Algebra & Number Theory

Volume 17
2023
No. 7

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2023 is US $485/year for the electronic version, and $705/year (+$65, 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.

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 17:7 (2023)
https://doi.org/10.2140/ant.2023.17.1239

Counting abelian varieties over finite fields via Frobenius densities

Jeffrey D. Achter, S. Ali Altuğ, Luis Garcia and Julia Gordon
Appendix by Wen-Wei Li and Thomas Rüd

Let $[X, \lambda]$ be a principally polarized abelian variety over a finite field with commutative endomorphism ring; further suppose that either X is ordinary or the field is prime. Motivated by an equidistribution heuristic, we introduce a factor $\nu_v([X, \lambda])$ for each place v of $\mathbb{Q}$, and show that the product of these factors essentially computes the size of the isogeny class of $[X, \lambda]$.

The derivation of this mass formula depends on a formula of Kottwitz and on analysis of measures on the group of symplectic similitudes and, in particular, does not rely on a calculation of class numbers.

1. Introduction

Let $[X, \lambda] \in \mathcal{A}_g(\mathbb{F}_q)$ be a principally polarized g-dimensional abelian variety over the finite field $\mathbb{F}_q = \mathbb{F}_{p^e}$. Its isogeny class $I([X, \lambda], \mathbb{F}_q)$ is finite; our goal is to understand the (weighted by automorphism group) cardinality $\widetilde{\#} I([X, \lambda], \mathbb{F}_q)$.

A random matrix heuristic might suggest the following. Let $f_{X/\mathbb{F}_q}(T)$ be the characteristic polynomial of Frobenius of X. It is well known that $f_{X/\mathbb{F}_q}(T) \in \mathbb{Z}[T]$. Following Gekeler [2003], for a rational prime $\ell \nmid p \operatorname{disc}(f)$, one can define a number

$$\nu_\ell([X, \lambda], \mathbb{F}_q) = \lim_{n\to\infty} \frac{\#\{\gamma \in \mathrm{GSp}_{2g}(\mathbb{Z}_\ell/\ell^n) : \mathrm{charpoly}_\gamma(T) = f_{X/\mathbb{F}_q}(T) \bmod \ell^n\}}{\#\mathrm{GSp}_{2g}(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_{\mathrm{GSp}_{2g}}(\mathbb{Z}/\ell^n)}, \tag{1-1}$$

where GSp_{2g} is the group of symplectic similitudes of a symplectic space of dimension $2g$, and $\mathbb{A}_{\mathrm{GSp}_{2g}}$ is the space of characteristic polynomials of these similitudes.

For $\ell \nmid p \operatorname{disc}(f)$, in which case the conjugacy class is determined by the characteristic polynomial (see Lemma 3.1), we interpret $\nu_\ell[X, \lambda]$ as the deviation of the size of the conjugacy class with characteristic polynomial $f_{X/\mathbb{F}_q}(T)$ from the average size of a conjugacy class in $\mathrm{GSp}(\mathbb{Z}_\ell)$.

For $\ell \mid \operatorname{disc}(f_{X/\mathbb{F}_q}(T))$, since the characteristic polynomial need not determine a unique conjugacy class in $\mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$, a slightly more involved definition of $\nu_\ell[X, \lambda]$ is needed; see (4-1). Similarly, we define quantities $\nu_p([X, \lambda], \mathbb{F}_q)$ and $\nu_\infty([X, \lambda], \mathbb{F}_q)$ using, respectively, equidistribution considerations for σ-conjugacy classes in $\mathrm{GSp}_{2g}(\mathbb{Q}_q)$ and the Sato–Tate measure on the compact form USp_{2g}.

MSC2010: primary 11G10; secondary 14G15, 20G25, 22E35.
Keywords: abelian variety, isogeny class, orbital integral.

Careless optimism might lead one to hope that $\widetilde{\#}I([X,\lambda],\mathbb{F}_q)$ is given by the product of the average archimedean and p-adic masses with the local deviations:

$$\widetilde{\#}I([X,\lambda],\mathbb{F}_q) \propto \nu_\infty([X,\lambda],\mathbb{F}_q)\prod_\ell \nu_\ell([X,\lambda],\mathbb{F}_q). \tag{1-2}$$

This argument is (at best) superficially plausible. Nonetheless, in this paper we give a pure-thought proof of the following theorem:

Theorem A. *Let $[X,\lambda]$ be a principally polarized abelian variety over $\mathbb{F}_q$ with commutative endomorphism ring. Suppose that either X is ordinary or that $\mathbb{F}_q=\mathbb{F}_p$ is the prime field. Then*

$$\widetilde{\#}I([X,\lambda],\mathbb{F}_q) = q^{\frac{1}{2}\dim(\mathcal{A}_g)}\tau_T\,\nu_\infty([X,\lambda],\mathbb{F}_q)\prod_\ell \nu_\ell([X,\lambda],\mathbb{F}_q). \tag{1-3}$$

Here $\dim(\mathcal{A}_g)=\frac{1}{2}g(g+1)$ and τ_T is the Tamagawa number of the algebraic torus associated with $[X,\lambda]$ in Section 2A.

As we have mentioned, this formulation is inspired by Gekeler [2003], who proves Theorem A for an ordinary elliptic curve E over a finite prime field $\mathbb{F}_p$. (In the case $g=1$ considered by Gekeler, τ_T equals 1.) Roughly speaking, the strategy there is to compute the terms ν_ℓ explicitly, and show that the right-hand side of (1-3) actually computes, via Euler products, the value at $s=1$ of a suitable L-function. One concludes via the analytic class number formula and the known description of the isogeny class $I(E,\mathbb{F}_q)$ as a torsor under the class group of the quadratic imaginary order attached to the Frobenius of E. This strategy was redeployed in [Achter and Williams 2015] and [Gerhard and Williams 2019] for certain ordinary abelian varieties.

More recently, Achter and Gordon [2017] showed directly that the right-hand side of (1-3) actually computes the product of the volume of a certain (adelic) quotient and an orbital integral on GL_2. Thanks to the work of Langlands [1973], and Dirichlet's class number formula, one has a direct proof that this product computes the size of the isogeny class of the elliptic curve.

In fact, Langlands' formula, originally developed to count points on modular curves over finite fields, has been generalized by Kottwitz [1992] to an essentially arbitrary Shimura variety of PEL type. We review this latter formula in Proposition 2.1, and refer to it as the Langlands–Kottwitz formula below. It comes as a product of an (adelic) volume of a torus and an orbital integral, this time over GSp_{2g}. Although the orbital integral in the Langlands–Kottwitz formula clearly decomposes as a product of local terms, the volume term, however, appears as a global quantity (a class number in the case of GL_2; see [Achter and Gordon 2017, Lemma A.4]). Thus an Euler product expression for $\widetilde{\#}I([X,\lambda],\mathbb{F}_q)$ such as the one in (1-3) is, at least, not immediate.

The content of the present paper is to prove that the Euler product given by the right-hand side of (1-3) is indeed equal to the product of the global volume and the orbital integral given by Kottwitz's formula. We establish this by a delicate analysis of the interplay between various measures on the relevant spaces.

This paper is the logical extension of [Achter and Gordon 2017], which worked out these details for the case with governing group $\mathrm{GSp}_2=\mathrm{GL}_2$. The reader will correctly expect that the structure of the argument is largely similar. However, the cohomological and combinatorial intricacies of symplectic similitude groups in comparison to general linear groups — in particular, the tori are much more complex and conjugacy and stable conjugacy need not coincide — mean that each stage is considerably more involved.

We highlight three particular issues that make the generalization from elliptic curves to higher rank not straightforward.

The first is already mentioned above — the difference between conjugacy and stable conjugacy in GSp_{2g} when $g > 1$. This issue is discussed in detail in Section 3, and leads to Definition 4.1, which (as we prove in Section 3) coincides with (1-1) when $\ell \nmid p\,\mathrm{disc}(f)$.

The second is the fundamental lemma for base change, which is used to relate a Gekeler-style ratio at p to the twisted orbital integral. The complicated function one generally gets as a result of base change is the reason we have to assume that X is ordinary if $q \neq p$; this is discussed in detail in Section 4C.

The last is that the tori in GSp_{2g} for $g \geq 2$ are significantly more complicated than those for $g = 1$. The global calculation in Section 5 reflects this complexity, and involves the Tamagawa number of the algebraic torus T. This number is well known to be 1 for $g = 1$, but for general g we have to leave it as an (unknown) constant; Thomas Rüd and (independently) Wen-Wei Li obtained suggestive partial results and kindly agreed to present them in the Appendix.

Perhaps not surprisingly, (1-3) can also be interpreted as a Smith–Minkowski–Siegel type mass formula (in the sense of Tamagawa–Weil) with explicit local masses (see [Gan and Yu 2000]). Here the underlying group, of course, is GSp_{2g} and the masses calculate sizes of the relevant isogeny classes. Although this point of view is interesting in its own right we do not pursue it further in this paper. We would, however, like to note that the appearance of Tamagawa numbers is natural in this context.

The present work is, of course, part of a long and thriving discourse on the size of isogeny classes of abelian varieties over finite fields. Deuring [1941] computes the size of an isogeny class of elliptic curves as a class number. Waterhouse [1969] reinterprets and extends Deuring's work to a much larger class of abelian varieties. The key point in his study is to consider separately the ℓ-primary components of the kernel of an isogeny, and model them using the various Tate modules (for ℓ prime to the characteristic) and the Dieudonné module. Both of these approaches have been revisited and enhanced in recent times. Yu and collaborators have undertaken a detailed analysis of, for example, the number of supersingular abelian varieties over finite fields; their answers are often expressed in terms of a mass formula which comes from a local lattice-theoretic perspective on the isogeny problem (e.g., [Xue and Yu 2021; Yu 2012]). In a somewhat different direction, Marseglia [2021] and Howe [2022] are often able to express the size of an isogeny class of principally polarized abelian varieties as a sum of suitably generalized class numbers.

In the approach taken here, the orbital integral in the Langlands–Kottwitz formula (Proposition 2.1) does the work of assimilating information from the different local lattice calculations. Class numbers necessarily arise from our formula, but are not built-in; see Section 6 for this emergence.

