem A.1.

Lemma 6.13. *Assume that* $\pi \in \mathrm{Irr}_{\mathrm{gen,meta}} M$ *and* $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ *is irreducible and tempered. Then the bilinear form* $B\big(W, M\big(\frac{1}{2}\big)W^{\wedge}, \frac{1}{2}\big)$ *does not vanish identically on* $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp} \times \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$.

Proof. Since the image of the restriction map $\mathbb{W}^{\psi_{N_M}}(\pi) \to C(N_M\backslash\varrho(\mathcal{P}), \psi_{N_M})$ contains $C_c^{\infty}(N_M\backslash\varrho(\mathcal{P}), \psi_{N_M})$, it follows that for any $\varphi \in C_c^{\infty}(T'_{\mathbb{M}'} \times N'_{\mathbb{M}'})$ and $\phi \in C_c^{\infty}(U)$ there exists $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}$ (necessarily in $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp}$) of the form (5-2) such that $\varphi(t, n) = W^M(\eta_M((t w_0^{\mathbb{M}'} n)^*))$. It follows from Lemma 6.10 that the linear map $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp} \to C_c^{\infty}(T')$ given by $W \mapsto Y^{\psi}(W_{1/2}, \cdot)$ is onto. Therefore, by (5-3), the lemma amounts to the nonvanishing of the linear form $W^{\wedge} \mapsto Y^{\psi^{-1}}(M^*W^{\wedge}, t)$ on $\mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$ for some $t \in T'$. This follows from Theorem A.1 and the last part of Corollary 5.3. □

6E. By (5-4), Lemma 6.1 and Lemma 6.13, we conclude the following corollary.

Corollary 6.14. *Suppose that* $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}} M$ *is good and* $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ *is tempered. Then*

$$B\big(W, M\big(\tfrac{1}{2}\big)W^{\wedge}, \tfrac{1}{2}\big) = c_{\pi} A_e^{\psi}(M^*W) A_e^{\psi^{-1}}(M^*W^{\wedge})$$

for all $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp}$ *and* $W^{\wedge} \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$. *Moreover, the linear form* $A_e^{\psi}(M^*W)$ *does not vanish identically on* $\mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp}$.

In other words, taking into account the reduction step of Section 3F we have reduced Theorem 3.4 to the statement below.

Proposition 6.15. *Assume* $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}} M$ *is good and* $\tilde{\pi} = \mathcal{D}_{\psi^{-1}}(\pi)$ *is tempered. Then for any* $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^{\circ}_{\sharp}$ *and* $W^{\wedge} \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)$, *we have*

$$B\big(W, M\big(\tfrac{1}{2}\big)W^{\wedge}, \tfrac{1}{2}\big) = \epsilon_{\pi} A_e^{\psi}(M^*W) A_e^{\psi^{-1}}(M^*W^{\wedge}). \tag{6-12}$$

The proposition will eventually be proved in Section 8 after further reductions, using the results of [Lapid and Mao 2014; 2015b].

7. Applications of functional equations

In this section we use the functional equations established in [Lapid and Mao 2014] to analyze $B(W, M(s)W^{\wedge}, s)$.

7A. We first apply a functional equation proved in [Lapid and Mao 2014, Appendix B]. To that end we introduce a variant of B. We define $\underline{B}(W, W^{\vee}, s)$ to be the right-hand side of (6-11) whenever the integral defining $\{\cdot, \cdot\}$ and the double integrals over $V_M^{\sharp}\backslash N^{\sharp}$ and U are absolutely convergent.

Clearly for $g \in \mathbb{M}'$, with $|\det g| = 1$,

$$\{W(\cdot\,\eta_M(g)),\, W^\vee(\cdot\,\eta_M(g))\} = \{W, W^\vee\}.$$

Thus, $\underline{B}(W, W^\vee, s)$ is equal to

$$\int_{N'_{\mathbb{M}'}}\int_U\int_U \{W_s(\cdot\,\eta_M(n)w_U v),\, W^\vee_{-s}(\cdot\,w_U u)\}\psi_U(v)^{-1}\psi_U(u)\psi_{N'_{\mathbb{M}'}}(n)\,dv\,du\,dn$$
$$= \int_{N'_{\mathbb{M}'}}\int_U\int_U \{W_s(\cdot w_U v),\, W^\vee_{-s}(\cdot\eta_M(n)w_U u)\}\psi_U(v)^{-1}\psi_U(u)\psi_{N'_{\mathbb{M}'}}(n)^{-1}\,dv\,du\,dn$$
$$= \int_{V_M^\sharp\backslash N^\sharp}\int_U \{W_s(\cdot\,w_U v),\, W^\vee_{-s}(\cdot\,w_U v_2)\}\psi_U(v)^{-1}\psi_{N^\sharp}(v_2)\,dv\,dv_2. \tag{7-1}$$

By (6-11), for any $\pi \in \operatorname{Irr}_{\text{gen}} M$, $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, $W^\vee \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee))$ and $-\operatorname{Re} s \gg 1$, we have

$$\underline{B}(W, W^\vee, s) = B(W, W^\vee, s).$$

On the other hand, we have the following result.

Proposition 7.1 [Lapid and Mao 2014, Appendix B]. *Let $\pi \in \operatorname{Irr}_{\text{temp}} M$.*

(1) *For* $\operatorname{Re} s \gg 1$, $\underline{B}(W, W^\vee, s)$ *is well defined for any* $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee)\big)^\circ$.

(2) *For* $-\operatorname{Re} s \gg 1$, $\underline{B}(W, W^\vee, s)$ *is well defined for any* $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ$, $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi^\vee)\big)$.

(3) *For* $-\operatorname{Re} s \gg 1$, *we have*

$$\underline{B}(W, M(s)W^\wedge, s) = \underline{B}(M(s)W, W^\wedge, -s)$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ$, $W^\wedge \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)^\circ$.

Combined with the above we obtain:

Corollary 7.2. *For* $-\operatorname{Re} s \gg 1$, *we have*

$$B(W, M(s)W^\wedge, s) = \underline{B}(M(s)W, W^\wedge, -s)$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, $W^\wedge \in \operatorname{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)^\circ$.

7B. ***An identity of Whittaker functions on*** $\mathbf{GL_{2n}}$**.** A key fact in the formal argument for the case $\mathbf{n} = 1$ in [Lapid and Mao 2017, §7] was that for any unitarizable $\pi \in \operatorname{Irr}_{\text{gen}} \mathrm{GL}_2$, the expression

$$\int_{F^*} W\left(\begin{pmatrix} t & \\ & 1\end{pmatrix}\right) W^\vee\left(\begin{pmatrix} t & \\ & 1\end{pmatrix}\right) dt$$

defines a GL_2-invariant bilinear form on $\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi) \times \mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi^\vee)$.

In the general case, we encountered (in the definition of $\{\cdot,\cdot\}$) a similar integral

$$A_{\mathbf{n}}(W, W^\vee) = \int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(\eta_{\mathbb{M}}(g))W^\vee(\eta_{\mathbb{M}}(g))|\det g|^{1-\mathbf{n}}\,dg,$$

where $W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)$ and $W^\vee \in \mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi^\vee)$.

Let J be the subspace of $\mathrm{Mat}_{\mathbf{n}}$ consisting of the matrices whose first column is zero. Note that for any $g \in \mathbb{M}'$, $X \in \mathrm{Mat}_{\mathbf{n}}$, $Y \in J$, $n_1, n_2 \in N'_{\mathbb{M}'}$, we have

$$A_{\mathbf{n}}\big(\pi(\eta_{\mathbb{M}}(g)\ell_{\mathbb{M}}(X+Y)\eta^\vee_{\mathbb{M}}(n_1))W, \pi^\vee(\eta_{\mathbb{M}}(g)\ell_{\mathbb{M}}(X)\eta^\vee_{\mathbb{M}}(n_2))W^\vee\big) = \psi_{N'_{\mathbb{M}'}}(n_1 n_2^{-1})|\det g|^{\mathbf{n}-1}A_{\mathbf{n}}(W, W^\vee) \quad (7\text{-}2)$$

(with $\ell_{\mathbb{M}}$, $\eta_{\mathbb{M}}$ and $\eta^\vee_{\mathbb{M}}$ defined in Section 2A). We also have the following relation, for $w_{2\mathbf{n},\mathbf{n}} := \begin{pmatrix} & I_{\mathbf{n}} \\ I_{\mathbf{n}} & \end{pmatrix}$.

Theorem 7.3 [Lapid and Mao 2014, Theorem 1.3]. *Let $\pi \in \mathrm{Irr}_{\mathrm{temp}}\,\mathbb{M}$. Then for any $W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)$, $W^\vee \in \mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi^\vee)$, we have*

$$A_{\mathbf{n}}(W, W^\vee) = \int_J\int_J\int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(w_{2\mathbf{n},\mathbf{n}}\eta_{\mathbb{M}}(g)\ell_{\mathbb{M}}(X)) \times W^\vee(w_{2\mathbf{n},\mathbf{n}}\eta_{\mathbb{M}}(g)\ell_{\mathbb{M}}(Y))|\det g|^{\mathbf{n}-1}\,dg\,dX\,dY. \quad (7\text{-}3)$$

The integrals on both sides are absolutely convergent.

Note that it is not a priori clear that the right-hand side of (7-3) satisfies (7-2) for $X \in \mathrm{Mat}_{\mathbf{n}} - J$.

We can slightly rephrase Theorem 7.3 as follows. First, we write the right-hand side of (7-3) as

$$\int_J\int_J\int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(\eta^\vee_{\mathbb{M}}(g)w_{2\mathbf{n},\mathbf{n}}\ell_{\mathbb{M}}(X))W^\vee(\eta^\vee_{\mathbb{M}}(g)w_{2\mathbf{n},\mathbf{n}}\ell_{\mathbb{M}}(Y))|\det g|^{\mathbf{n}-1}\,dg\,dX\,dY.$$

We may multiply $w_{2\mathbf{n},\mathbf{n}}$ on the left by any element $x \in \eta^\vee_{\mathbb{M}}(\mathbb{M}')$ with $|\det x| = 1$. In particular, we take $w'_{2\mathbf{n},\mathbf{n}} := \begin{pmatrix} & I_{\mathbf{n}} \\ w_0^{\mathbb{M}'} & \end{pmatrix}$ instead of $w_{2\mathbf{n},\mathbf{n}}$. Let $\boldsymbol{\epsilon}_2 \in \ell_{\mathbb{M}}(\epsilon_{1,1} + J)$, where $\epsilon_{1,1}$ is the matrix in $\mathrm{Mat}_{\mathbf{n}}$ with 1 in the upper left corner and zero elsewhere. Since

$$A_{\mathbf{n}}(\pi(\boldsymbol{\epsilon}_2)W, \pi^\vee(\boldsymbol{\epsilon}_2)W^\vee) = A_{\mathbf{n}}(W, W^\vee),$$

we infer that $A_{\mathbf{n}}(W, W^\vee)$ equals

$$\int_J\int_J\int_{N'_{\mathbb{M}'}\backslash\mathbb{M}'} W(\eta^\vee_{\mathbb{M}}(g)w'_{2\mathbf{n},\mathbf{n}}\ell_{\mathbb{M}}(X)\boldsymbol{\epsilon}_2) \times W^\vee(\eta^\vee_{\mathbb{M}}(g)w'_{2\mathbf{n},\mathbf{n}}\ell_{\mathbb{M}}(Y)\boldsymbol{\epsilon}_2)|\det g|^{\mathbf{n}-1}\,dg\,dX\,dY.$$

Using Bruhat decomposition for g, the integral can be rewritten as

$$\int_J \int_J \int_{N'_{\mathbb{M}'}} \int_{T'_{\mathbb{M}'}} W(\eta^\vee_{\mathbb{M}}(t) w'_{2\mathbf{n},\mathbf{n}} \eta_{\mathbb{M}}(n) \ell_{\mathbb{M}}(X) \boldsymbol{\epsilon}_2) \\ \times W^\vee(\eta^\vee_{\mathbb{M}}(t) w'_{2\mathbf{n},\mathbf{n}} \eta_{\mathbb{M}}(n) \ell_{\mathbb{M}}(Y) \boldsymbol{\epsilon}_2) |\det t|^{\mathbf{n}-1} \delta_{B'_{\mathbb{M}'}}(t)^{-1}\, dt\, dn\, dX\, dY.$$

Since $\ell_{\mathbb{M}}(\epsilon_{1,1}+J)$ is invariant under conjugation by $\eta_{\mathbb{M}}(N'_{\mathbb{M}'})$, by changing variables in X and Y we get

$$A_{\mathbf{n}}(W, W^\vee) = \int_J \int_J \int_{N'_{\mathbb{M}'}} \int_{T'_{\mathbb{M}'}} W(\eta^\vee_{\mathbb{M}}(t) \boldsymbol{\epsilon}_3 \ell_{\mathbb{M}}(X) \eta_{\mathbb{M}}(n)) \\ \times W^\vee(\eta^\vee_{\mathbb{M}}(t) \boldsymbol{\epsilon}_3 \ell_{\mathbb{M}}(Y) \eta_{\mathbb{M}}(n)) |\det t|^{\mathbf{n}-1} \delta_{B'_{\mathbb{M}'}}(t)^{-1}\, dt\, dn\, dX\, dY, \quad (7\text{-}4)$$

where $\boldsymbol{\epsilon}_3 = w'_{2\mathbf{n},\mathbf{n}} \boldsymbol{\epsilon}_2$.

7C. In the first expression of (7-1) for $\underline{B}$, we have an inner integral of the form

$$\int_{N'_{\mathbb{M}'}} \{W_s(\cdot\, \eta_M(n)), W^\vee_{-s}\} \psi_{N'_{\mathbb{M}'}}(n)\, dn.$$

This leads us to apply the above functional equation in the following setting. Let $\pi \in \mathrm{Irr}_{\mathrm{temp}}\, \mathbb{M}$. Define the bilinear form $D(W, W^\vee)$ on $\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi) \times \mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi^\vee)$ by

$$D(W, W^\vee) := \int_{N'_{\mathbb{M}'}} A_{\mathbf{n}}(\pi(\eta_{\mathbb{M}}(n)) W, W^\vee) \psi_{N'_{\mathbb{M}'}}(n)\, dn. \quad (7\text{-}5)$$

It is shown in [Lapid and Mao 2014, Appendix A] that the integral is absolutely convergent.

Remark 7.4. For $(W, W^\vee) \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi)) \times \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee)\big)$, we have

$$D(\delta_P^{-1/2} W \circ \varrho, \delta_P^{-1/2} W^\vee \circ \varrho) = D(\delta_P^{-1/2} W_s \circ \varrho, \delta_P^{-1/2} W^\vee_{-s} \circ \varrho) \\ = \int_{N'_{\mathbb{M}'}} \{W_s(\cdot\, \eta_M(n)), W^\vee_{-s}\} \psi_{N'_{\mathbb{M}'}}(n)\, dn. \quad (7\text{-}6)$$

For convenience, set

$$\bar{R} = \boldsymbol{\epsilon}_3 (\eta_{\mathbb{M}}(N'_{\mathbb{M}'}) \ltimes \ell_{\mathbb{M}}(J)) \boldsymbol{\epsilon}_3^{-1} = w'_{2\mathbf{n},\mathbf{n}} (\eta_{\mathbb{M}}(N'_{\mathbb{M}'}) \ltimes \ell_{\mathbb{M}}(J)) w'^{-1}_{2\mathbf{n},\mathbf{n}} \\ = \left\{ \left(\begin{smallmatrix} I_{\mathbf{n}} & \\ x & n^t \end{smallmatrix} \right) : x \in J, n \in N'_{\mathbb{M}'} \right\} \subset N^t_{\mathbb{M}} \cap \mathcal{P}^*.$$

Definition 7.5. Let $\mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi)_\natural$ be the linear subspace of $\mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi)$ consisting of W such that

$$W(\cdot\, \boldsymbol{\epsilon}_3)|_{\mathcal{P}^*} \in C_c^\infty(N_{\mathbb{M}} \backslash \mathcal{P}^*, \psi^{-1}_{N_{\mathbb{M}}}) \quad \text{and} \quad W(\cdot\, \boldsymbol{\epsilon}_3)|_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'}) \ltimes \bar{R}} \in C_c^\infty(\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'}) \ltimes \bar{R}).$$

It is easy to see that this definition is independent of the choice of $\boldsymbol{\epsilon}_2$ (and correspondingly $\boldsymbol{\epsilon}_3$).

We note that $\mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi)_\natural$ is nonzero, since the restriction of $\mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi)$ to $\mathcal{P}^*$ contains $C_c^\infty(N_{\mathbb{M}}\backslash\mathcal{P}^*, \psi_{N_{\mathbb{M}}}^{-1})$ and $\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}) \ltimes \bar{R} \subset B_{\mathbb{M}}^t \cap \mathcal{P}^*$. On the other hand, even if we assume that π is supercuspidal, $\mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi)_\natural$ is a proper subspace of $\mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi)$ since the set $N_{\mathbb{M}} \cdot (\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}) \ltimes \bar{R})$ is not closed.

Let $N_{\mathbb{M}}^\flat = (N_{\mathbb{M}}^\sharp)^*$ and let $\psi_{N_{\mathbb{M}}^\flat}$ be the character on $N_{\mathbb{M}}^\flat$ given by

$$\psi_{N_{\mathbb{M}}^\flat}(m) = \psi_{N_{\mathbb{M}}^\sharp}(m^*).$$

Note that $N_{\mathbb{M}}^\flat$ consists of the elements of $N_{\mathbb{M}}$ whose $(\mathbf{n}+1)$-st column vanishes above the diagonal.

Proposition 7.6. *Let $\pi \in \operatorname{Irr}_{\mathrm{temp}} \mathbb{M}$. For any $W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)$ and $W^\vee \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}^{-1}}(\pi^\vee)_\natural$, $D(W, W^\vee)$ is equal to the absolutely convergent integral*

$$\int_{\eta_{\mathbb{M}}^\vee(N'_{\mathbb{M}'})\backslash N_{\mathbb{M}}^\flat} \int_{\eta_{\mathbb{M}}^\vee(N'_{\mathbb{M}'})\backslash N_{\mathbb{M}}^\flat} \int_{T'_{\mathbb{M}'}} W(\eta_{\mathbb{M}}^\vee(t)\boldsymbol{\epsilon}_3 r_1) W^\vee(\eta_{\mathbb{M}}^\vee(t)\boldsymbol{\epsilon}_3 r_2) \\ \times |\det t|^{\mathbf{n}-1} \delta_{B'_{\mathbb{M}'}}(t)^{-1} \psi_{N_{\mathbb{M}}^\flat}(r_1^{-1} r_2)\, dt\, dr_2\, dr_1,$$

where the integrand is compactly supported in all variables.

Proof. From (7-4) and (7-5), we get that $D(W, W^\vee)$ is equal to

$$\int_{N'_{\mathbb{M}'}} \left(\int_J \int_J \int_{N'_{\mathbb{M}'}} \int_{T'_{\mathbb{M}'}} W(\eta_{\mathbb{M}}^\vee(t)\boldsymbol{\epsilon}_3 \ell_{\mathbb{M}}(X) \eta_{\mathbb{M}}(n_2 n_1)) W^\vee(\eta_{\mathbb{M}}^\vee(t)\boldsymbol{\epsilon}_3 \ell_{\mathbb{M}}(Y) \eta_{\mathbb{M}}(n_2)) \\ \times |\det t|^{\mathbf{n}-1} \delta_{B'_{\mathbb{M}'}}(t)^{-1}\, dt\, dn_2\, dX\, dY \right) \psi_{N'_{\mathbb{M}'}}(n_1)\, dn_1.$$

By the condition on $W^\vee$, the integrand is compactly supported in t, Y and n_2. Since $\eta_{\mathbb{M}}^\vee(t)$ normalizes $\bar{R}$, it is also compactly supported in X, n_1 in view of Lemma 7.7 below. A change of variable $n_1 \mapsto n_2^{-1} n_1$ gives the identity in the proposition. □

Lemma 7.7. *The restriction to $N_{\mathbb{M}}^t \cap (\mathcal{P} \cup \mathcal{P}^*)$ of any $W \in C^{\mathrm{sm}}(N_{\mathbb{M}}\backslash\mathbb{M}, \psi_{N_{\mathbb{M}}})$ is compactly supported.*

Proof. This is standard. By passing to $W(\cdot^*)$, it is enough to consider the support in $N_{\mathbb{M}}^t \cap \mathcal{P}$. Let $n = uak$ with $a = \operatorname{diag}(a_1, \dots, a_{2\mathbf{n}})$ be the Iwasawa decomposition of $n \in N_{\mathbb{M}}^t \cap \mathcal{P}$. Then $a_{2\mathbf{n}} = 1$ and $|n_{i,j}| \le |a_i \dots a_{2\mathbf{n}}|$ for all $1 \le j \le i \le 2\mathbf{n}$. Thus, if $W(n) \ne 0$ then the a_i are bounded in terms of W, and hence n is bounded. The lemma follows. □

7D. We apply Proposition 7.6 to get a new expression for $\underline{B}(W, W^\vee, s)$. Before stating the result, we first introduce an integral that appears in the new expression.

As in [Lapid and Mao 2015b, §7.4], define

$$\Delta(t) := |t_1|^{-\mathbf{n}} \delta_B^{1/2}(\varrho(t)), \quad t = \operatorname{diag}(t_1, \dots, t_{2\mathbf{n}}) \in T_{\mathbb{M}}.$$

In particular, when $t = \eta_{\mathbb{M}}^\vee(t')$ with $t' \in T'_{\mathbb{M}'}$, we have $\Delta(t) = \delta_{B'}(\varrho'(t))^{1/2}$.