Notation. We work over a finite field $\mathbb{F}_q = \mathbb{F}_{p^e}$ of characteristic p. We will often drop the field from all notation. We denote by $\mathbb{Q}_q$ the degree e unramified extension of $\mathbb{Q}_p$, and by $\mathbb{Z}_q$ its ring of integers; the p-th power automorphism of $\mathbb{F}_q$ lifts to the Frobenius automorphism σ of $\mathbb{Q}_q$.

Fix a positive integer g. Let $V = \mathbb{Z}^{\oplus 2g}$, endowed with basis $x_1, \dots, x_g, y_1, \dots, y_g$, and equip it with the symplectic form such that $\langle x_i, y_j\rangle = \delta_{ij}$ and $\langle x_i, x_j\rangle = \langle y_i, y_j\rangle = 0$. Let G be the group of similitudes $G = \mathrm{GSp}(V, \langle\cdot,\cdot\rangle) \cong \mathrm{GSp}_{2g}$; it has dimension $2g^2+g+1$ and rank $r = g+1$. Its derived group is $G^{\mathrm{der}} \cong \mathrm{Sp}_{2g}$, and $G/G^{\mathrm{der}} \cong \mathbb{G}_m$. Let $\eta : G \to \mathbb{G}_m$ be the corresponding surjection (the multiplier map). We write T_{spl} for the split torus of diagonal matrices in G.

If K is a field, $\alpha \in G(K)$, and $\Gamma \subseteq G(K)$, we let ${}^{\Gamma}\alpha = \{\beta^{-1}\alpha\beta : \beta \in \Gamma\}$ be the orbit of α under Γ. Since G has simply connected derived group, the stable conjugacy class, or stable orbit, of α is those elements of $G(K)$ which are conjugate to α as elements of $G(\bar{K})$ [Kottwitz 1982, p. 785].

For $\alpha \in G(\mathbb{Q}_q)$, the twisted, or σ-, conjugacy class of α is $\{\beta^{-1}\alpha\beta^{\sigma} : \beta \in G(\mathbb{Q}_q)\}$; and the stable twisted conjugacy class of α is $G(\mathbb{Q}_q) \cap \{\beta^{-1}\alpha\beta^{\sigma} : \beta \in G(\overline{\mathbb{Q}}_q)\}$.

For an element $\gamma \in G(\mathbb{Q}_\ell)$, where ℓ is an arbitrary prime, the Weyl discriminant of γ is denoted by $D(\gamma)$: $D(\gamma) = \prod_{\alpha\in\Phi}(1-\alpha(\gamma))$, where the product is over all roots of G (see Section 6A1 for details).

2. Background

2A. ***The Kottwitz formula.*** The key formula we need is developed by Kottwitz [1992]. In fact, the special case we need is detailed in [Kottwitz 1990, Section 12]. By way of establishing necessary notation, we review the relevant part of this work here.

Let $\mathcal{A}_g$ denote the moduli space of principally polarized abelian varieties of dimension g. An isogeny between two principally polarized abelian varieties $[X, \lambda], [Y, \mu] \in \mathcal{A}_g(\mathbb{F}_q)$ is an isogeny $\phi : X \to Y$ such that $m\phi^*\mu = n\cdot\lambda$ for some nonzero integers m and n. The isogeny class $I([X, \lambda], \mathbb{F}_q)$ is the set of all principally polarized abelian varieties $[Y, \mu]/\mathbb{F}_q$ admitting such an isogeny (over $\mathbb{F}_q$), and its weighted cardinality is

$$\widetilde{\#} I([X, \lambda], \mathbb{F}_q) = \sum_{[Y,\mu]\in I([X,\lambda],\mathbb{F}_q)} \frac{1}{\#\mathrm{Aut}(Y, \mu)}.$$

The abelian variety $X/\mathbb{F}_q$ admits a Frobenius endomorphism $\varpi_{X/\mathbb{F}_q}$, with characteristic polynomial $f_{X/\mathbb{F}_q}(T)$ of degree $2g$. (By [Tate 1966], this polynomial determines the isogeny class of X as an unpolarized abelian variety.)

For each $\ell \neq p$, $H^1(X_{\overline{\mathbb{F}}_q}, \mathbb{Z}_\ell)$ (the dual of the Tate module) is a free $\mathbb{Z}_\ell$-module of rank $2g$, endowed with a symplectic pairing $\langle\cdot,\cdot\rangle_\lambda$ induced by the polarization. The Frobenius endomorphism $\varpi_{X/\mathbb{F}_q}$ induces an element $\gamma_{X/\mathbb{F}_q,\ell} \in \mathrm{GSp}(H^1(X_{\overline{\mathbb{F}}_q}, \mathbb{Z}_\ell), \langle\cdot,\cdot\rangle_\lambda)$, and thus an element of $G(\mathbb{Z}_\ell)$, well-defined up to conjugacy. Moreover, there is an equality of characteristic polynomials $f_{\gamma_{X/\mathbb{F}_q,\ell}}(T) = f_{X/\mathbb{F}_q}(T)$. Simultaneously considering all finite primes $\ell \neq p$, we obtain an adelic similitude $\gamma_{[X,\lambda]} \in G(\mathbb{A}_f^p)$. (Alternatively one can, of course, directly consider the action of $\varpi_{X/\mathbb{F}_q}$ on $H^1(X_{\overline{\mathbb{F}}_q}, \widehat{\mathbb{Z}}^p) = \lim_{\overleftarrow{p\nmid n}} H^1(X_{\overline{\mathbb{F}}_q}, \mathbb{Z}/n)$.)

Similarly, the crystalline cohomology group $H^1_{\mathrm{cris}}(X, \mathbb{Q}_q)$ is endowed with an integral structure $H^1_{\mathrm{cris}}(X, \mathbb{Z}_q)$ and a σ-linear endomorphism F. It determines, up to σ-conjugacy, an element $\delta_{X/\mathbb{F}_q}$ of $G(\mathbb{Q}_q)$ with multiplier $\eta(\delta_{X/\mathbb{F}_q}) = p$.

The e-th iterate of F is linear, and in fact F^e is the endomorphism of $H^1_{\mathrm{cris}}(X, \mathbb{Q}_q)$ induced by $\varpi_{X/\mathbb{F}_q}$.

Let $T_{[X,\lambda]}/\mathbb{Q}$ represent the automorphism group of $[X, \lambda]$ in the category of abelian varieties up to $\mathbb{Q}$-isogeny. Concretely, the polarization λ induces a (Rosati) involution $(\dagger)$ on $\mathrm{End}(X) \otimes \mathbb{Q}$; and for each $\mathbb{Q}$-algebra R, we have

$$T_{[X,\lambda]}(R) = \{\alpha \in (\mathrm{End}(X) \otimes R)^\times : \alpha\alpha^{(\dagger)} \in R^\times\}.$$

By Tate's theorem [1966], for $\ell \neq p$, $T_{[X,\lambda]}(\mathbb{Q}_\ell)$ is isomorphic to $G_{\gamma_{X/\mathbb{F}_q,\ell}}(\mathbb{Q}_\ell)$, the centralizer of $\gamma_{X/\mathbb{F}_q,\ell}$, and $T_{[X,\lambda]}(\mathbb{Q}_q)$ is isomorphic to $G_{\delta_{X/\mathbb{F}_q}\sigma}(\mathbb{Q}_p)$, the twisted centralizer of $\delta_{X/\mathbb{F}_q}$ in $G(\mathbb{Q}_q)$.

A direct analysis of the effect of isogenies on the first cohomology groups of abelian varieties shows:

Proposition 2.1 [Kottwitz 1990]. *The weighted cardinality of the isogeny class of* $[X, \lambda] \in \mathcal{A}_g(\mathbb{F}_q)$ *is*

$$\widetilde{\#}I([X,\lambda],\mathbb{F}_q) = \mathrm{vol}(T_{[X,\lambda]}(\mathbb{Q})\backslash T_{[X,\lambda]}(\mathbb{A}_f)) \cdot \int_{G_{\gamma_{X/\mathbb{F}_q}}(\mathbb{A}_f^p)\backslash G(\mathbb{A}_f^p)} \mathbb{1}_{G(\widehat{\mathbb{Z}}_f^p)}(g^{-1}\gamma_{X/\mathbb{F}_q} g)\,dg \cdot \int_{G_{\delta_{X/\mathbb{F}_q}\sigma}(\mathbb{Q}_p)\backslash G(\mathbb{Q}_q)} \mathbb{1}_{G(\mathbb{Z}_q)\,\mathrm{diag}(p,\dots,p,1,\dots,1)G(\mathbb{Z}_q)}(h^{-1}\delta_{X/\mathbb{F}_q}h^\sigma)\,dh. \quad (2\text{-}1)$$

In the orbital and twisted orbital integrals in (2-1), we choose the Haar measures on G which assign volume 1 to $G(\widehat{\mathbb{Z}}^p)$ and to $G(\mathbb{Z}_q)$, respectively. The choice of measure on T does not matter here, as long as the same measure is used to calculate the global volume. We define the specific measure on T in Section 5. It coincides with the canonical measure at all but finitely many places.

This formula appears in [Kottwitz 1990, p. 205]; see also [Kottwitz 1992] for its generalization to a much larger class of PEL Shimura varieties. As in [Achter and Gordon 2017, 2.4], the weighted cardinality accounts for the fact that we have not introduced a rigidifying level structure, and thus our objects admit nontrivial, albeit finite, automorphism groups.

Remark 2.2. Using Honda–Tate theory, one can find $\gamma_{X/\mathbb{F}_q,0} \in G(\mathbb{Q})$, well-defined up to $G(\overline{\mathbb{Q}})$-conjugacy, such that $\gamma_{X/\mathbb{F}_q,0}$ and $\gamma_{X/\mathbb{F}_q,\ell}$ are conjugate in $G(\overline{\mathbb{Q}}_\ell)$; see Kottwitz [1990, p. 206; 1992, p. 422]. Similarly, $\gamma_{X/\mathbb{F}_q,0}$ and $\mathrm{N}\,\delta_{X/\mathbb{F}_q}$ are conjugate in $G(\overline{\mathbb{Q}}_q)$, where N denotes the norm map

$$G(\mathbb{Q}_q) \to G(\mathbb{Q}_q), \quad \alpha \mapsto \alpha\alpha^\sigma \cdots \alpha^{\sigma^{e-1}}.$$

In particular, the characteristic polynomial of $\gamma_{X/\mathbb{F}_q,0}$ is $f_{X/\mathbb{F}_q}(T)$. In fact, by adjusting $\delta_{X/\mathbb{F}_q}$ in its twisted conjugacy class, we henceforth can and will assume that

$$\mathrm{N}(\delta_{X/\mathbb{F}_q}) \in G(\mathbb{Q}_p) \subset G(\mathbb{Q}_q) \quad (2\text{-}2)$$

[Kottwitz 1982, p. 206]. Then the group variety $T_{[X,\lambda]/\mathbb{F}_q}$ is isomorphic to the centralizer of $\mathrm{N}(\gamma_{X/\mathbb{F}_q,0})$ in G.