For now, let ϵ_4 be an arbitrary element in $N_{\mathbb{M}}$. (We will fix it in the next section in order to make the formulas look nicer.) Let $T'' = \eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'}) \times Z_{\mathbb{M}}$. For $W \in C^{\mathrm{sm}}(N\backslash G, \psi_N)$, $t \in T''$, let

$$\begin{aligned} E^\psi(W,t) &:= \Delta(t)^{-1} \int_{\eta_M(N'_{\mathbb{M}'})\backslash N^\sharp} W(\varrho(t\epsilon_4\epsilon_3)w_U v)\psi_{N^\sharp}(v)^{-1}\,dv \\ &= \Delta(t)^{-1}\psi_{N_{\mathbb{M}}}(t\epsilon_4 t^{-1}) \int_{\eta_M(N'_{\mathbb{M}'})\backslash N^\sharp} W(\varrho(t\epsilon_3)w_U v)\psi_{N^\sharp}(v)^{-1}\,dv \\ &= \Delta(t)^{-1}\psi_{N_{\mathbb{M}}}(t\epsilon_4 t^{-1}) \int_U \int_{\bar R} W(\varrho(tv\epsilon_3)w_U u) \\ &\qquad\qquad \times \psi_{N^\flat_{\mathbb{M}}}(\epsilon_3^{-1} v\epsilon_3)^{-1}\psi_U(u)^{-1}\,dv\,du, \qquad (7\text{-}7) \end{aligned}$$

provided that the integral converges. (It is easy to check that the integrand above is invariant under $\eta_M(N'_{\mathbb{M}'})$ when $t \in T''$.) Similarly, define $E^{\psi^{-1}}(W^\wedge, t)$ for $W^\wedge \in C^{\mathrm{sm}}(N\backslash G, \psi_N^{-1})$.

Note that for any $t \in T''$, $E^\psi(W,t)$ is $(N^\sharp, \psi_{N^\sharp})$-equivariant in W. It is also clear from the definition that whenever defined,

$$E^\psi(W_s, tz) = E^\psi(W_s, t)|\det z|^{s+1/2}\omega_\pi(z), \quad z \in Z_{\mathbb{M}},\ t \in T'', \qquad (7\text{-}8)$$

for $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ with $\pi \in \mathrm{Irr}_{\mathrm{gen}}\,\mathbb{M}$, where ω_π is the central character of π.

Definition 7.8. Let $\mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural$ be the linear subspace of $\mathrm{Ind}(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi))^\circ$ spanned by the functions which vanish outside $Pw_U N$ and on the big cell are given by

$$W(u'mw_U u) = \delta_P^{1/2}(m)W^M(m)\phi(u), \quad m \in M,\ u, u' \in U,$$

with $\phi \in C_c^\infty(U)$ and $W^M \circ \varrho \in \mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi)_\natural$.

This space is clearly nonzero since $\mathbb{W}^{\psi^{-1}_{N_{\mathbb{M}}}}(\pi)_\natural$ is nonzero.

Lemma 7.9. *Let $\pi \in \mathrm{Irr}_{\mathrm{gen}}\, M$. For $\mathrm{Re}\, s \gg 1$, the integral (7-7) defining $E^\psi(W_s, t)$ converges for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ and $t \in T''$ uniformly for (s,t) in a compact set. Hence, $E^\psi(W_s,t)$ is holomorphic for $\mathrm{Re}\, s \gg 1$ and continuous in t. If $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ$ then $E^{\psi^{-1}}(W_s^\wedge, t)$ is entire in s and locally constant in t, uniformly in s. If moreover $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural$ then $E^{\psi^{-1}}(W_s^\wedge, t)$ is compactly supported in $t \in \eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})$, uniformly in s.*

Proof. For the first assertion, since $T_{\mathbb{M}}$ normalizes the group $\bar R$ and U is normalized by M, we need to check the convergence, locally uniformly in $m \in M$, of

$$\int_{\bar R}\int_U |W_s(\varrho(v)w_U um)|\,du\,dv.$$

This expression is clearly locally constant in m, so we can assume $m=e$. Since $\bar{R} \subset N^t_{\mathbb{M}} \cap \mathcal{P}^*$, the integral above is a "partial integration" of the double integral (6-6b) in Lemma 6.8 (with $g=e$), thus converges by Remark 2.2.

By Lemma 7.7, when $W^\wedge \in \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ$, the integrand in the last expression of (7-7) (with $W^\wedge_s$ in place of W) is compactly supported in v and u, uniformly in s and locally uniformly in t. It is then clear that $E^{\psi^{-1}}(W^\wedge_s, t)$ is locally constant in t. The second part follows. Furthermore, if $W^\wedge \in \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural$, it is clear from the definition of the latter space that the integrand in (7-7) (with $W^\wedge$ in place of W) is compactly supported in v and $t \in \eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})$. The lemma follows. □

7E.

Proposition 7.10. *Let $\pi \in \operatorname{Irr}_{\mathrm{temp}} M$. Then for* $\operatorname{Re} s \gg 1$, *we have*

$$\underline{B}(W, W^\vee, s) = \int_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})} E^{\psi}(W_s, t) E^{\psi^{-1}}(W^\vee_{-s}, t)\, \frac{dt}{|\det t|} \tag{7-9}$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$ and $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee)\big)^\circ_\natural$, where (by Lemma 7.9) *the integrand on the right-hand side is continuous and compactly supported.*

Proof. First note that the element ϵ_4 has no effect on the validity of (7-9), so we may ignore it from the consideration. By Proposition 7.1, if $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee)\big)^\circ$, the integrals defining $\underline{B}(W, W^\vee, s)$ are absolutely convergent when $\operatorname{Re} s \gg 1$. If moreover $W^\vee \in \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi^\vee)\big)^\circ_\natural$, then by the relation (7-6), (the first expression of) (7-1) and Proposition 7.6, $\underline{B}(W, W^\vee, s)$ is equal to

$$\begin{aligned}\int_U \int_U \bigg(\int_{T'_{\mathbb{M}'}} \int_{\eta^\vee_{\mathbb{M}}(N'_{\mathbb{M}'})\backslash N^\flat_{\mathbb{M}}} \int_{\eta^\vee_{\mathbb{M}}(N'_{\mathbb{M}'})\backslash N^\flat_{\mathbb{M}}} & W_s(\varrho(\eta^\vee_{\mathbb{M}}(t)\epsilon_3 r_1) w_U u_1) \\ & \times W^\vee_{-s}(\varrho(\eta^\vee_{\mathbb{M}}(t)\epsilon_3 r_2) w_U u_2) \delta_{B'_{\mathbb{M}'}}(t)^{-1} \delta_P(\eta^\vee_M(t))^{-1} |\det t|^{\mathbf{n}-1} \\ & \times \psi_U(u_1)^{-1} \psi_{N^\flat_{\mathbb{M}}}(r_1)^{-1} \psi_U(u_2) \psi_{N^\flat_{\mathbb{M}}}(r_2)\, dr_1\, dr_2\, dt \bigg)\, du_1\, du_2.\end{aligned}$$

Using the fact that for $t \in T'_{\mathbb{M}'}$

$$\delta_{B'_{\mathbb{M}'}}(t)^{-1} \delta_P(\eta^\vee_M(t))^{-1} |\det t|^{\mathbf{n}-1} = \delta_{B'}(\varrho'(t))^{-1} |\det t|^{-1},$$

to finish the proof of Proposition 7.10 it suffices to show the convergence of

$$\begin{aligned}\int_{T'_{\mathbb{M}'}} \bigg(\int_{\eta_M(N'_{\mathbb{M}'})\backslash N^\sharp} & |W_s(\varrho(\eta^\vee_{\mathbb{M}}(t)\epsilon_3) w_U v_1)|\, dv_1 \\ & \times \int_{\eta_M(N'_{\mathbb{M}'})\backslash N^\sharp} |W^\vee_{-s}(\varrho(\eta^\vee_{\mathbb{M}}(t)\epsilon_3) w_U v_2)|\, dv_2 \bigg) \delta_{B'}(\varrho'(t))^{-1} \frac{dt}{|\det t|}.\end{aligned}$$

It follows from Lemma 7.9 that the integrals over v_1 and v_2 converge for a fixed t and that the integral over v_2 is compactly supported as a function of t. The argument in Lemma 7.9 also shows that the resulting integrals over v_1 and v_2 are smooth in t. Thus, the integral converges. □

Combining Proposition 7.10 with Corollary 7.2 and (7-8) we get the following.

Proposition 7.11. *Let $\pi \in \mathrm{Irr}_{\mathrm{temp}} M$. Then for $-\operatorname{Re} s \gg 1$ and any*

$$W \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}}(\pi)\big)^{\circ}_{\sharp},$$
$$W^{\wedge} \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)^{\circ}_{\natural},$$

we have

$$B(W, M(s)W^{\wedge}, s) = \int_{Z_{\mathbb{M}}\backslash T''} E^{\psi}(M_s^* W, t) E^{\psi^{-1}}(W_s^{\wedge}, t) \frac{dt}{|\det t|}, \tag{7-10}$$

where the integrand is continuous and compactly supported.

8. Proof of Proposition 6.15

Proposition 7.11 is an important (and hard-earned) step towards the proof of Proposition 6.15. However, it is still insufficient. First, it is not valid at the point $s = \frac{1}{2}$, which is pertinent for the left-hand side of (6-12). Second, the functions E^{ψ} have to be investigated in order to match the right-hand side of (6-12). Also, so far we have only used the fact that $\pi \in \mathrm{Irr}_{\mathrm{meta}} \mathbb{M}$ to reduce to special sections, but it should be clear from the nature of our formula that this fact has to enter the computation in a more substantial and quantitative way. Fortunately, these issues were taken up in [Lapid and Mao 2015b, §11]. We recall these results and explain how they are used to conclude the proof of Proposition 6.15 from Proposition 7.11.

8A. Let $\mathfrak{d} = \mathrm{diag}(1, -1, \dots, (-1)^{\mathbf{n}-1}) \in \mathrm{Mat}_{\mathbf{n}}$. We now fix

$$\boldsymbol{\epsilon}_4 = \ell_{\mathbb{M}}\big(-\tfrac{1}{2}\mathfrak{d} w_0^{\mathbb{M}'}\big) \in N_{\mathbb{M}}.$$

This element is denoted by ϵ' in [Lapid and Mao 2015b, p. 9550] with the parameter $\mathfrak{a} = -\frac{1}{2}$ in the notation of that paper. We also fix $\boldsymbol{\epsilon}_2 = \ell_{\mathbb{M}}(\mathfrak{d})$ (and correspondingly $\boldsymbol{\epsilon}_3 = w'_{2\mathbf{n},\mathbf{n}}\boldsymbol{\epsilon}_2$; see Section 7B). The reason for fixing the elements this way is (partly) due to the special form of the character $\psi_{\overline{U}}$ on $\overline{U}$ given by

$$\psi_{\overline{U}}(\bar{v}) = \psi_U((\varrho(\boldsymbol{\epsilon}_4\boldsymbol{\epsilon}_3)w_U)^{-1}\bar{v}\varrho(\boldsymbol{\epsilon}_4\boldsymbol{\epsilon}_3)w_U)^{-1} = \psi(\bar{v}_{2\mathbf{n}+1,1}), \quad \bar{v} \in \overline{U}. \tag{8-1}$$

We rewrite the first expression on the right-hand side of (7-7) by splitting the integral into integrals over U and $\eta_M(N'_{\mathbb{M}'})\backslash N^{\sharp}_M$, making the change of variables

$$v \mapsto (\varrho(\boldsymbol{\epsilon}_4\boldsymbol{\epsilon}_3)w_U)^{-1}\bar{v}\varrho(\boldsymbol{\epsilon}_4\boldsymbol{\epsilon}_3)w_U$$

in the integral over U, and conjugating the variable in the second integral by w_U. We get, for $\operatorname{Re} s \gg 1$,

$$E^{\psi}(W_s, t) = \Delta(t)^{-1} \int_{\eta^{\vee}_{\mathbb{M}}(N'_{\mathbb{M}'})\backslash N^{\flat}_{\mathbb{M}}} \int_{\overline{U}} W_s(\varrho(t)\bar{v}\varrho(\boldsymbol{\epsilon}_4\boldsymbol{\epsilon}_3 r)w_U)\psi_{\overline{U}}(\bar{v})\psi_{N^{\flat}_{\mathbb{M}}}(r)^{-1}\, d\bar{v}\, dr$$
$$= \Delta(t)^{-1} \int_{\bar{R}} \int_{\overline{U}} W_s(\varrho(t)\bar{v}\varrho(\boldsymbol{\epsilon}_4 r\boldsymbol{\epsilon}_3)w_U)\psi_{\overline{U}}(\bar{v})\psi_{\bar{R}}(r)\, d\bar{v}\, dr, \qquad (8\text{-}2)$$

where $\psi_{\bar{R}}(r) = \psi_{N^{\flat}_{\mathbb{M}}}(\boldsymbol{\epsilon}_3^{-1} r \boldsymbol{\epsilon}_3)^{-1}$.

We now quote the first pertinent result from [Lapid and Mao 2015b, §11]. Let S be the subtorus $\prod_{i=1}^{\mathbf{n}-1} T_i$ of T'' (of codimension two), where T_i is the one-dimensional torus

$$T_i := \{\operatorname{diag}(\overbrace{z^{-1}, \dots, z^{-1}}^{2\mathbf{n}-i}, \overbrace{z, \dots, z}^{i}) : z \in F^*\}.$$

For any $f \in C_c^{\infty}(S)$ and $g \in C(T'')$, we write $f * g(\cdot) = \int_S f(t) g(\cdot\, t)\, dt$.

Theorem 8.1. *For any $W \in C^{\mathrm{sm}}(N\backslash G, \psi_N)$ which is left-invariant under a compact open subgroup of Z_M and any $f \in C_c^{\infty}(S)$, the function $f * E^{\psi}(W_s, t)$ extends to an entire function in s which is locally constant in t, uniformly in s. Moreover, if $\pi \in \operatorname{Irr}_{\mathrm{meta,temp}} \mathbb{M}$ then*

$$f * E^{\psi}(M_s^* W, t)\big|_{s=1/2} = \epsilon_{\pi}^{\mathbf{n}} A_e^{\psi}(M^* W) \int_S f(t')\, dt'$$

for any $W \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))$, $t \in T''$ and $f \in C_c^{\infty}(S)$.

The first part is [Lapid and Mao 2015b, Corollary 11.9], except for a slight correction which we now explain. Recall some notation from that paper. As in Section 6 of [ibid.] let $\mathcal{Z}$ be the unipotent subgroup of $\mathbb{M}$ given by

$$\mathcal{Z} = \{m \in \mathbb{M} : m_{i,i} = 1\ \forall i,\ m_{i,j} = 0 \text{ if either } (j > i \text{ and } i + j > 2\mathbf{n}) \text{ or } (i > j \text{ and } i + j \le 2\mathbf{n} + 1)\},$$

and let $\psi_{\mathcal{Z}}$ be its character

$$\psi_{\mathcal{Z}}(m) = \psi(m_{1,2} + \cdots + m_{\mathbf{n}-1,\mathbf{n}} - m_{\mathbf{n}+2,\mathbf{n}+1} - \cdots - m_{2\mathbf{n},2\mathbf{n}-1}), \quad m \in \mathcal{Z}. \qquad (8\text{-}3)$$

The group $\varrho(\mathcal{Z})$ stabilizes the character $\psi_{\overline{U}}$. Let $\mathfrak{E} = \varrho(\mathcal{Z}) \ltimes \overline{U}$ with the character

$$\psi_{\mathfrak{E}}(\varrho(m)\bar{u}) = \psi_{\mathcal{Z}}(m)\psi_{\overline{U}}^{-1}(\bar{u}), \quad m \in \mathcal{Z},\ \bar{u} \in \overline{U}.$$

Also, let $\mathcal{Z}^+ = \mathcal{Z} \cap N_{\mathbb{M}}$, $V_{\Delta} = \mathcal{Z} \cap N_{\mathbb{M}}^t$ and

$$N_{\mathbb{M},\Delta} = \{\ell_{\mathbb{M}}(X)^t : X \in J,\ X_{i,j} = 0 \text{ if } i + j > \mathbf{n} + 1\}.$$

We have $\mathcal{Z} = \mathcal{Z}^+ \cdot V_{\Delta}$ and $\bar{R} = V_{\Delta} \cdot N_{\mathbb{M},\Delta}$. In Section 10.4 of [ibid.] it was erroneously claimed that the second product is semidirect. Fortunately, this does not have any bearing on the argument. All what matters is that the character $\psi_{\bar{R}}$ is given by $\psi_{\bar{R}}(vn) = \psi_{V_\Delta}(v)$, $v \in V_\Delta$, $n \in N_{\mathbb{M},\Delta}$, where $\psi_{V_\Delta}(r)$ is defined right after (7.11) of [ibid.]. Thus, the middle integral in (11.11) of [ibid.] should be taken over $V_\Delta \cdot N_{\mathbb{M},\Delta}$, i.e., over $\bar{R}$ in our notation. (Also, the variable v should be replaced by r.) With this correction, the expression in (11.11) of [ibid.], evaluated at $\varrho(\epsilon_3)w_U$, is $(f\Delta) * E^\psi(W_s, t)$ and its analytic continuation is given by the expression in (11.10) of [ibid.], which amounts to a finite sum.

Meanwhile, from the definition of A_e^ψ in (6-1) we have

$$A_e^\psi(W) = \int_{\varrho(V_{\mathbb{M}}^*)\backslash\gamma V_- \gamma^{-1}} W(v\gamma\epsilon_1)\psi_{V_-}((\gamma\epsilon_1)^{-1}v\gamma\epsilon_1)^{-1}\,dv.$$

We can integrate instead over the group $\overline{U}^\urcorner$ in Section 11.5 of [ibid.], consisting of elements $\bar{u}$ in $\overline{U}$ such that $\bar{u}_{2\mathbf{n}+i,j} = 0$ whenever $i \geq \mathbf{n}$ and $j \leq \mathbf{n}$. The character $\hat{\psi}_{\overline{U}^\urcorner}(\bar{u}) := \psi_{V_-}((\gamma\epsilon_1)^{-1}\bar{u}\gamma\epsilon_1)^{-1}$ on $\overline{U}^\urcorner$ also matches the one defined in Sections 11.3 and 11.5 of [ibid.] (with the parameter $\mathfrak{a} = -\frac{1}{2}$). Thus, $A_e^\psi(W)$ is $\mathcal{T}'(W)(\gamma\epsilon_1)$ in the notation there. The second part of Theorem 8.1 amounts to the first statement of [ibid., Corollary 11.10] upon taking

$$\epsilon_1 = (\hat{w}\gamma)^{-1}\varrho(\epsilon_3)w_U = \ell_M((-1)^{\mathbf{n}}\mathfrak{d}),$$

where

$$\hat{w} := \eta(w'_{U'})\varrho(w'_{2\mathbf{n},\mathbf{n}}) = \begin{pmatrix} & I_{\mathbf{n}} & & \\ & & & w_{\mathbf{n}} \\ -w_{\mathbf{n}} & & & \\ & & I_{\mathbf{n}} & \end{pmatrix}$$

is as in the bottom of p. 9523 of [ibid.].

Theorem 8.1 is a crucial step. Recall that the Langlands quotient of

$$\mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}}(\pi), \tfrac{1}{2}\big),$$

i.e., the image under M^*, admits a unique $(N^\sharp, \psi_{N^\sharp})$-equivariant functional. Thus, $E^\psi(M^*W, t)$ (which is technically not defined) is a priori proportional to $A_e^\psi(M^*W)$ and the constant of proportionality is a function depending on t. Theorem 8.1 essentially says that in fact this function is a constant which can be computed explicitly. The main input is that the Langlands quotient admits a realization in $C^{\mathrm{sm}}(H\backslash G)$, where $H \simeq \mathrm{Sp}_{\mathbf{n}} \times \mathrm{Sp}_{\mathbf{n}}$ is the centralizer of $\varrho(E)$ in G. (Recall that $E = \mathrm{diag}(1, -1, \dots, 1, -1)$.) We refer to [Lapid and Mao 2015b] for more details.

We may thus conclude the following corollary.

Corollary 8.2. *Suppose that $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}}\,\mathbb{M}$. Then for any $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp$, $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)^\circ_\natural$ we have*

$$B\big(W, M(\tfrac{1}{2})W^\wedge, \tfrac{1}{2}\big) = \epsilon_\pi^{\mathbf{n}} A_e^\psi(M^*W) \int_{Z_{\mathbb{M}}\backslash T''} E^{\psi^{-1}}(W^\wedge_{1/2}, t)\,\frac{dt}{|\det t|}. \tag{8-4}$$

Proof. By Lemma 7.9, for any $W^\wedge \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}^{-1}}(\pi)\big)_\natural^\circ$ there exists $K_0 \in \mathcal{CSGR}(T'')$ such that $E^{\psi^{-1}}(W_s^\wedge, \cdot) \in C(T'')^{K_0}$ for all s and $E^{\psi^{-1}}(W_s^\wedge, \cdot)$ is compactly supported modulo $\mathbb{Z}_{\mathbb{M}}$ uniformly in s. Suppose that $f \in C_c^\infty(S)$ is supported in $S \cap K_0$ and let $f^\vee(t) := f(t^{-1})$. By Proposition 7.11, for $-\mathrm{Re}\, s \gg 1$ and $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_\sharp^\circ$ we have

$$B(W, M(s)W^\wedge, s) \int_S f(t)\, dt$$
$$= \int_{Z_{\mathbb{M}}\backslash T''} E^\psi(M_s^* W, t) f^\vee * E^{\psi^{-1}}(W_s^\wedge, t) \frac{dt}{|\det t|}$$
$$= \int_{Z_{\mathbb{M}}\backslash T''} f * E^\psi(M_s^* W, t) E^{\psi^{-1}}(W_s^\wedge, t) \frac{dt}{|\det t|}. \qquad (8\text{-}5)$$

From the first part of Theorem 8.1 (i.e., [Lapid and Mao 2015b, Corollary 11.9]) we infer that both sides of (8-5) are meromorphic functions in s and the identity holds whenever $M(s)$ is holomorphic. Specializing to $s = \frac{1}{2}$ and using the second part of Theorem 8.1 (i.e., [ibid., Corollary 11.10]) the corollary follows. □

8B. It remains to compute the integral on the right-hand side of (8-4). Again, this is essentially done in [Lapid and Mao 2015b] and it relies heavily on the fact that $\pi \in \mathrm{Irr}_{\mathrm{meta}}\, \mathbb{M}$.