It turns out that, moreover, one can find a rational element $\gamma_0 \in G(\mathbb{Q})$ such that γ_0 is $G(\mathbb{Q}_\ell)$-conjugate to $\gamma_{X/\mathbb{F}_q,\ell}$ for every $\ell \neq p$ (see [Kisin 2017, p. 889]). Consequently, in (2-1) we could replace $\gamma_{X/\mathbb{F}_q}$ with a global object γ_0; but we will never use this fact in this paper.

In the remainder of this paper we fix a principally polarized abelian variety $[X, \lambda]/\mathbb{F}_q$ with commutative endomorphism ring $\mathrm{End}(X)$. (For example, any simple, ordinary abelian variety necessarily has a commutative endomorphism ring [Waterhouse 1969, Theorem 7.2].) By Tate's theorem, the commutativity of $\mathrm{End}(X)$ is equivalent to the condition that $T_{[X,\lambda]}$ is a maximal torus in G.

To ease notation slightly, we will write δ_0 and T for $\delta_{X/\mathbb{F}_q}$ and $T_{[X,\lambda]}$, respectively. If ℓ is a fixed, notationally suppressed prime, we will sometimes write γ_0 for $\gamma_{X/\mathbb{F}_q,\ell}$; by Remark 2.2, one may equally well let γ_0 be the image of some choice γ_0 in $G(\mathbb{Q})$ (though we will not be using it).

2B. ***Structure of the centralizer.*** For future use, we record some information about the centralizer $T = T_{[X,\lambda]}$. Recall that X is a g-dimensional abelian variety with commutative endomorphism ring. Then T is a maximal torus in G, and $K := \mathrm{End}(X)^0 = \mathrm{End}(X)\otimes\mathbb{Q}$ is a CM-algebra of degree $2g$ over $\mathbb{Q}$. Then K is isomorphic to a direct sum $K \cong \bigoplus_{i=1}^t K_i$ of CM-fields, and the Rosati involution on $\mathrm{End}(X)$ induces a positive involution $a \mapsto \bar{a}$ on K, which in turn restricts to complex conjugation on each component K_i. Let $K^+ \subset K$ be the subalgebra fixed by the positive involution. Then $K^+ \cong \bigoplus_{i=1}^t K_i^+$, where K_i^+ is the maximal totally real subfield of K_i, and $[K^+ : \mathbb{Q}] = g$.

In general, if L is a field and M/L is a finite étale algebra, let $\boldsymbol{R}_{M/L}$ be Weil's restriction of scalars functor. The norm map $\mathrm{N}_{M/L}$ induces a map of tori $\boldsymbol{R}_{M/L}\mathbb{G}_m \to \mathbb{G}_m$, and the norm-one torus is the kernel of the map

$$1 \to \boldsymbol{R}^{(1)}_{M/L}\mathbb{G}_m \to \boldsymbol{R}_{M/L}\mathbb{G}_m \xrightarrow{\mathrm{N}_{M/L}} \mathbb{G}_m \to 1.$$

With these preparations we have

$$T^{\mathrm{der}} := T \cap G^{\mathrm{der}} \cong \boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m,$$

and T sits in the diagram

$$\begin{tikzcd}
1 \arrow[r] & T^{\mathrm{der}} \arrow[r] \arrow[d, "\sim"] & T \arrow[r] \arrow[d, hook] & \mathbb{G}_m \arrow[r] \arrow[d, hook] & 1 \\
1 \arrow[r] & \boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m \arrow[r] & \boldsymbol{R}_{K/\mathbb{Q}}\mathbb{G}_m \arrow[r, "\boldsymbol{R}_{K^+/\mathbb{Q}}\mathrm{N}_{K/K^+}"] & \boldsymbol{R}_{K^+/\mathbb{Q}}\mathbb{G}_m \arrow[r] & 1
\end{tikzcd} \tag{2-3}$$

On points, we have

$$T(\mathbb{Q}) = \{a \in K^\times : a\bar{a} \in \mathbb{Q}^\times\} = \left\{(a_1,\dots,a_t) \in \bigoplus K_i^\times : \text{there exists } c \in \mathbb{Q}^\times \text{ such that } a_i\bar{a}_i = c\right\},$$
$$T^{\mathrm{der}}(\mathbb{Q}) = \{a \in K^\times : a\bar{a} = 1\} = \left\{(a_1,\dots,a_t) \in \bigoplus K_i^\times : a_i\bar{a}_i = 1\right\}.$$

Let $\widetilde{T} = T^{\mathrm{der}} \times \mathbb{G}_m$. It is not hard to write down an explicit isogeny $\alpha : \widetilde{T} \to T$ and a complementary isogeny $\beta : T \to \widetilde{T}$ such that $\alpha\circ\beta$ is the squaring map. We choose the maps which, on points, are given by

$$T \xrightarrow{\beta} \widetilde{T} \xrightarrow{\alpha} T, \qquad a \overset{\beta}{\longmapsto} (a\bar{a}^{-1}, a\bar{a}), \quad (b,c) \overset{\alpha}{\longmapsto} bc.$$

2C. ***The Steinberg quotient.*** Recall that we have fixed a maximal split torus T_{spl} in G; let W be the Weyl group of G relative to T_{spl}. Let $T^{\mathrm{der}}_{\mathrm{spl}} = T_{\mathrm{spl}} \cap G^{\mathrm{der}}$, and let $A^{\mathrm{der}} = T^{\mathrm{der}}_{\mathrm{spl}}/W$ be the Steinberg quotient for the semisimple group G^{der}. It is isomorphic to the affine space of dimension $r-1=g$.

We let $\mathbb{A}_G = A^{\mathrm{der}} \times \mathbb{G}_m$ be the analogue of the Steinberg quotient for the reductive group G, and define a map

$$G \xrightarrow{\mathfrak{c}} \mathbb{A}_G, \quad \gamma \mapsto \big(\mathrm{tr}(\gamma), \mathrm{tr}\big(\textstyle\bigwedge^2 \gamma\big), \ldots, \mathrm{tr}\big(\textstyle\bigwedge^g \gamma\big), \eta(\gamma)\big). \tag{2-4}$$

Note that $\eta(\gamma) = \mathrm{tr}\big(\bigwedge^{g+1}(\gamma)\big)/\mathrm{tr}(\gamma)$; and if $\gamma \in G^{\mathrm{der}} \subset G$, then $\mathfrak{c}(\gamma) = (\mathfrak{c}^{\mathrm{der}}(\gamma), 1)$, where $\mathfrak{c}^{\mathrm{der}}$ is the usual Steinberg map.

2D. *Truncations.* Let ℓ be any finite prime (including $\ell = p$). Let $\pi_n = \pi_{\ell,n} : \mathbb{Z}_\ell \to \mathbb{Z}_\ell/\ell^n$ be the truncation map. For any $\mathbb{Z}_\ell$-scheme $\mathcal{X}$, we denote by $\pi_n^{\mathcal{X}}$ the corresponding map

$$\pi_n^{\mathcal{X}} : \mathcal{X}(\mathbb{Z}_\ell) \to \mathcal{X}(\mathbb{Z}_\ell/\ell^n)$$

induced by π_n. Given $S_n \subset \mathcal{X}(\mathbb{Z}_\ell/\ell^n)$, we will often set

$$\widetilde{S}_n = \pi_n^{-1}(S_n).$$

The projection maps π_n^G extend to a somewhat larger set of similitudes. Let $M(\mathbb{Z}_\ell)$ be the set of symplectic similitudes which stabilize the lattice $V \otimes \mathbb{Z}_\ell$. Then

$$M(\mathbb{Z}_\ell) = \mathrm{GSp}(V \otimes \mathbb{Q}_\ell) \cap \mathrm{End}(V \otimes \mathbb{Z}_\ell) \cong \mathrm{GSp}_{2g}(\mathbb{Q}_\ell) \cap \mathrm{Mat}_{2g}(\mathbb{Z}_\ell).$$

Inside this set, for each $d \geq 0$ we identify a subset

$$M(\mathbb{Z}_\ell)_d = \{A \in M(\mathbb{Z}_\ell) : \mathrm{ord}_\ell \det(A) \leq d\}.$$

Finally, let us denote by $M(\mathbb{Z}_\ell/\ell^n)_d$ the set

$$M(\mathbb{Z}_\ell/\ell^n)_d = \{A \in M(\mathbb{Z}_\ell/\ell^n) : \mathrm{ord}_\ell \det(A) \leq d\}.$$

Note that $M(\mathbb{Z}_\ell)_0 = G(\mathbb{Z}_\ell)$, and in the last definition, the condition on the determinant is not vacuous even if $d \gg n$, because it rules out the matrices of determinant zero.

With a certain amount of abuse, we introduce the following notion of "$M(\mathbb{Z}_\ell)_d$-conjugacy":

Definition 2.3. If $\gamma \in M(\mathbb{Z}_\ell)$ and in particular if $\gamma \in G(\mathbb{Z}_\ell)$, we will write $\gamma \sim_{M(\mathbb{Z}_\ell)_d} \gamma_0$ if there exists some $A \in M(\mathbb{Z}_\ell)_d$ such that $A\gamma = \gamma_0 A$.

Similarly, if $\bar{\gamma} \in M(\mathbb{Z}_\ell/\ell^n)$, we write $\bar{\gamma} \sim_{M(\mathbb{Z}_\ell/\ell^n)_d} \gamma_0$ if there exists some $\bar{A} \in M(\mathbb{Z}_\ell/\ell^n)_d$ such that $\bar{A}\gamma = \pi_n(\gamma_0)\bar{A}$.