Let

$$\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural\natural} = \{W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi) : W|_{\mathcal{P}^*} \in C_c^\infty(N_{\mathbb{M}}\backslash\mathcal{P}^*, \psi_{N_{\mathbb{M}}}) \text{ and}$$
$$W|_{\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}) \ltimes \mathcal{Z}} \in C_c^\infty(\mathcal{Z}^+\backslash\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'}) \ltimes \mathcal{Z}, \psi_{\mathcal{Z}})\}.$$

Then $\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural\natural}$ is invariant under $\pi(\boldsymbol{\epsilon}_4)$ and $\pi(u\boldsymbol{\epsilon}_3)W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural\natural}$ for any $W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_\natural$ and $u \in N_{\mathbb{M},\Delta}$. Note that $\boldsymbol{\epsilon}_4$ normalizes $\mathcal{Z}$, $\mathcal{Z}^+$ and $\psi_{\mathcal{Z}}$ (by direct computation). Similarly $\boldsymbol{\epsilon}_3(w'_{2\mathbf{n},\mathbf{n}})^{-1}$ stabilizes $(\mathcal{Z}, \psi_{\mathcal{Z}})$.

Theorem 8.3. *Let $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}}\, \mathbb{M}$. Then*

(1) [Lapid and Mao 2015b, Proposition 3.2; Matringe 2015]. *The integral*

$$\mathfrak{P}^{H_{\mathbb{M}}}(W) := \int_{N_{\mathbb{M}} \cap H_{\mathbb{M}} \backslash \mathcal{P} \cap H_{\mathbb{M}}} W(p)\, dp$$

converges and defines a nonzero $H_{\mathbb{M}}$-invariant functional on $\mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)$.

(2) [Lapid and Mao 2015b, Proposition 11.1]. *For any $W \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural\natural}$ we have*

$$\int_{\mathcal{Z}^+\backslash\mathcal{Z}} \int_{\eta_{\mathbb{M}}^\vee(T'_{\mathbb{M}'})} \Delta(t)^{-1} |\det t|^{\mathbf{n}} W(tr) \psi_{\mathcal{Z}}(r)^{-1}\, dt\, dr$$
$$= \int_{\mathcal{Z} \cap H_{\mathbb{M}} \backslash \mathcal{Z}} \mathfrak{P}^{H_{\mathbb{M}}}(\pi(n)W) \psi_{\mathcal{Z}}(n)^{-1}\, dn.$$

(3) [Lapid and Mao 2015b, Lemma 4.3]. *The integral*

$$L_W(g) := \int_{P\cap H\backslash H} \int_{N_{\mathbb{M}}\cap H_{\mathbb{M}}\backslash \mathcal{P}\cap H_{\mathbb{M}}} W(\varrho(p)hg)|\det p|^{-(\mathbf{n}+1)}\, dp\, dh$$
$$= \int_{H\cap \bar{U}} \mathfrak{P}^{H_{\mathbb{M}}}\big(\big(\delta_P^{-1/2} I\big(\tfrac{1}{2}, \bar{u}g\big)W\big)\circ\varrho\big)\, d\bar{u}$$

converges for any $W \in \mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}}(\pi), \tfrac{1}{2}\big)$ *and defines an intertwining map*

$$\mathrm{Ind}\big(\mathbb{W}^{\psi_{N_M}}(\pi), \tfrac{1}{2}\big) \to C^{\mathrm{sm}}(H\backslash G).$$

(4) [Lapid and Mao 2015b, second statement of Corollary 11.10]. *We have*

$$A_e^{\psi}(M^*W) = \epsilon_\pi^{\mathbf{n}+1} \int_{N_{\mathbb{M},\Delta}} \left(\int_{H\cap\mathfrak{E}\backslash\mathfrak{E}} L_W(v\varrho(\boldsymbol{\epsilon}_4 u \boldsymbol{\epsilon}_3) w_U)\psi_{\mathfrak{E}}^{-1}(v)\, dv \right) du. \tag{8-6}$$

The first part is based on a variant of Bernstein's theorem [Bernstein 1984] on $\mathcal{P}$-invariant distributions. The second part is proved by "root killing", a method which goes back at least to [Jacquet et al. 1979] and has been used constantly ever since. The third part is essentially a formal consequence of the first part. The last (and most complicated) part is closely related to Theorem 8.1 (with $\boldsymbol{\epsilon}_1$ as before). We refer to [Lapid and Mao 2015b] for more details.

Remark 8.4. While the above formula for $A_e^{\psi}(M^*W)$ is convenient for computing the right-hand side of (8-4), it is also helpful to write the integration with the variables at the right. We do that in (8-9) below.

Corollary 8.5. *Let* $\pi \in \mathrm{Irr}_{\mathrm{meta,temp}}\,\mathbb{M}$. *Then for any* $W \in \mathrm{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))_{\natural}^{\circ}$ *we have*

$$\int_{Z_{\mathbb{M}}\backslash T''} E^{\psi}(W_{1/2}, t)\, \frac{dt}{|\det t|} = \epsilon_\pi^{\mathbf{n}+1} A_e^{\psi}(M^*W). \tag{8-7}$$

Note it is not a priori clear that the left-hand side of (8-7) factors through M^*W.

Proof. We may assume without loss of generality that for $m \in \mathbb{M}$, $u, u' \in U$,

$$W_{1/2}(u'\varrho(m)w_U u) = W^{\mathbb{M}}(m)|\det m|^{1/2}\delta_P(\varrho(m))^{1/2}\phi(u) \tag{8-8}$$

with $W^{\mathbb{M}} \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural}$ and $\phi \in C_c^{\infty}(U)$. We evaluate the left-hand side I of (8-7) using the last expression in (7-7) (where we recall that the integrand is compactly supported by Lemma 7.9). Thus,

$$I = I' \int_U \phi(v)\psi_U^{-1}(v)\, dv$$

where

$$I' = \int_{\eta_{\mathbb{M}}^{\vee}(T'_{\mathbb{M}'})} \int_{\bar{R}} \Delta(t)^{-1}|\det t|^{\mathbf{n}} W^{\mathbb{M}}(t\boldsymbol{\epsilon}_4 r \boldsymbol{\epsilon}_3)\psi_{\bar{R}}(r)\, dr\, dt.$$

The integrand in I' is compactly supported because $W^{\mathbb{M}} \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural}$. Note that for $r \in V_\Delta = \mathcal{Z} \cap \bar{R}$, we have $\psi_{\bar{R}}(r) = \psi_{\mathcal{Z}}(r)^{-1}$. We write

$$\begin{aligned} I' &= \int_{N_{\mathbb{M},\Delta}} \left(\int_{V_\Delta} \int_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})} \Delta(t)^{-1} |\det t|^{\mathbf{n}} W^{\mathbb{M}}(t\epsilon_4 r u \epsilon_3) \psi_{\mathcal{Z}}(r)^{-1}\, dt\, dr \right) du \\ &= \int_{N_{\mathbb{M},\Delta}} \left(\int_{\mathcal{Z}^+ \backslash \mathcal{Z}} \int_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})} \Delta(t)^{-1} |\det t|^{\mathbf{n}} W^{\mathbb{M}}(t\epsilon_4 r u \epsilon_3) \psi_{\mathcal{Z}}(r)^{-1}\, dt\, dr \right) du \\ &= \int_{N_{\mathbb{M},\Delta}} \left(\int_{\mathcal{Z}^+ \backslash \mathcal{Z}} \int_{\eta^\vee_{\mathbb{M}}(T'_{\mathbb{M}'})} \Delta(t)^{-1} |\det t|^{\mathbf{n}} W^{\mathbb{M}}(t r \epsilon_4 u \epsilon_3) \psi_{\mathcal{Z}}(r)^{-1}\, dt\, dr \right) du \end{aligned}$$

(since ϵ_4 stabilizes $\psi_{\mathcal{Z}}$). For the double integral in the brackets we apply part (2) of the theorem above to $\pi(\epsilon_4 u \epsilon_3) W^{\mathbb{M}}$ (which is applicable since $W^{\mathbb{M}} \in \mathbb{W}^{\psi_{N_{\mathbb{M}}}}(\pi)_{\natural}$). We get

$$I' = \int_{N_{\mathbb{M},\Delta}} \left(\int_{\mathcal{Z} \cap H_{\mathbb{M}} \backslash \mathcal{Z}} \mathfrak{P}^{H_{\mathbb{M}}}(\pi(n\epsilon_4 u \epsilon_3) W^{\mathbb{M}}) \psi_{\mathcal{Z}}(n)^{-1}\, dn \right) du.$$

Thus,

$$I = \int_{N_{\mathbb{M},\Delta}} \left(\int_{\mathcal{Z} \cap H_{\mathbb{M}} \backslash \mathcal{Z}} \int_U \mathfrak{P}^{H_{\mathbb{M}}}(\pi(n\epsilon_4 u \epsilon_3) W^{\mathbb{M}}) \psi_{\mathcal{Z}}(n)^{-1} \phi(v) \psi_U^{-1}(v)\, dv\, dn \right) du.$$

From (8-8), $\left(\delta_P^{-1/2} I\left(\frac{1}{2}, \varrho(m) w_U v\right) W\right) \circ \varrho = \phi(v) \delta_P^{1/2}(\varrho(m)) |\det m|^{1/2} \pi(m) W^{\mathbb{M}}$ for any $v \in U$, $m \in \mathbb{M}$. Thus, I equals

$$\int_{N_{\mathbb{M},\Delta}} \left(\int_{\mathcal{Z} \cap H_{\mathbb{M}} \backslash \mathcal{Z}} \int_U \mathfrak{P}^{H_{\mathbb{M}}}\left(\left(\delta_P^{-1/2} I\left(\tfrac{1}{2}, \varrho(n\epsilon_4 u \epsilon_3) w_U v\right) W\right) \circ \varrho \right) \psi_{\mathcal{Z}}(n)^{-1} \psi_U(v)^{-1}\, dv\, dn \right) du.$$

Because $\epsilon_3^{-1} N_{\mathbb{M},\Delta} \epsilon_3 \subset N_{\mathbb{M}}^\flat$, the group $\varrho(\epsilon_3^{-1} N_{\mathbb{M},\Delta} \epsilon_3)$ stabilizes the character $\psi_U(w_U^{-1} \cdot w_U)$ on $\bar{U}$. Making a change of variable

$$v \mapsto (\varrho(n\epsilon_4 u \epsilon_3) w_U)^{-1} \bar{v} \varrho(n\epsilon_4 u \epsilon_3) w_U$$

on U we obtain

$$I = \int_{N_{\mathbb{M},\Delta}} \left(\int_{\varrho(\mathcal{Z} \cap H_{\mathbb{M}}) \backslash \mathfrak{E}} \mathfrak{P}^{H_{\mathbb{M}}}\left(\left(\delta_P^{-1/2} I\left(\tfrac{1}{2}, v \varrho(\epsilon_4 u \epsilon_3) w_U\right) W\right) \circ \varrho \right) \psi_{\mathfrak{E}}^{-1}(v)\, dv \right) du.$$

Integrating over $H \cap \bar{U}$ first, we get from part (3) of the theorem that

$$I = \int_{N_{\mathbb{M},\Delta}} \left(\int_{H \cap \mathfrak{E} \backslash \mathfrak{E}} L_W(v \varrho(\epsilon_4 u \epsilon_3) w_U) \psi_{\mathfrak{E}}^{-1}(v)\, dv \right) du.$$

Finally, by part (4) of the theorem, this is equal to $\epsilon_\pi^{\mathbf{n}+1} A_e^{\psi}(M^* W)$ as required. $\square$

8C.

Proof of Proposition 6.15. From Corollaries 8.2 and 8.5 we get that (6-12) holds for $(W, W^\wedge) \in \operatorname{Ind}(\mathbb{W}^{\psi_{N_M}}(\pi))^\circ_\sharp \times \operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural$, namely

$$B\big(W, M\big(\tfrac{1}{2}\big)W^\wedge, \tfrac{1}{2}\big) = \epsilon_\pi A^\psi_e(M^* W) A^{\psi^{-1}}_e(M^* W^\wedge).$$

On the other hand, as in the proof of Lemma 6.13, it follows from the definition (7-7) and the fact that the image of the restriction map $\mathbb{W}^{\psi^{-1}_{N_M}}(\pi) \to C(N_M \backslash \varrho(\mathcal{P}^*), \psi^{-1}_{N_M})$ contains $C^\infty_c(N_M \backslash \varrho(\mathcal{P}^*), \psi^{-1}_{N_M})$ that the linear map $\operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural \to C^\infty_c(T'_{\mathbb{M}'})$ given by $W^\wedge \mapsto E^{\psi^{-1}}(W^\wedge_{1/2}, \cdot)$ is onto. Therefore, by Corollary 8.5 the linear form $A^{\psi^{-1}}_e(M^* W^\wedge)$ is nonvanishing on $\operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural$. By Corollary 6.14 we conclude that (6-12) holds for all $W^\wedge$ $\big($not necessarily in $\operatorname{Ind}\big(\mathbb{W}^{\psi^{-1}_{N_M}}(\pi)\big)^\circ_\natural\big)$. Proposition 6.15 follows. $\square$

By the discussion before Proposition 6.15, this concludes the proof of Theorem 3.4 and thus the proof of Theorem 1.3.

8D. *Central character of the descent.* Assume that $\pi \in \operatorname{Irr}_{\mathrm{meta,temp}} \mathbb{M}$. Let us give another expression for $A^\psi_e(M^* W)$, which we use to prove a formula for the so-called "central sign" of the descent $\mathcal{D}_\psi(\pi)$.

Denote by $\boldsymbol{\epsilon}_5$ the element $\varrho(\boldsymbol{\epsilon}_4 \boldsymbol{\epsilon}_3) w_U$. A change of variables in u and v in (8-6) gives

$$A^\psi_e(M^* W) = \epsilon^{\mathbf{n}+1}_\pi \int_{N^\sharp_{\mathbb{M},\Delta}} \left(\int_{(H \cap \mathfrak{E})^{\boldsymbol{\epsilon}_5} \backslash \mathfrak{E}^{\boldsymbol{\epsilon}_5}} L_W(\boldsymbol{\epsilon}_5 v \varrho(u)) \psi^{-1}_{\mathfrak{E}^{\boldsymbol{\epsilon}_5}}(v)\, dv \right) du. \tag{8-9}$$

Here

$$\begin{aligned} N^\sharp_{\mathbb{M},\Delta} &= \{\ell_{\mathbb{M}}(X) : X_{i,j} = 0 \text{ if either } j > i \text{ or } i = \mathbf{n}\}, \\ \mathfrak{E}^{\boldsymbol{\epsilon}_5} &= \boldsymbol{\epsilon}_5^{-1} \mathfrak{E} \boldsymbol{\epsilon}_5 = \varrho(\mathcal{Z}^\sharp) \ltimes U, \\ \psi_{\mathfrak{E}^{\epsilon_5}}(\varrho(m)u) &= \psi_{\mathcal{Z}^\sharp}(m)\psi_U(u), \quad m \in \mathcal{Z}^\sharp,\ u \in U, \end{aligned}$$

with

$$\begin{aligned} \mathcal{Z}^\sharp &= \left\{ \left(\begin{smallmatrix} n_1 & v_1 \\ v_2 & n_2 \end{smallmatrix}\right) : n_1, n_2 \in N'_{\mathbb{M}'},\ I_{\mathbf{n}} + v_1,\ I_{\mathbf{n}} + v_2 \in N'_{\mathbb{M}'} \right\}, \\ \psi_{\mathcal{Z}^\sharp}(m) &= \psi(-m_{1,2} - \cdots - m_{\mathbf{n}-1,\mathbf{n}} + m_{\mathbf{n}+1,\mathbf{n}+2} + \cdots + m_{2\mathbf{n}-1,2\mathbf{n}}), \quad m \in \mathcal{Z}^\sharp. \end{aligned}$$

From this expression we get the next proposition.

Proposition 8.6. *Let* $\pi \in \operatorname{Irr}_{\mathrm{gen,meta}} \mathbb{M}$, *and* $\tilde{\pi} = \mathcal{D}_\psi(\pi)$. *Then*

$$\tilde{\pi}(\widetilde{-I_{2\mathbf{n}}}) = \gamma_{\psi^{-1}}((-1)^{\mathbf{n}}) \epsilon_\pi.$$

Recall that $\tilde{\pi}(\widetilde{-I_{2\mathbf{n}}})/\gamma_{\psi^{-1}}((-1)^{\mathbf{n}})$ is the central sign of $\tilde{\pi}$ (with respect to ψ^{-1}) introduced by Gan and Savin [2012]. (However, we do not assume $\tilde{\pi}$ is irreducible.)

Proof. We need to show $A^{\psi}_{\sharp}(M^*W, \widetilde{-I_{2\mathbf{n}}}) = \gamma_{\psi^{-1}}((-1)^{\mathbf{n}})\epsilon_\pi A^{\psi}_{e}(M^*W)$. For the moment assume that π is tempered. By Lemma 3.1 and (8-9), $A^{\psi}_{\sharp}(M^*W, \widetilde{-I_{2\mathbf{n}}})$ equals

$$\gamma_{\psi^{-1}}((-1)^{\mathbf{n}})\epsilon_\pi^{\mathbf{n}+1}\int_{N^{\sharp}_{\mathbb{M},\Delta}}\left(\int_{(H\cap\mathfrak{E})^{\epsilon_5}\backslash\mathfrak{E}^{\epsilon_5}} L_W(\boldsymbol{\epsilon}_5 v\varrho(u)\eta(-I_{2\mathbf{n}}))\psi^{-1}_{\mathfrak{E}^{\epsilon_5}}(v)\,dv\right)du.$$

Conjugating $v\varrho(u)$ by $\eta(-I_{2\mathbf{n}})$ we obtain

$$\gamma_{\psi^{-1}}((-1)^{\mathbf{n}})\epsilon_\pi^{\mathbf{n}+1}\int_{N^{\sharp}_{\mathbb{M},\Delta}}\left(\int_{(H\cap\mathfrak{E})^{\epsilon_5}\backslash\mathfrak{E}^{\epsilon_5}} L_W(\boldsymbol{\epsilon}_5 \eta(-I_{2\mathbf{n}})v\varrho(u))\psi^{-1}_{\mathfrak{E}^{\epsilon_5}}(v)\,dv\right)du.$$

Here we need to verify that $\eta(-I_{2\mathbf{n}})$ stabilizes $(\mathfrak{E}^{\epsilon_5}, \psi_{\mathfrak{E}^{\epsilon_5}})$ (which is clear) and normalizes H^{ϵ_5}. The second fact follows from the observation that $\boldsymbol{\epsilon}_5\eta(-I_{2\mathbf{n}}) = \varrho(\boldsymbol{a})\boldsymbol{\epsilon}_5$, where

$$\boldsymbol{a} = \operatorname{diag}\left(\tfrac{1}{2}\mathfrak{d}, 2\mathfrak{d}^*\right)w_0^{\mathbb{M}} \in H_{\mathbb{M}} w_0^{\mathbb{M}}.$$

Since $L_W(\varrho(w_0^{\mathbb{M}})\,\cdot\,) = \epsilon_\pi L_W(\,\cdot\,)$ [Lapid and Mao 2015b, (4.8)], we get

$$A^{\psi}_{\sharp}(M^*W, \widetilde{-I_{2\mathbf{n}}}) = \gamma_{\psi^{-1}}((-1)^{\mathbf{n}})\epsilon_\pi^{\mathbf{n}}\int_{N^{\sharp}_{\mathbb{M},\Delta}}\left(\int_{(H\cap\mathfrak{E})^{\epsilon_5}\backslash\mathfrak{E}^{\epsilon_5}} L_W(\boldsymbol{\epsilon}_5 v\varrho(u))\psi^{-1}_{\mathfrak{E}^{\epsilon_5}}(v)\,dv\right)du.$$

The claim now follows from the comparison with (8-9). Finally, the same argument as in Section 3F using the classification result Theorem 3.8 gives the proposition for all $\pi \in \operatorname{Irr}_{\mathrm{gen,meta}}\mathbb{M}$. □

Appendix: Nonvanishing of Bessel functions

Let G be a split group over a p-adic field.[5] Let $B = A \ltimes N$ be a Borel subgroup of G. Let $G^\circ = Bw_0B$ be the open Bruhat cell where w_0 is the longest element of the Weyl group. Fix a nondegenerate continuous character ψ_N of N. For any $\pi \in \operatorname{Irr}_{\mathrm{gen},\psi_N}(G)$, the Bessel function $\mathbb{B}_\pi = \mathbb{B}_\pi^{\psi_N}$ of π with respect to ψ_N is the locally constant function on G° given by the relation

$$\int_N^{\mathrm{st}} W(gn)\psi_N(n)^{-1}\,dn = \mathbb{B}_\pi(g)W(e) \tag{A-1}$$

for any $W \in \mathbb{W}^{\psi_N}(\pi)$; see [Lapid and Mao 2013]. In this section we prove the following result.

Theorem A.1. *For any tempered $\pi \in \operatorname{Irr}_{\mathrm{gen},\psi_N}(G)$ the function $\mathbb{B}_\pi$ is not identically zero on G°.*

[5] The notation in the appendix is different from the body of the paper.

The argument is similar to the one in [Ichino and Zhang 2014]. Fix a tempered $\pi \in \mathrm{Irr}_{\mathrm{gen},\psi_N}(G)$ and realize it on its Whittaker model $\mathbb{W}^{\psi_N}(\pi)$. Similarly realize $\pi^\vee$ on $\mathbb{W}^{\psi_N^{-1}}(\pi^\vee)$. Thus, we get a pairing $(\cdot,\cdot)$ on $\mathbb{W}^{\psi_N}(\pi)\times\mathbb{W}^{\psi_N^{-1}}(\pi^\vee)$.