When n is small relative to d, truncations of $M(\mathbb{Z}_\ell)_d$- conjugate elements might not be $M(\mathbb{Z}_\ell/\ell^n)_d$-conjugate (since, e.g., all the elements $A \in M(\mathbb{Z}_\ell)$ satisfying $A\gamma = \gamma_0 A$ might project to 0 mod ℓ^n). Of course, this does not happen when $n \gg d$. We also note that trivially, if $\gamma \sim_{M(\mathbb{Z}_\ell)_{d_0}} \gamma_0$ for some d_0, then $\gamma \sim_{M(\mathbb{Z}_\ell)_d} \gamma_0$ for all $d \geq d_0$. The analogous statement holds for $\bar{\gamma} \in G(\mathbb{Z}_\ell/\ell^n)$ as long as $n \gg d$.

2E. *Measures and integrals.* As in [Achter and Gordon 2017], we need to explicitly work out the relationships between several different natural measures on the ℓ-adic points of varieties, especially groups and group orbits. The definitions introduced in [Achter and Gordon 2017, §3] (where a little more historical perspective is briefly reviewed) go through with minimal changes. We recall the relevant notation here.

Serre–Oesterlé measure: Serre [1981, §3], observed that for a smooth p-adic submanifold Y of $\mathbb{Z}_p^m$ of dimension d, there is a limit $\lim_{n\to\infty}|Y_n|p^{-nd}$, where Y_n is the reduction of Y modulo p^n (in our notation, $Y_n = \pi_n(Y)$). Moreover, Serre pointed out that this limit can be understood as the *volume* of Y with respect to a certain measure, which is canonical. The definition of this measure for more general sets Y was elaborated on by Oesterlé [1982] and by Veys [1992]. We refer to this measure as the Serre–Oesterlé measure, and denote it by μ^{SO}.

Measures on groups: Once and for all, we fix the measure $|dx|_\ell$ on the affine line $\mathbb{A}^1_{\mathbb{Q}_\ell}$ to be the translation-invariant measure such that $\mathrm{vol}_{|dx|_\ell}(\mathbb{Z}_\ell) = 1$. Then there are two fundamentally different approaches to defining measure. The first is, for any smooth algebraic variety $\mathcal{X}$ over $\mathbb{Q}_\ell$ with a nonvanishing top degree differential form ω on it, one gets the associated measure $|d\omega|_\ell$ on $\mathcal{X}(\mathbb{Q}_\ell)$. In particular, for a reductive group G, there is a canonical differential form ω_G, defined in the greatest generality by Gross [1997]. This gives a canonical measure $|d\omega_G|_\ell$ on $G(\mathbb{Q}_\ell)$. When G is split over $\mathbb{Q}$, this measure has an alternative description using point-counting over the finite field (i.e., it coincides with the Serre–Oesterlé measure μ_G^{SO} defined above):

$$\int_{G(\mathbb{Z}_\ell)} |d\omega_G|_\ell = \frac{\#G(\mathbb{Z}/\ell)}{\ell^{\dim(G)}}. \tag{2-5}$$

This observation is originally due to A. Weil [1982], and is actually built into his definition of integration on adeles. Weil's classical observation is precisely what makes this paper possible.

For groups, there is a second approach. Start with a Haar measure and normalize it so that some given maximal subgroup has volume 1. The choice of a "canonical" compact subgroup in this approach could lead to very interesting considerations (and is one of the main points of [Gross 1997]), but in our situation only one easy case is needed. For $G(\mathbb{Q}_\ell)$, the relevant maximal subgroup is $G(\mathbb{Z}_\ell)$; we denote such a Haar measure on $G(\mathbb{Q}_\ell)$ by μ_G^{can}.

Geometric measure on orbits: This is a measure constructed in [Frenkel et al. 2010] on a fibre of the Steinberg map $\mathfrak{c}: G \to \mathbb{A}_G$. Let ω_G be a volume form on G, and let ω_A be the volume form $\bigwedge dx_i \wedge (dx/|x|)$ on $\mathbb{A}_G \cong \mathbb{A}^{\mathrm{rank}(G)-1} \times \mathbb{G}_m$. On the fibre $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma))$, factor ω_G as

$$\omega_G = \omega^{\mathrm{geom}}_{\mathfrak{c}(\gamma)} \wedge \omega_A;$$

integrating $|\omega^{\mathrm{geom}}_{\mathfrak{c}(\gamma)}|$ defines a measure μ^{geom} on $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma))$.

Suppose ϕ is a locally constant compactly supported function on $G(\mathbb{Q}_\ell)$. Recall the family $\gamma_{X/\mathbb{F}_q,\ell}$ (and δ_0), whose centralizers are the sets of $\mathbb{Q}_\ell$-points of the algebraic torus $T := T_{[X,\lambda]}$. We use two different measures on the orbit ${}^{G(\mathbb{Q}_\ell)}\gamma_{X/\mathbb{F}_q,\ell} \cong T_{[X,\lambda]}(\mathbb{Q}_\ell)\backslash G(\mathbb{Q}_\ell)$ to define an integral. When ℓ is fixed, we will often denote the element $\gamma_{X/\mathbb{F}_q,\ell}$ by γ_0; we define $\mu_{\gamma_0}^{\mathrm{Tama}}$ as the quotient measure $\mu_G^{\mathrm{can}}/\mu_T^{\mathrm{Tama}}$, where μ_T^{Tama} is the Tamagawa measure on T defined below in Section 5A, and let $\mu_{\gamma_0}^{\mathrm{geom}}$ be the geometric measure reviewed above. (Since the orbit of γ_0 is an open subset of $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0))$, the restriction of the geometric measure from $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0))$ to the orbit makes sense.) Then for $\bullet \in \{\mathrm{Tama}, \mathrm{geom}\}$, set

$$O^\bullet_{\gamma_0} := \int_{T(\mathbb{Q}_\ell)\backslash G(\mathbb{Q}_\ell)} \phi(g^{-1}\gamma_0 g)\, d\mu^\bullet_{\gamma_0}.$$

3. Conjugacy

3A. *Integral conjugacy.* To relate the right-hand side of (2-1) to the ratios ν_ℓ of (1-1), we interpret the orbital integral as the volume of the intersection of the $G(\mathbb{Q}_\ell)$-orbit of γ_0 with $G(\mathbb{Z}_\ell)$. For almost all ℓ, $G(\mathbb{Z}_\ell) \cap^{G(\mathbb{Q}_\ell)} \gamma_0 =^{G(\mathbb{Z}_\ell)} \gamma_0$:

Lemma 3.1. *Suppose $\gamma_0 \in G(\mathbb{Z}_\ell)$ and $\ell \nmid D(\gamma_0)$. If $\gamma \in G(\mathbb{Z}_\ell)$, then*

$$\gamma \sim_{G(\mathbb{Q}_\ell)} \gamma_0 \iff \gamma \sim_{G(\mathbb{Z}_\ell)} \gamma_0.$$

Proof. The hypothesis on γ_0 implies that the centralizer G_{γ_0} is a smooth torus over $\mathbb{Z}_\ell$, and thus the transporter from G_γ to G_{γ_0} is smooth over $\mathbb{Z}_\ell$ (e.g., [Conrad 2014, Proposition 2.1.2]).

Since γ and γ_0 are conjugate in $G(\mathbb{Q}_\ell)$, they have the same characteristic polynomial, and thus their reductions $\bar{\gamma}_0 = \pi_1^G(\gamma_0)$ and $\bar{\gamma} = \pi_1^G(\gamma)$ are stably conjugate in $G(\mathbb{Z}_\ell/\ell)$. By Lang's theorem, $\bar{\gamma}_0$ and $\bar{\gamma}$ are conjugate in $G(\mathbb{Z}_\ell/\ell)$; by smoothness of the transporter scheme, γ and γ_0 are conjugate in $G(\mathbb{Z}_\ell)$. □

If $\bar{\gamma}_0$ is *not* regular, then the set $G(\mathbb{Z}_\ell) \cap^{G(\mathbb{Q}_\ell)} \gamma_0$ generally consists of several different $G(\mathbb{Z}_\ell)$-orbits. Nonetheless, the number of distinct orbits is bounded; and membership in ${}^{G(\mathbb{Q}_\ell)}\gamma_0$ can be detected at a finite truncation level.

Lemma 3.2. *Suppose $\gamma_0 \in G(\mathbb{Z}_\ell)$ is regular semisimple. There exists an integer $e = e(\gamma_0)$ such that, if $n \gg 0$ and $d > e$, then for $\gamma \in G(\mathbb{Z}_\ell/\ell^n)$, the following conditions are equivalent*:

(1) $\gamma \sim_{M(\mathbb{Z}_\ell/\ell^n)_d} \gamma_0 \bmod \ell^n$, *and*

(2) *there exists some $\tilde{\gamma} \in G(\mathbb{Z}_\ell)$ such that $\tilde{\gamma} \bmod \ell^n = \gamma$ and $\tilde{\gamma} \sim_{G(\mathbb{Q}_\ell)} \gamma_0$.*

The statement is also true with $G(\mathbb{Z}_\ell)$ replaced with $M(\mathbb{Z}_\ell)$ everywhere.

Proof. We prove the original statement.

The intersection of $G(\mathbb{Z}_\ell)$ with the $G(\mathbb{Q}_\ell)$-orbit of γ_0 is a finite union of $G(\mathbb{Z}_\ell)$-orbits, since it is compact (recall that γ_0 is regular semisimple) and the $G(\mathbb{Z}_\ell)$-orbits are open in this intersection; let $g_1, \dots, g_s$ be representatives of these orbits, and let $A_i \in G(\mathbb{Q}_\ell)$ be elements satisfying $A_i g_i A_i^{-1} = \gamma_0$, so that $A_i g_i = \gamma_0 A_i$. We clear denominators; for each i, let $X_i \in M(\mathbb{Z}_\ell)$ be a scalar multiple of A_i. Then $X_i g_i = \gamma_0 X_i$, and we set

$$e(\gamma_0) = \max_{i \in \{1,\dots,s\}} \{|\mathrm{ord}(\det X_i)|\}.$$

Now, suppose $n > 2d(\gamma_0)$, where $d(\gamma_0)$ is the valuation of the discriminant of γ_0, $e \geq e(\gamma_0)$, and $n \gg e$. We want to prove that with these assumptions, an element $\gamma \in G(\mathbb{Z}_\ell/\ell^n)$ satisfies

$$\gamma \sim_{M(\mathbb{Z}_\ell/\ell^n)_e} \pi_n(\gamma_0)$$

if and only if there exists a lift $\tilde{\gamma} \in G(\mathbb{Z}_\ell)$ such that $\pi_n(\tilde{\gamma}) = \gamma$ and $\tilde{\gamma} \sim_{G(\mathbb{Q}_\ell)} \gamma_0$.

One direction is easy: suppose there exists $\tilde{\gamma} \in G(\mathbb{Z}_\ell)$ such that $\tilde{\gamma} \bmod \ell^n = \gamma$ and $\tilde{\gamma} \sim_{G(\mathbb{Q}_\ell)} \gamma_0$. Then there exists $i \in \{1, \dots, s\}$ such that $\tilde{\gamma} \sim_{G(\mathbb{Z}_\ell)} g_i$. Therefore there exists $Y \in G(\mathbb{Z}_\ell)$ such that $Y\tilde{\gamma} = g_i Y$.

Recall that, as above, there exists $X_i \in M(\mathbb{Z}_\ell)$ such that $X_i g_i = \gamma_0 X_i$. Then $Z := \pi_n^M(X_i Y)$ lies in $M(\mathbb{Z}_\ell/\ell^n)_e$ and satisfies the condition $Z\gamma = \pi_n(\gamma_0)Z$.

The other direction is a special case of Hensel's lemma. Since Hensel's lemma in this generality, though well-known, is surprisingly hard to find in the literature, we provide a detailed explanation with references.

For each n, let

$$R_{\gamma_0}(\mathbb{Z}_\ell/\ell^n) = \left\{(A,\gamma) : A \in M(\mathbb{Z}_\ell/\ell^n),\ \gamma \in G(\mathbb{Z}_\ell/\ell^n),\ A\gamma = \pi_n^G(\gamma_0)A\right\} \subset M(\mathbb{Z}_\ell/\ell^n) \times G(\mathbb{Z}_\ell/\ell^n),$$

where π_n^G is the projection from Section 2D. This is a system of $(2g)^2$ equations in $8g^2$ variables (namely, the matrix entries of A and γ). Now Hensel's lemma as stated in [Bourbaki 1985, III.4.5., Corollaire 3, p. 271] applies directly, as follows. Let $n(\gamma_0)$ be the valuation of the minor formed by the first $(2g)^2$ columns of the Jacobian matrix of this system of equations at γ_0. By Hensel's lemma, if $n > 2n(\gamma_0)$ and $(A,\gamma) \in R_{\gamma_0}(\mathbb{Z}_\ell/\ell^n)$, then there exists some $\widetilde{\gamma} \in G(\mathbb{Z}_\ell)$ such that $\pi_n(\widetilde{\gamma}) = \gamma$ and $\widetilde{\gamma} \sim_{M(\mathbb{Z}_\ell)} \gamma_0$.

Since the core argument simply relies on the solvability, via Hensel's lemma, of a system of equations over $\mathbb{Z}_\ell$, it is also valid if $G(\mathbb{Z}_\ell)$ is replaced by $M(\mathbb{Z}_\ell)$. □

Remark 3.3. We observe (though we do not need this observation in this paper) that $n(\gamma_0)$ in fact equals the valuation of the discriminant of γ_0, e.g., by the argument provided in [Kottwitz 2005, §7.2].

For $\gamma_0 \in G(\mathbb{Z}_\ell)$, let

$$C_{(d,n)}(\gamma_0) = \{\gamma \in G(\mathbb{Z}_\ell/\ell^n) : \gamma \sim_{M(\mathbb{Z}_\ell/\ell^n)_d} \gamma_0\}. \tag{3-1}$$

If $d = 0$, this coincides with the usual conjugacy class of $\pi_n(\gamma_0)$. As in Section 2D, let $\widetilde{C}_{(d,n)}(\gamma_0) = (\pi_n^G)^{-1}(C_{(d,n)}(\gamma_0))$ be the set of lifts of elements of $C_{(d,n)}(\gamma_0)$ to $G(\mathbb{Z}_\ell)$.

We also extend this notation to elements $\gamma_0 \in M(\mathbb{Z}_\ell)$:

$$C_{(d,n)}(\gamma_0) = \{\gamma \in M(\mathbb{Z}_\ell/\ell^n) : \gamma \sim_{M(\mathbb{Z}_\ell/\ell^n)_d} \gamma_0\}.$$

(If $\gamma_0 \in G(\mathbb{Z}_\ell) \subset M(\mathbb{Z}_\ell)$, the two notions coincide and thus there is no ambiguity.)

Corollary 3.4. (a) *Suppose $\gamma_0 \in G(\mathbb{Z}_\ell)$. There exists $d = d(\gamma_0)$ such that, if $n \gg 0$, then*

$$C_{(d,n)}(\gamma_0) = \pi_n(G(\mathbb{Z}_\ell) \cap {}^{G(\mathbb{Q}_\ell)}\gamma_0).$$

Moreover,

$$G(\mathbb{Z}_\ell) \cap {}^{G(\mathbb{Q}_\ell)}\gamma_0 = \bigcap_{n \geq 0} \widetilde{C}_{(d,n)}(\gamma_0).$$

(b) *Suppose $\gamma_0 \in M(\mathbb{Z}_\ell)$. There exists $d = d(\gamma_0)$ such that, if $n \gg 0$, then*

$$C_{(d,n)}(\gamma_0) = \pi_n(M(\mathbb{Z}_\ell) \cap {}^{G(\mathbb{Q}_\ell)}\gamma_0).$$

Moreover,

$$M(\mathbb{Z}_\ell) \cap {}^{G(\mathbb{Q}_\ell)}\gamma_0 = \bigcap_{n \geq 0} \widetilde{C}_{(d,n)}(\gamma_0).$$

Proof. This is a direct consequence of Lemma 3.2. □

3B. ***Stable (twisted) conjugacy.*** In this section, we further assume that $[X, \lambda]$ is a principally polarized abelian variety with commutative endomorphism ring for which $\frac{1}{2}$ is not a slope of the Newton polygon of X. (Again, any ordinary simple principally polarized abelian variety satisfies these hypotheses.)

Recall the definition of K and K^+, as well as the discussion of T, from Section 2B. By a prime of K (or K^+) lying over p we mean a prime $\mathfrak{p}$ of some K_i (respectively, K_i^+) lying over p, and we write $K_\mathfrak{p}$ for $K_{i,\mathfrak{p}}$. With this convention, we then have $K \otimes \mathbb{Q}_p \cong \bigoplus K_\mathfrak{p}$.

Lemma 3.5. *Let $\mathfrak{p}^+$ be a prime of K^+ lying over p. Then $\mathfrak{p}^+$ splits in K.*

Proof. This is standard. We work in the category of p-divisible groups up to isogeny. Then $X[p^\infty]$ has height $2g$, and comes equipped with an action by $K \otimes_\mathbb{Q} \mathbb{Q}_p$.

Corresponding to the decomposition $K^+ \otimes_\mathbb{Q} \mathbb{Q}_p \cong \bigoplus_{\mathfrak{p}^+|p} K^+_{\mathfrak{p}^+}$ we have the decomposition

$$X[p^\infty] = \bigoplus X[\mathfrak{p}^{+\infty}].$$

Moreover, $X[\mathfrak{p}^{+\infty}]$ is a p-divisible group of height $2[K^+_{\mathfrak{p}^+} : \mathbb{Q}_p]$, and self-dual (because $\mathfrak{p}^+\mathcal{O}_K$ is stable under the Rosati involution). We now fix one $\mathfrak{p}^+$, and show that it must split in K.

Since $K^+_{\mathfrak{p}^+}$ is a field (and not just a $\mathbb{Q}_p$-algebra) of dimension $\frac{1}{2}\,\mathrm{ht}(X[\mathfrak{p}^{+\infty}])$, $X[\mathfrak{p}^{+\infty}]$ has at most two slopes. Since by hypothesis $\frac{1}{2}$ is not a slope of X, $X[\mathfrak{p}^{+\infty}]$ has exactly two slopes, say $\lambda = a/b$ and $1-a/b$, where $\gcd(a, b) = 1$. Let m be the multiplicity of λ as a slope of $X[\mathfrak{p}^{+\infty}]$; then $mb = [K^+_{\mathfrak{p}^+} : \mathbb{Q}_p]$. The endomorphism algebra of $X[\mathfrak{p}^{+\infty}]$ (again, in the category of p-divisible groups up to isogeny) is isomorphic to

$$\mathrm{End}(X[\mathfrak{p}^{+\infty}])^0 \cong \mathrm{Mat}_m(D_\lambda) \oplus \mathrm{Mat}_m(D_{1-\lambda}),$$

where D_λ is the central simple $\mathbb{Q}_p$-algebra with Brauer invariant λ. In particular, any subfield L of $\mathrm{End}(X[\mathfrak{p}^{+\infty}])^0$ satisfies $[L : \mathbb{Q}_p] \le mb = [K^+_{\mathfrak{p}^+} : \mathbb{Q}_p]$. Since $K \otimes_{K^+} K^+_{\mathfrak{p}^+}$ acts on $X[\mathfrak{p}^{+\infty}]$, we must have $K \otimes_{K^+} K^+_{\mathfrak{p}^+} \cong K^+_{\mathfrak{p}^+} \oplus K^+_{\mathfrak{p}^+}$, as claimed. □

Corollary 3.6.

$$T^{\mathrm{der}}_{\mathbb{Q}_p} \cong \bigoplus_{\mathfrak{p}^+} \boldsymbol{R}_{K^+_{\mathfrak{p}^+}/\mathbb{Q}_p}\mathbb{G}_m.$$

Proof. Since $T^{\mathrm{der}} = \boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m$, using Lemma 3.5 we find

$$T^{\mathrm{der}}_{\mathbb{Q}_p} = \boldsymbol{R}_{K^+\otimes\mathbb{Q}_p/\mathbb{Q}_p}\boldsymbol{R}^{(1)}_{K\otimes\mathbb{Q}_p/K^+\otimes\mathbb{Q}_p}\mathbb{G}_m \cong \bigoplus_{\mathfrak{p}^+|p} \boldsymbol{R}_{K^+_{\mathfrak{p}^+}/\mathbb{Q}_p}\boldsymbol{R}^{(1)}_{K\otimes_{K^+}K^+_{\mathfrak{p}^+}/K^+_{\mathfrak{p}^+}}\mathbb{G}_m.$$

If L is any field then $\boldsymbol{R}_{L\oplus L/L}\mathbb{G}_m \cong \mathbb{G}_{m,L} \oplus \mathbb{G}_{m,L}$; the norm map $\boldsymbol{R}_{L\oplus L/L}\mathbb{G}_m \to \mathbb{G}_m$ is given by multiplication of components; and so $\boldsymbol{R}^{(1)}_{L\oplus L/L}\mathbb{G}_m$ is isomorphic to $\mathbb{G}_{m,L}$, where the latter is embedded in the former via (id, inv). □

Recall that the twisted conjugacy class of $\mathrm{N}(\delta_0)$ is defined over $\mathbb{Q}_p$, and consequently contains a $G(\mathbb{Q}_p)$-rational representative [Kottwitz 1982, p. 799, Theorem 4.4]; and we have adjusted δ_0 in its twisted conjugacy class to ensure that $\gamma_0 := \mathrm{N}(\delta_0) \in G(\mathbb{Q}_p)$ (2-2).