Fix $W_0^\vee \in \mathbb{W}^{\psi_N^{-1}}(\pi^\vee)$ such that

$$\int_N^{\mathrm{st}} (\pi(n)W, W_0^\vee)\psi_N(n)^{-1}\,dn = W(e)$$

for all $W\in\mathbb{W}^{\psi_N}(\pi)$. This is possible by [Lapid and Mao 2015a, Propositions 2.3 and 2.10]. Then for any $g\in G$ we have

$$W(g) = \int_N^{\mathrm{st}} (\pi(ng)W, W_0^\vee)\psi_N(n)^{-1}\,dn. \tag{A-2}$$

Similarly, fix $W_0\in\mathbb{W}^{\psi_N}(\pi)$ such that for all $W^\vee\in\mathbb{W}^{\psi_N^{-1}}(\pi^\vee)$,

$$\int_N^{\mathrm{st}} (W_0, \pi^\vee(n)W^\vee)\psi_N(n)\,dn = W^\vee(e). \tag{A-3}$$

Set $\Phi(g) = (\pi(g)W_0, W_0^\vee)$. Let N^{der} be the derived group of N. Also let Ξ be the Harish-Chandra function on G (see, e.g., [Waldspurger 2003]).

Lemma A.2. *The function*

$$g\mapsto \int_{N^{\mathrm{der}}}\int_{N^{\mathrm{der}}} \Phi(n_1gn_2)\,dn_1\,dn_2$$

on G° is locally L^1 on G. Moreover,

$$g\mapsto \int_{N^{\mathrm{der}}}\int_{N^{\mathrm{der}}} \Xi(n_1gn_2)\,dn_1\,dn_2$$

is locally L^1 on G.

Proof. The argument is exactly as in [Ichino and Zhang 2014, Lemma A.4] using the convergence of $\int_{N^{\mathrm{der}}}\Xi(n)\,dn$ [Sakellaridis and Venkatesh 2012, Lemma 6.3.1]. □

Remark A.3. Note that $g\mapsto\int_{N^{\mathrm{der}}}\int_{N^{\mathrm{der}}}\Xi(n_1gn_2)\,dn_1\,dn_2$ is locally constant on G°. Thus, its local integrability on G implies its convergence for any $g\in G^\circ$.

For any $f\in C_c^\infty(G)$ let $L_f(W)=\int_G f(g)W(g)\,dg$ for $W\in\mathbb{W}^{\psi_N}(\pi)$. Then $L_f\in\pi^\vee$. Let L_f^* be the corresponding element in $\mathbb{W}^{\psi_N^{-1}}(\pi^\vee)$ and set $B_\pi(f)=L_f^*(e)$. The distribution $f\mapsto B_\pi(f)$ is called the *Bessel distribution*. It is nonzero: we can choose $f\in C_c^\infty(G)$ such that L_f is nontrivial, and then, by translating f if necessary we can arrange that $L_f^*(e)\neq 0$.

Note that by (A-3),

$$B_\pi(f) = \int_N^{\mathrm{st}} (W_0, \pi^\vee(n)L_f^*)\psi_N(n)\,dn = \int_N^{\mathrm{st}}\left(\int_G f(g)W_0(gn^{-1})\,dg\right)\psi_N(n)\,dn,$$

and therefore by (A-2) we have

$$B_\pi(f) = \int_N^{\mathrm{st}} \left(\int_G f(g) \left(\int_N^{\mathrm{st}} \Phi(n_1 g n_2^{-1}) \psi_N(n_1)^{-1}\, dn_1 \right) dg \right) \psi_N(n_2)\, dn_2. \tag{A-4}$$

Let A^1 be the maximal compact subgroup of A. Fix $\Omega_0 \in \mathcal{CSGR}(N)$, which is invariant under conjugation by A^1 (e.g., take $\Omega_0 = N \cap K$). Fix an element $a \in A$ such that $|\alpha(a)| > 1$ for all $\alpha \in \Delta_0$. Consider the sequence $\Omega_n = a^n \Omega_0 a^{-n} \in \mathcal{CSGR}(N)$, $n = 1, 2, \ldots$. Any Ω_n is invariant under conjugation by A^1, $\Omega_1 \subset \Omega_2 \subset \cdots$ and $\bigcup \Omega_n = N$.

Let $A^d = A \cap G^{\mathrm{der}}$. Consider the family

$$A_n = \{t \in A^d : |\alpha(t) - 1| \le q^{-n} \text{ for all } \alpha \in \Delta_0\} \in \mathcal{CSGR}(A^d),$$

which forms a basis of neighborhoods of 1 for A^d. We only consider n sufficiently large so that the image of A_n under the homomorphism $t \in A \mapsto (\alpha(t))_{\alpha \in \Delta_0} \in (F^*)^{\Delta_0}$ is $\prod_{\alpha \in \Delta_0} (1 + \varpi^n \mathcal{O})$, where ϖ is a uniformizer of $\mathcal{O}$.

For any $\alpha \in \Delta_0$ choose a parameterization $x_\alpha : F \to N_\alpha$ such that $\psi_N \circ x_\alpha$ is trivial on $\mathcal{O}$ but not on $\varpi^{-1}\mathcal{O}$. Let N_n be the group generated by $\langle x_\alpha(\varpi^{-n}\mathcal{O}), \alpha \in \Delta_0 \rangle$ and the derived group N^{der} of N. The following lemma is clear.

Lemma A.4. *For any $u \in N$ we have*

$$(\mathrm{vol}\, A_n)^{-1} \int_{A_n} \psi_N(tut^{-1})\, dt = \begin{cases} \psi_N(u) & \text{for } u \in N_n, \\ 0 & \text{otherwise.} \end{cases}$$

Next we prove:

Lemma A.5. *Suppose that f and Φ are bi-invariant under A_n. Then we have*

$$B_\pi(f) = \int_G f(g) \alpha_n(g)\, dg$$

where

$$\alpha_n(g) = \int_{N_n} \int_{N_n} \Phi(n_1 g n_2^{-1}) \psi_N(n_1)^{-1} \psi_N(n_2)\, dn_2\, dn_1.$$

Remark A.6. Of course we cannot conclude from the lemma by itself that B_π is given by a locally L^1 function (though this is conjectured to be the case.)

Proof. We start with (A-4). Let m be such that

$$B_\pi(f) = \int_{\Omega_m} \int_G \left(\int_N^{\mathrm{st}} f(g) \Phi(n_1 g n_2^{-1}) \psi_N(n_1)^{-1}\, dn_1 \right) dg\, \psi_N(n_2)\, dn_2.$$

Then for any $t \in A_n$ we have

$$\begin{aligned} B_\pi(f) &= \int_{\Omega_m} \int_G \left(\int_N^{\mathrm{st}} f(gt)\Phi(n_1 g n_2^{-1} t)\psi_N(n_1)^{-1}\, dn_1 \right) dg\, \psi_N(n_2)\, dn_2 \\ &= \int_{\Omega_m} \int_G \left(\int_N^{\mathrm{st}} f(gt)\Phi(n_1 g t n_2^{-1})\psi_N(n_1)^{-1}\, dn_1 \right) dg\, \psi_N(t n_2 t^{-1})\, dn_2 \\ &= \int_{\Omega_m} \int_G \left(\int_N^{\mathrm{st}} f(g)\Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\, dn_1 \right) dg\, \psi_N(t n_2 t^{-1})\, dn_2. \end{aligned}$$

Averaging over $t \in A_n$ we get

$$B_\pi(f) = \int_{N_n \cap \Omega_m} \int_G \left(\int_N^{\mathrm{st}} f(g)\Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\, dn_1 \right) dg\, \psi_N(n_2)\, dn_2$$

by Lemma A.4. Thus, for m' sufficiently large (depending on f and m) we have

$$B_\pi(f) = \int_{N_n \cap \Omega_m} \int_G \int_{\Omega_{m'}} f(g)\Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\, dn_1\, dg\, \psi_N(n_2)\, dn_2.$$

Once again, for any $t \in A_n$,

$$\begin{aligned} B_\pi(f) &= \int_{N_n \cap \Omega_m} \int_G \int_{\Omega_{m'}} f(tg)\Phi(t n_1 g n_2^{-1})\psi_N(n_1)^{-1}\, dn_1\, dg\, \psi_N(n_2)\, dn_2 \\ &= \int_{N_n \cap \Omega_m} \int_G \int_{\Omega_{m'}} f(tg)\Phi(n_1 t g n_2^{-1})\psi_N(t^{-1} n_1 t)^{-1}\, dn_1\, dg\, \psi_N(n_2)\, dn_2 \\ &= \int_{N_n \cap \Omega_m} \int_G \int_{\Omega_{m'}} f(g)\Phi(n_1 g n_2^{-1})\psi_N(t^{-1} n_1 t)^{-1}\, dn_1\, dg\, \psi_N(n_2)\, dn_2. \end{aligned}$$

As before, averaging over A_n we get

$$B_\pi(f) = \int_{N_n \cap \Omega_m} \int_G \int_{N_n \cap \Omega_{m'}} f(g)\Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\, dn_1\, dg\, \psi_N(n_2)\, dn_2.$$

Since m and m' can be chosen arbitrarily large, the lemma follows from the convergence of

$$\int_{N_n} \int_G \int_{N_n} |f(g)\Phi(n_1 g n_2^{-1})|\, dn_1\, dg\, dn_2,$$

i.e., Lemma A.2. □

We prove the following analogously.

Lemma A.7. *For $g \in G^\circ$, let $\mathbb{B}_\pi^{A_n}(g) := \mathrm{vol}(A_n)^{-2} \int_{A_n} \int_{A_n} \mathbb{B}_\pi(t_1 g t_2)\, dt_1\, dt_2$. Suppose that Φ is bi-invariant under A_n. Then*

$$\mathbb{B}_\pi^{A_n}(g) W_0(e) = \alpha_n(g)$$

for all $g \in G^\circ$.

Proof. First note that for any compact subset C of G° we have

$$\int_{\Omega_m}\left(\int_N^{\mathrm{st}} \Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\right)\psi_N(n_2)\,dn_2 = \mathbb{B}_\pi(g)W_0(e)$$

for all $g \in C$ and m sufficiently large. Indeed, by (A-2) the inner stable integral is $W_0(gn_2^{-1})$. Thus, the relation above follows from [Lapid and Mao 2013] and (A-1).