Lemma 3.7. *The stable conjugacy class of γ_0 consists of a single conjugacy class, and the stable σ-conjugacy class of δ_0 consists of a single σ-conjugacy class.*

Proof. To prove the first claim, it suffices (by [Kottwitz 1982, p. 788]) to show that $H^1(\mathbb{Q}_p, T)$ vanishes. By taking the long exact sequence of cohomology of the top row of (2-3), and then invoking Hilbert 90 and Corollary 3.6, we find that $H^1(\mathbb{Q}_p, T)$ does in fact vanish.

For the second claim, it similarly suffices to show that the first cohomology of the twisted centralizer $G_{\delta_0,\sigma}$ vanishes [Kottwitz 1982, p. 805]. However, the twisted centralizer of an element is always an inner form of the (usual) centralizer of its norm [Kottwitz 1982, Lemma 5.8]. In our case, the centralizer $T = G_{\gamma_0}$ is a torus, and thus admits no nontrivial inner forms. We conclude again that $H^1(\mathbb{Q}_p, G_{\delta_0,\sigma})$ is trivial. □

4. Ratios

4A. *Definitions.* For $\ell \neq p$, we define a local ratio $\nu_\ell([X, \lambda])$ designed to measure the extent to which the conjugacy class of $\gamma_{X_0/\mathbb{F}_q}$, as an element of $G(\mathbb{Z}_\ell/\ell)$, is more or less prevalent than a randomly chosen conjugacy class. More precisely, to measure this probability, we consider the finite group $G(\mathbb{Z}_\ell/\ell^n)$ for sufficiently large n, and recall that our notion of "conjugacy" in this group is not the usual conjugacy but the relation $\sim_{M(\mathbb{Z}_\ell/\ell^n)_e}$ defined above in Section 2D. For $\ell = p$, the element $\gamma_{X_0/\mathbb{F}_q}$ is not in $G(\mathbb{Z}_p)$, and we use $M(\mathbb{Z}_p)$ instead; but this has no effect on the definition since our notion of "conjugacy" in $G(\mathbb{Z}_p/p^n)$ already uses $M(\mathbb{Z}_p)$.

Recall the definition of $C_{(d,n)}(\gamma_0)$ from (3-1), and that $C_n(\gamma_0) := C_{(0,n)}(\gamma_0)$ is the actual conjugacy class of $\pi_n(\gamma_0)$ in $G(\mathbb{Z}_\ell/\ell^n)$.

Definition 4.1. For each finite place ℓ, including $\ell = p$, using the shorthand $\gamma_0 := \gamma_{X/\mathbb{F}_q,\ell} \in M(\mathbb{Z}_\ell)$, set

$$\nu_\ell([X, \lambda]) = \nu_\ell([X, \lambda], \mathbb{F}_q) = \lim_{d\to\infty} \lim_{n\to\infty} \frac{\#C_{(d,n)}(\gamma_0)}{\#G(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)}. \tag{4-1}$$

At infinity, define

$$\nu_\infty([X, \lambda]) = \nu_\infty([X, \lambda], \mathbb{F}_q) = \frac{|D(\gamma_0)|_\infty^{\frac{1}{2}}}{(2\pi)^g}, \tag{4-2}$$

where $|\cdot|_\infty$ is the usual real absolute value.

Remark 4.2. So far we have avoided using the fact that there exists a rational element $\gamma_0 \in G(\mathbb{Q})$ as in Remark 2.2, and treated γ_0 as an element of $G(\mathbb{A}_f)$. We can continue doing so, and then for (4-2) simply *define* the archimedean absolute value of its discriminant by $|D(\gamma_0)|_\infty := \prod_\ell |D(\gamma_{X/\mathbb{F}_q,\ell})|_\ell^{-1}$.

The ratios stabilize for large enough d and n, and thus the limits are not, strictly speaking, necessary. In fact, for $\ell \neq 2$, p and not dividing the discriminant of γ_0, the ratios stabilize right away, at $d = 0$ and $n = 1$, as the next two lemmas show.

Lemma 4.3. *If $\ell \nmid D(\gamma_0)$ and $n \gg 0$, then $C_{(d,n)}(\gamma_0) = C_n(\gamma_0)$.*

Proof. Recall that all our notation assumes that ℓ is fixed. Clearly $C_n(\gamma_0) \subseteq C_{(d,n)}(\gamma_0)$. Conversely, suppose n is sufficiently large as in Lemma 3.2 and that $\gamma \in C_{(d,n)}(\gamma_0)$. Then there exists some $\widetilde{\gamma} \in \pi_n^{-1}(\gamma)$ such that $\widetilde{\gamma} \sim_{G(\mathbb{Q}_\ell)} \gamma_0$. By Lemma 3.1, $\widetilde{\gamma} \sim_{G(\mathbb{Z}_\ell)} \gamma_0$, and so $\gamma \sim_{G(\mathbb{Z}_\ell/\ell^n)} \gamma_0$. □

Lemma 4.4. *If $\ell \nmid D(\gamma_0)$, $\ell \neq 2$, and $d \gg_{\gamma_0} 0$, then for $n \geq 1$,*

$$\frac{\#C_{(d,n)}(\gamma_0)}{\#G(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)} = \frac{\#\{\gamma \in G(\mathbb{Z}_\ell/\ell) : \gamma \sim \gamma_0\}}{\#G(\mathbb{Z}_\ell/\ell)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell)}. \tag{4-3}$$

Proof. By Lemma 4.3, the left-hand side of (4-3) is

$$\frac{\#G(\mathbb{Z}_\ell/\ell^n)/\#G_{\gamma_0}(\mathbb{Z}_\ell/\ell^n)}{\#G(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)} = \frac{\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)}{\#G_{\gamma_0}(\mathbb{Z}_\ell/\ell^n)}.$$

Since $\pi_1(\gamma_0)$ is regular, the centralizer G_{γ_0} is smooth over $\mathbb{Z}_\ell$ of relative dimension $g+1$. Since the same is true of the scheme $\mathbb{A}_G/\mathbb{Z}_\ell$, the result now follows. □

4B. ***From ratios to integrals.*** Fix a prime ℓ (possibly $\ell = p$ or $\ell = 2$). (In this subsection, as above, all quantities depend on this notationally suppressed prime.) Recall (2-4) the canonical map $\mathfrak{c} : G \to A$ from G to its Steinberg quotient. The fibres of this map over regular points are stable orbits of regular semisimple elements. Define a system of neighbourhoods of $\mathfrak{c}(\gamma_0)$ inside $\mathbb{A}_G(\mathbb{Z}_\ell)$ by

$$\widetilde{U}_n(\gamma_0) = \widetilde{\pi_n^{\mathbb{A}_G}(\mathfrak{c}(\gamma_0))} = (\pi_n^{\mathbb{A}_G})^{-1}\big(\pi_n^{\mathbb{A}_G}(\mathfrak{c}(\gamma_0))\big).$$

In other words,

$$\widetilde{U}_n(\gamma_0) = \{a = (a_1, \ldots, a_g, \eta) \in \mathbb{A}_G(\mathbb{Z}_\ell) \mid a_i \equiv a_i(\gamma_0) \bmod \ell^n,\ \eta \equiv \eta(\gamma_0) \bmod \ell^n\}.$$

These definitions and (3-1) are summarized by the diagram

$$\begin{tikzcd} \widetilde{C}_{(d,n)}(\gamma_0) \subset M(\mathbb{Z}_\ell) \arrow[r, "\mathfrak{c}"] \arrow[d, "\pi_n^M"] & \mathbb{A}_G(\mathbb{Z}_\ell) \supset \widetilde{U}_n(\gamma_0) \arrow[d, "\pi_n^{\mathbb{A}_G}"] \\ C_{(d,n)}(\gamma_0) \subset M(\mathbb{Z}_\ell/\ell^n) \arrow[r, "\mathfrak{c}_n"] & \mathbb{A}_G(\mathbb{Z}/\ell^n) \ni \pi_n^{\mathbb{A}_G}(\mathfrak{c}(\gamma_0)) \end{tikzcd} \tag{4-4}$$

where $\mathfrak{c}_n : G(\mathbb{Z}/\ell^n) \to \mathbb{A}_G(\mathbb{Z}/\ell^n)$ is the map sending an element to the coefficients of its characteristic polynomial mod ℓ^n. The diagram of maps commutes since reduction mod ℓ^n is a ring homomorphism, and the map $\mathfrak{c}$ is polynomial in the matrix entries of γ. (The diagram of subsets need not commute, though.) We also note that when $\ell \neq p$, the sets $\widetilde{C}_{(d,n)}(\gamma_0)$ and $C_{(d,n)}(\gamma_0)$ are contained in $G(\mathbb{Z}_\ell)$ and $G(\mathbb{Z}_\ell/\ell^n)$, respectively, since $\mathrm{ord}_\ell(\det(\gamma_0)) = 0$ in this case, and this is also true for all elements that are congruent to any conjugate of γ_0.

By definition of the geometric measure, for any open subset $B \subset G(\mathbb{Z}_\ell)$ we have

$$\mathrm{vol}_{\mu^{\mathrm{geom}}}\big(B \cap \mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0))\big) = \lim_{n\to\infty} \frac{\mathrm{vol}_{|d\omega_G|}\big(\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0)) \cap B\big)}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}. \tag{4-5}$$

Recall that each stable orbit $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma))$ is a finite disjoint union of rational orbits. Each rational orbit being an open subset of the stable orbit, we may and do define geometric measure on each rational orbit, by restriction.

In simple terms, the sets $\widetilde{U}_n$ form a system of neighbourhoods of the point $\mathfrak{c}(\gamma_0) \in \mathbb{A}_G$; the set $\widetilde{C}_{(d,n)}(\gamma_0)$ can be thought of as the intersection of a neighbourhood of the orbit of γ_0 with $G(\mathbb{Z}_\ell)$; the set $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0))$ is the stable orbit of γ_0. The following lemma gives the precise relationships between all these sets.

Lemma 4.5. (a) *Let $\ell \neq p$. For large enough d and n (depending on γ_0), we have*

$$\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0)) \cap G(\mathbb{Z}_\ell) = \bigcup_{\gamma' \sim_{G(\overline{\mathbb{Q}}_\ell)} \gamma_0} \widetilde{C}_{d,n}(\gamma'), \tag{4-6}$$

where γ' runs over a set of representatives of $G(\mathbb{Q}_\ell)$-conjugacy classes in the stable conjugacy class of γ_0 whose $\mathbb{Q}_\ell$-orbits intersect $G(\mathbb{Z}_\ell)$, so that we may take the elements γ' to lie in $G(\mathbb{Z}_\ell)$.

(b) *When n is sufficiently large (depending on γ_0), the sets $\widetilde{C}_{d,n}(\gamma')$ above are disjoint.*

(c) *Let μ_G^{SO} be the Serre–Oesterlé measure on $G(\mathbb{Q}_p) \cap M(\mathbb{Z}_p)$, viewed as a submanifold of $M(\mathbb{Z}_p)$. Then $\mathrm{vol}_{\mu_G^{\mathrm{SO}}}(\widetilde{C}_{d,n}(\gamma_0)) = \ell^{-n \dim(G)} \# C_{d,n}(\gamma_0)$; in particular, $\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{d,n}(\gamma_0)) = \ell^{-n \dim(G)} \# C_{d,n}(\gamma_0)$ if $\ell \neq p$.*

Proof. (a) This is an easy consequence of the fact that two regular semisimple elements of $G(\mathbb{Q}_\ell)$ are stably conjugate if and only if their characteristic polynomials coincide. In our notation,

$$\mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0)) = \bigsqcup_{\gamma' \sim_{G(\overline{\mathbb{Q}}_\ell)} \gamma_0} ({}^{G(\mathbb{Q}_\ell)}\gamma'),$$

where ${}^{G(\mathbb{Q}_\ell)}\gamma'$ denotes the rational conjugacy class of γ' in $G(\mathbb{Q}_\ell)$ as before. Now, we will describe both the left-hand side and the right-hand side of (4-6) as the set of elements $\gamma \in G(\mathbb{Z}_\ell)$ whose characteristic polynomial is congruent to that of $\gamma_0 \bmod \ell^n$. Indeed, on the left-hand side, by definition, $\gamma \in \mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0)) \cap G(\mathbb{Z}_\ell)$ if and only if $\pi_n^{\mathbb{A}_G}(\mathfrak{c}(\gamma)) = \pi_n^{\mathbb{A}_G}(\mathfrak{c}(\gamma_0))$. By the commutativity of (4-4), this is equivalent to $\mathfrak{c}_n(\pi_n^G(\gamma)) = \mathfrak{c}_n(\pi_n^G(\gamma_0))$, i.e., the characteristic polynomials of γ and γ_0 are congruent mod ℓ^n. On the right-hand side, given $\gamma' \in G(\mathbb{Z}_\ell)$, by Lemma 3.2, for d and n large enough,[1] we have that $\gamma \in \widetilde{C}_{(d,n)}(\gamma')$ if and only if there exists $\gamma'' \in G(\mathbb{Z}_\ell)$ such that $\gamma'' \equiv \gamma' \bmod \ell^n$ and γ'' is $G(\mathbb{Q}_\ell)$-conjugate to γ. Taking the union of these sets as γ' runs over the set of integral representatives of $G(\mathbb{Q}_\ell)$-conjugacy classes in the stable class of γ_0, we obtain the set of all elements $\gamma \in G(\mathbb{Z}_\ell)$ that are congruent modulo ℓ^n to an element of $G(\mathbb{Z}_\ell)$ that is stably conjugate to γ_0, i.e., to an element having the same characteristic polynomial as γ_0. This means that $\mathfrak{c}_n(\pi_n^G(\gamma)) = \mathfrak{c}_n(\pi_n^G(\gamma_0))$, completing the proof of the first statement.

(b) Since the orbits of regular semisimple elements are closed in the ℓ-adic topology, distinct orbits have disjoint neighbourhoods.

[1] *Large enough* depends on γ', but only through its discriminant. Since stably conjugate elements have the same discriminant, ultimately this only depends on γ_0.

(c) The map $\pi_n^M : \widetilde{C}_{(d,n)}(\gamma_0) \to C_{(d,n)}(\gamma_0)$ is surjective, so $\widetilde{C}_{(d,n)}(\gamma_0)$ can be thought of as a disjoint union of fibres of π_n^M. Since M is a smooth scheme over $\mathbb{Z}_\ell$, each fibre of π_n^M has volume $\ell^{-n\dim(G)}$ with respect to the measure μ^{SO} (see [Serre 1981]). The first statement follows. Moreover, as discussed above in Section 2E, on $G(\mathbb{Z}_\ell)$, the measures μ^{SO} and $\mu_{|\omega_G|}$ coincide. For $\ell \neq p$, we have $\widetilde{C}_{(d,n)}(\gamma_0) \subset G(\mathbb{Z}_\ell)$, which completes the proof. □

Recall that ϕ_0 is the characteristic function of $G(\mathbb{Z}_\ell)$.

Corollary 4.6. *Let $\ell \neq p$. Then there exists $d(\gamma_0)$ such that for $d \geq d(\gamma_0)$*

$$O_{\gamma_0}^{\mathrm{geom}}(\phi_0) = \lim_{n\to\infty} \frac{\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{(d,n)}(\gamma_0))}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}.$$

Proof. The orbital integral, by definition, calculates the volume of the set of integral points in the rational orbit of γ_0, with respect to the geometric measure on the orbit. Using Lemma 4.5(a)–(b), we write $\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0)) \cap G(\mathbb{Z}_\ell) = \bigsqcup_{\gamma'} \widetilde{C}_{d,n}(\gamma')$, where γ' are as in that lemma, with γ_0 being one of the elements γ'. The union on the right-hand side of (4-6) is a disjoint union of neighbourhoods of the individual orbits, intersected with $G(\mathbb{Z}_\ell)$. The statement follows from the equality (4-5), applied to the set $B := \widetilde{C}_{d,n}(\gamma_0)$. □

Corollary 4.7. *For $\ell \neq p$, the Gekeler ratio (4-1) is related to the geometric orbital integral by*

$$\nu_\ell([X,\lambda]) = \frac{\ell^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell)} O_{\gamma_0}^{\mathrm{geom}}(\phi_0).$$

Proof. Note that at a finite level n (and for d large enough so that the equalities in all the previous lemmas hold), the denominator in (4-1) is

$$\frac{\#G(\mathbb{Z}_\ell/\ell^n)}{\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)} = \frac{\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell^n)\#\mathbb{G}_m(\mathbb{Z}/\ell^n)}{\ell^{(\mathrm{rank}(G)-1)n}\#\mathbb{G}_m(\mathbb{Z}/\ell^n)} = \frac{\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell^n)}{\ell^{(\mathrm{rank}(G)-1)n}} = \frac{\ell^{(\dim(G)-1)(n-1)}\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell)}{\ell^{(\mathrm{rank}(G)-1)n}}.$$

By Lemma 4.5(c), we have $\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{d,n}(\gamma_0)) = \#C_{d,n}(\gamma_0)/\ell^{n\dim(G)}$, and by definition of the measure on the Steinberg quotient, $\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0)) = \ell^{-n\,\mathrm{rank}(G)}$ (here we are using the fact that $|\eta(\gamma_{X,\ell})| = 1$ for $\ell \neq p$, so the absolute value of the $\mathbb{G}_m$-coordinate is 1 on $\widetilde{U}_n$).

Then for a given level n, we have

$$\begin{aligned}\frac{\#C_{(d,n)}(\gamma_0)}{\#G(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)} &= \frac{\ell^{n\dim(G)}\,\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{(d,n)}(\gamma_0))\ell^{(\mathrm{rank}(G)-1)n}}{\ell^{(\dim(G)-1)(n-1)}\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell)} \\ &= \frac{\ell^{\dim(G)-1}}{\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell)} \frac{\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{(d,n)}(\gamma_0))}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}.\end{aligned}$$

The result now follows from Corollary 4.6. □

4C. ***Calculation at p.*** Recall that we have fixed a maximal split torus $T_{\mathrm{spl}} \subset G$. For any cocharacter $\lambda \in X_*(T_{\mathrm{spl}})$ (and any power $q = p^e$ of p), let $\psi_\lambda = \psi_{\lambda,q}$ be the characteristic function of the double coset

$$D_{\lambda,q} = G(\mathbb{Z}_q)\lambda(p)G(\mathbb{Z}_q).$$

By the Cartan decomposition, the collection of all ψ_λ is a basis for $\mathcal{H}_G = \mathcal{H}_{G,\mathbb{Q}_q}$, the Hecke algebra of functions on $G(\mathbb{Q}_q)$ which are bi-$G(\mathbb{Z}_q)$-invariant.

Let μ_0 be the cocharacter $p \mapsto \mathrm{diag}(p, \dots, p, 1, \dots, 1)$; it is the cocharacter associated to the Shimura variety $\mathcal{A}_g$. Define

$$\psi_{q,p} = \psi_{\mu_0,q} = \mathbb{1}_{G(\mathbb{Z}_q)\,\mathrm{diag}(p,\dots,p,1,\dots,1)G(\mathbb{Z}_q)} \quad \text{and} \quad \phi_{q,p} = \psi_{e\mu_0,p} = \mathbb{1}_{G(\mathbb{Z}_p)\,\mathrm{diag}(q,\dots,q,1,\dots,1)G(\mathbb{Z}_p)}.$$

Recall that $\delta_0 = \delta_{X/\mathbb{F}_q}$ represents the absolute Frobenius of X, and that $\gamma_0 := \mathrm{N}\,\delta_0$ lies in $G(\mathbb{Q}_p) \cap M(\mathbb{Z}_p)$ (2-2).