Now, for any $t \in A_n$ we have

$$\begin{aligned}\mathbb{B}_\pi(gt)W_0(e) &= \int_{\Omega_m}\left(\int_N^{\mathrm{st}} \Phi(n_1 g t n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\right)\psi_N(n_2)\,dn_2\\ &= \int_{\Omega_m}\left(\int_N^{\mathrm{st}} \Phi(n_1 g n_2^{-1} t)\psi_N(n_1)^{-1}\,dn_1\right)\psi_N(t^{-1}n_2 t)\,dn_2\\ &= \int_{\Omega_m}\left(\int_N^{\mathrm{st}} \Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\right)\psi_N(t^{-1}n_2 t)\,dn_2.\end{aligned}$$

Thus, by Lemma A.4, we have

$$\begin{aligned}&\mathrm{vol}(A_n)^{-1}\int_{A_n}\mathbb{B}_\pi(gt)\,dt\;W_0(e)\\ &\qquad= \int_{N_n\cap\Omega_m}\left(\int_N^{\mathrm{st}} \Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\right)\psi_N(n_2)\,dn_2.\end{aligned}$$

For g in a compact set C of G° and for m' sufficiently large (depending on C and m) we can write this as

$$\int_{N_n\cap\Omega_m}\int_{\Omega_{m'}} \Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\,\psi_N(n_2)\,dn_2.$$

Now, for any $t_1 \in A_n$,

$$\begin{aligned}&\mathrm{vol}(A_n)^{-1}\int_{A_n}\mathbb{B}_\pi(t_1 g t_2)\,dt_2\;W_0(e)\\ &\qquad= \int_{N_n\cap\Omega_m}\int_{\Omega_{m'}} \Phi(n_1 t_1 g n_2^{-1})\psi_N(n_1)^{-1}\,dn_1\,\psi_N(n_2)\,dn_2\\ &\qquad= \int_{N_n\cap\Omega_m}\int_{\Omega_{m'}} \Phi(t_1 n_1 g n_2^{-1})\psi_N(t_1 n_1 t_1^{-1})^{-1}\,dn_1\,\psi_N(n_2)\,dn_2\\ &\qquad= \int_{N_n\cap\Omega_m}\int_{\Omega_{m'}} \Phi(n_1 g n_2^{-1})\psi_N(t_1 n_1 t_1^{-1})^{-1}\,dn_1\,\psi_N(n_2)\,dn_2.\end{aligned}$$

Averaging over $t_1 \in A_n$ we get

$$\mathbb{B}_\pi^{A_n}(g)W_0(e) = \int_{N_n\cap\Omega_m}\int_{N_n\cap\Omega_{m'}} \Phi(n_1 g n_2^{-1})\psi_N(n_1)^{-1}\psi_N(n_2)\,dn_1\,dn_2.$$

The lemma follows from the convergence of

$$\int_{N_n}\int_{N_n} |\Phi(n_1 g n_2^{-1})|\, dn_1\, dn_2,$$

i.e., Remark A.3. $\square$

Proof of Theorem A.1. Since $B_\pi(f) \neq 0$, we have $\alpha_n|_{G^\circ} \not\equiv 0$ for $n \gg 1$ by Lemma A.5. Hence, by Lemma A.7, $\mathbb{B}_\pi^{A_n}\big|_{G^\circ} \not\equiv 0$ and therefore $\mathbb{B}_\pi\big|_{G^\circ} \not\equiv 0$. $\square$

Remark A.8. Theorem A.1 and its proof are also valid for $\widetilde{\mathrm{Sp}}_n$.

Acknowledgements

Part of this work was done while the authors were hosted by the Erwin Schrödinger Institute in Vienna for the program "Research in Teams". We are grateful to the ESI, and especially Joachim Schwermer, for their hospitality. We also thank Atsushi Ichino and Marko Tadić for helpful correspondence.

References

[Aizenbud and Gourevitch 2009] A. Aizenbud and D. Gourevitch, "Generalized Harish-Chandra descent, Gelfand pairs, and an Archimedean analog of Jacquet–Rallis's theorem", *Duke Math. J.* **149**:3 (2009), 509–567. MR Zbl

[Baruch 2005] E. M. Baruch, "Bessel functions for $\mathrm{GL}(n)$ over a p-adic field", pp. 1–40 in *Automorphic representations, L-functions and applications: progress and prospects*, edited by J. W. Cogdell et al., Ohio State Univ. Math. Res. Inst. Publ. **11**, de Gruyter, Berlin, 2005. MR Zbl

[Bernstein 1984] J. N. Bernstein, "P-invariant distributions on $\mathrm{GL}(N)$ and the classification of unitary representations of $\mathrm{GL}(N)$ (non-Archimedean case)", pp. 50–102 in *Lie group representations* (College Park, MD, 1982/1983), vol. 2, edited by R. Herb et al., Lecture Notes in Math. **1041**, Springer, Berlin, 1984. MR Zbl

[Cogdell et al. 2001] J. W. Cogdell, H. H. Kim, I. I. Piatetski-Shapiro, and F. Shahidi, "On lifting from classical groups to GL_N", *Publ. Math. Inst. Hautes Études Sci.* 93 (2001), 5–30. MR Zbl

[Cogdell et al. 2004] J. W. Cogdell, H. H. Kim, I. I. Piatetski-Shapiro, and F. Shahidi, "Functoriality for the classical groups", *Publ. Math. Inst. Hautes Études Sci.* 99 (2004), 163–233. MR Zbl

[Cogdell et al. 2011] J. W. Cogdell, I. I. Piatetski-Shapiro, and F. Shahidi, "Functoriality for the quasisplit classical groups", pp. 117–140 in *On certain L-functions* (West Lafayette, IN, 2007), edited by J. Arthur et al., Clay Math. Proc. **13**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Friedberg and Jacquet 1993] S. Friedberg and H. Jacquet, "Linear periods", *J. Reine Angew. Math.* **443** (1993), 91–139. MR Zbl

[Gan and Savin 2012] W. T. Gan and G. Savin, "Representations of metaplectic groups, I: epsilon dichotomy and local Langlands correspondence", *Compos. Math.* **148**:6 (2012), 1655–1694. MR Zbl

[Ginzburg et al. 1998] D. Ginzburg, S. Rallis, and D. Soudry, "L-functions for symplectic groups", *Bull. Soc. Math. France* **126**:2 (1998), 181–244. MR Zbl

[Ginzburg et al. 1999] D. Ginzburg, S. Rallis, and D. Soudry, "On a correspondence between cuspidal representations of GL_{2n} and $\widetilde{\mathrm{Sp}}_{2n}$", *J. Amer. Math. Soc.* **12**:3 (1999), 849–907. MR Zbl

[Ginzburg et al. 2011] D. Ginzburg, S. Rallis, and D. Soudry, *The descent map from automorphic representations of* $\mathrm{GL}(n)$ *to classical groups*, World Scientific Publishing, Hackensack, NJ, 2011. MR Zbl

[Hiraga et al. 2008] K. Hiraga, A. Ichino, and T. Ikeda, "Formal degrees and adjoint γ-factors", *J. Amer. Math. Soc.* **21**:1 (2008), 283–304. MR Zbl

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

[Ichino and Zhang 2014] A. Ichino and W. Zhang, "Spherical characters for a strongly tempered pair", (2014). Appendix to W. Zhang, "Fourier transform and the global Gan–Gross–Prasad conjecture for unitary groups", *Ann. of Math. (2)*, **180**:3 (2014), 971–1049. MR Zbl

[Ichino et al. 2017] A. Ichino, E. Lapid, and Z. Mao, "On the formal degrees of square-integrable representations of odd special orthogonal and metaplectic groups", *Duke Math. J.* (online publication January 2017).

[Jacquet and Rallis 1996] H. Jacquet and S. Rallis, "Uniqueness of linear periods", *Compositio Math.* **102**:1 (1996), 65–123. MR Zbl

[Jacquet and Shalika 1981a] H. Jacquet and J. A. Shalika, "On Euler products and the classification of automorphic representations, I", *Amer. J. Math.* **103**:3 (1981), 499–558. MR Zbl

[Jacquet and Shalika 1981b] H. Jacquet and J. A. Shalika, "On Euler products and the classification of automorphic forms, II", *Amer. J. Math.* **103**:4 (1981), 777–815. MR Zbl

[Jacquet et al. 1979] H. Jacquet, I. I. Piatetski-Shapiro, and J. Shalika, "Automorphic forms on GL(3), I", *Ann. of Math.* (2) **109**:1 (1979), 169–212. MR Zbl

[Jiang and Soudry 2004] D. Jiang and D. Soudry, "Generic representations and local Langlands reciprocity law for p-adic SO_{2n+1}", pp. 457–519 in *Contributions to automorphic forms, geometry, and number theory* (Baltimore, 2002), edited by H. Hida et al., Johns Hopkins University Press, Baltimore, MD, 2004. MR Zbl

[Kaplan 2015] E. Kaplan, "Complementary results on the Rankin–Selberg gamma factors of classical groups", *J. Number Theory* **146** (2015), 390–447. MR Zbl

[Kneser 1967] M. Kneser, "Semi-simple algebraic groups", pp. 250–265 in *Algebraic number theory* (Brighton, 1965), edited by J. W. S. Cassels and A. Fröhlich, Thompson, Washington, DC, 1967. MR

[Lapid and Mao 2013] E. Lapid and Z. Mao, "Stability of certain oscillatory integrals", *Int. Math. Res. Not.* **2013**:3 (2013), 525–547. MR Zbl

[Lapid and Mao 2014] E. Lapid and Z. Mao, "On a new functional equation for local integrals", pp. 261–294 in *Automorphic forms and related geometry: assessing the legacy of I. I. Piatetski-Shapiro* (New Haven, CT, 2012), edited by J. W. Cogdell et al., Contemp. Math. **614**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[Lapid and Mao 2015a] E. Lapid and Z. Mao, "A conjecture on Whittaker–Fourier coefficients of cusp forms", *J. Number Theory* **146** (2015), 448–505. MR Zbl

[Lapid and Mao 2015b] E. Lapid and Z. Mao, "Model transition for representations of metaplectic type", *Int. Math. Res. Not.* **2015**:19 (2015), 9486–9568. MR Zbl

[Lapid and Mao 2017] E. Lapid and Z. Mao, "Whittaker–Fourier coefficients of cusp forms on $\widetilde{\mathrm{Sp}}_n$: reduction to a local statement", *Amer. J. Math.* **139**:1 (2017), 1–55.

[Matringe 2015] N. Matringe, "On the local Bump–Friedberg L-function", *J. Reine Angew. Math.* **709** (2015), 119–170. MR Zbl

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

[Waldspurger 1981] J.-L. Waldspurger, "Sur les coefficients de Fourier des formes modulaires de poids demi-entier", *J. Math. Pures Appl.* (9) **60**:4 (1981), 375–484. MR Zbl

[Waldspurger 1985] J.-L. Waldspurger, "Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie", *Compositio Math.* **54**:2 (1985), 173–242. MR Zbl

[Waldspurger 2003] J.-L. Waldspurger, "La formule de Plancherel pour les groupes p-adiques (d'après Harish-Chandra)", *J. Inst. Math. Jussieu* **2**:2 (2003), 235–333. MR Zbl

Communicated by Peter Sarnak
Received 2016-08-07 Accepted 2016-12-16

erez.m.lapid@gmail.com *Department of Mathematics, Weizmann Institute of Science, 7610001 Rehovot, Israel*

zmao@rutgers.edu *Department of Mathematics and Computer Science, Rutgers University, Newark, NJ 07102, United States*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 11 No. 3 2017