Lemma 4.8. *Let $[X, \lambda]/\mathbb{F}_q$ be a principally polarized abelian variety. Suppose that either X is ordinary or that $q = p$ (and thus $e = 1$). Then*

$$^{G(\mathbb{Q}_p)}\gamma_0 \cap M(\mathbb{Z}_p) \subseteq D_{e\mu_0,p}.$$

Proof. We identify $X_*(T_{\mathrm{spl}})$ with

$$\{\underline{a} = (a_1, \dots, a_{2g}) \in \mathbb{Z}^{2g} : a_i + a_{g+i} = a_j + a_{g+j} \text{ for } 1 \le i, j \le g\}. \tag{4-7}$$

Suppose $\gamma \in D_{\underline{a},p} \subseteq G(\mathbb{Q}_p)$. Then $\mathrm{ord}_p(\eta(\gamma))$ is the common value of $a_i + a_{g+i}$; and γ stabilizes $V \otimes \mathbb{Z}_p$ — that is, $\gamma \in M(\mathbb{Z}_p)$ — if and only if each a_i satisfies $a_i \ge 0$.

Let $f(\underline{a}) = \#\{i : a_i = 0\}$. If $\alpha \in D_{\underline{a}} \cap M(\mathbb{Z}_p)$, then $f(\underline{a})$ is the rank of $\pi_1(\alpha)$ as an endomorphism of V/pV.

With these preparations, suppose $\gamma \in {}^{G(\mathbb{Q}_p)}\gamma_0 \cap M(\mathbb{Z}_p)$. Note that we have $a_i + a_{g+i} = e$.

First, suppose X is ordinary. Then exactly g eigenvalues of γ are p-adic units. Consequently, if $\gamma \in D_{\underline{a}}$, then $f(\underline{a}) = g$. The only $\underline{a}$ as in (4-7) compatible with the symmetry and integrality constraints is $(e, \dots, e, 0, \dots, 0)$.

Second, suppose X has arbitrary Newton polygon but that $e = 1$. Again, the only $\underline{a}$ such that $a_i + a_{g+i} = e = 1$ and each a_i is nonnegative is $(1, \dots, 1, 0, \dots, 0)$. $\square$

Lemma 4.9. *Suppose that $[X, \lambda]/\mathbb{F}_q$ is an ordinary, simple, principally polarized abelian variety. Then*

$$TO_{\delta_0}(\psi_{q,p}) = O_{\gamma_0}(\phi_{q,p}).$$

Proof. There is a base change map $b = b_{G,\mathbb{Q}_q/\mathbb{Q}_p} : \mathcal{H}_{G,\mathbb{Q}_q} \to \mathcal{H}_{G,\mathbb{Q}_p}$. The fundamental lemma asserts that if $\psi \in \mathcal{H}_{G,\mathbb{Q}_q}$, then stable twisted orbital integrals for ψ match with stable orbital integrals for $b\psi$. For our δ_0 and γ_0, the adjective *stable* is redundant (Lemma 3.7), the case of the fundamental lemma we need is [Clozel 1990, Theorem 1.1], and we have

$$TO_{\delta_0}(\psi_{q,p}) = O_{\gamma_0}(b\psi_{q,p}). \tag{4-8}$$

While we will stop short of computing $b\psi_{q,p}$, we will find a function that agrees with it on the orbit $^{G(\mathbb{Q}_p)}\gamma_0$.

The Satake transformation is an algebra homomorphism $\mathfrak{S} : \mathcal{H}_{G,\mathbb{Q}_q} \to \mathcal{H}_{T_{\mathrm{spl}},\mathbb{Q}_q}$ which maps $\mathcal{H}_{G,\mathbb{Q}_q}$ isomorphically onto the subring $\mathcal{H}^W_{T_{\mathrm{spl}},\mathbb{Q}_q}$ of invariants under the Weyl group. It is compatible with base change, in the sense that there is a commutative diagram

$$\begin{tikzcd} \mathcal{H}_{G,\mathbb{Q}_q} \arrow[r, "\mathfrak{S}"] \arrow[d, "b"] & \mathcal{H}_{T_{\mathrm{spl}},\mathbb{Q}_q} \cong \mathbb{C}[X_*(T_{\mathrm{spl}})] \arrow[d, "b"] \\ \mathcal{H}_{G,\mathbb{Q}_p} \arrow[r, "\mathfrak{S}"] & \mathcal{H}_{T_{\mathrm{spl}},\mathbb{Q}_p} \end{tikzcd}$$

We exploit the following data about the Satake transform and the base change map.

Under the canonical identification of $X_*(T_{\mathrm{spl}})$ and $X^*(\widehat{T}_{\mathrm{spl}})$, the character group of the dual torus, $\lambda \in X_*(T_{\mathrm{spl}})$ gives rise to a character of $\widehat{T}_{\mathrm{spl}}$, and thus a representation V_λ of $\widehat{G}$; let χ_λ be its trace. Then

$$\mathfrak{S}(\psi_{\mu,q}) = q^{\langle\mu,\rho\rangle}\chi_\mu + \sum_{\lambda<\mu} a(\mu,\lambda)\chi_\lambda$$

for certain numbers $a(\mu,\lambda)$, where as usual ρ is the half-sum of positive roots [Gross 1998, (3.9)].

On one hand, following Gross [1998, (3.15)] and Kottwitz [1984a, (2.1.3)], we observe that the weight $\mu_0 = (1, \dots, 1, 0, \dots, 0)$ is minuscule, and therefore

$$\mathfrak{S}(\psi_{\mu_0,q}) = q^{\langle\mu_0,\rho\rangle}\chi_{\mu_0}.$$

If we think of elements of $\mathbb{C}[X_*(T_{\mathrm{spl}})]^W$ as polynomials in $2g$ variables $z_1, \dots, z_{2g}$, then (essentially by definition of the highest weight and the fact that the multiplicity of the highest weight in an irreducible representation is 1 — in our case the representation in question is in fact the oscillator representation [Gross 1998, (3.15)]), we find that the leading term of $\mathfrak{S}(\psi_{\mu_0,q})$ is $q^{\langle\mu_0,\rho\rangle}z_1\cdots z_g$. By definition, the base change map takes $f \in \mathbb{C}[z_1, \dots, z_{2g}, z_1^{-1}, \dots, z_{2g}^{-1}]^W$ to $f(z_1^e, \dots, z_{2g}^e)$. Then

$$b(\mathfrak{S}(\psi_{q,p})) = q^{\langle\mu_0,\rho\rangle}z_1^e\cdots z_g^e + \sum_{\lambda<e\mu_0} a(e\mu_0,\lambda)\chi_\lambda.$$

On the other hand, we have

$$\mathfrak{S}(\phi_{q,p}) = p^{\langle e\mu_0,\rho\rangle}\chi_{e\mu_0} + \sum_{\lambda<e\mu_0} b(e\mu_0,\lambda)\chi_\lambda = q^{\langle\mu_0,\rho\rangle}z_1^e\cdots z_g^e + \sum_{\lambda<e\mu_0} c(e\mu_0,\lambda)\chi_\lambda.$$

In these formulas, $a(e\mu_0,\lambda)$, $b(e\mu_0,\lambda)$ and $c(e\mu_0,\lambda)$ are coefficients of lower-weight monomials that are ultimately irrelevant to our calculation. In particular,

$$\phi_{q,p} - \mathfrak{S}^{-1}\big(b(\mathfrak{S}(\psi_{q,p}))\big)$$

vanishes on $D_{e\mu_0,p} = G(\mathbb{Z}_p)e\mu_0(p)G(\mathbb{Z}_p)$.

The last point to note is that the intersection of the support of this difference $\phi_{q,p} - \mathfrak{S}^{-1}\big(b(\mathfrak{S}(\psi_{q,p}))\big)$ with the orbit of γ_0 is contained in $M(\mathbb{Z}_p)$. Once we have shown this, the desired result follows from the fundamental lemma (4-8) combined with Lemma 4.8. We start by observing that since the multiplier is a multiplicative map, it is constant on double $G(\mathbb{Z}_p)$-cosets. Therefore, for any double coset $D_{\underline{a},p}$ such that $D_{\underline{a},p} \cap {}^{G(\mathbb{Q}_p)}\gamma_0 \neq \varnothing$, we have $a_i + a_{g+i} = e$. Now suppose $\lambda \leftrightarrow \underline{a}$ is a dominant weight satisfying this condition and further satisfying $\lambda \leq e\mu_0$. Then we have $a_1 \geq a_2 \geq \cdots \geq a_g$ and $a_g \geq 0$ because λ

is dominant; and on the other hand, $e-a_1 \geq e-a_2 \geq \cdots \geq e-a_g$, and $e-a_g \geq 0$ because of the condition $\lambda \leq e\mu_0$. Therefore, in particular, $a_{g+1}, \ldots, a_{2g}$ are nonnegative, and thus $D_{\lambda,p} \subset M(\mathbb{Z}_p)$ (and in fact, we have also shown that $a_1 = \cdots = a_g$). □

Lemma 4.10. *Suppose that either X is ordinary or that $q = p$. Then there exists $d(\gamma_0)$ such that*

$$O_{\gamma_0}^{\mathrm{geom}}(\phi_{q,p}) = \lim_{n\to\infty} \frac{\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{d(\gamma_0),n}(\gamma_0))}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}.$$

Proof. Suppose that X is ordinary (but q is an arbitrary power of p). By Lemma 3.7, $\mathfrak{c}^{-1}(\mathfrak{c}(\gamma_0))$ is a single $G(\mathbb{Q}_p)$-conjugacy class; the same argument shows this is true for elements in a small neighbourhood of $\mathfrak{c}(\gamma_0)$. So, using (4-5), $O_{\gamma_0}^{\mathrm{geom}}(\phi_{q,p})$ equals $\lim_{n\to\infty}\big(\mathrm{vol}_{|d\omega_G|}\big(\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0))\cap D_{e\mu_0,p}\big)/\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))\big)$. By Lemma 4.8, we have

$$\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0))\cap D_{e\mu_0,p} = \mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0))\cap M(\mathbb{Z}_p).$$

Therefore, all we need to show is that for large enough d and n, we have

$$\mathfrak{c}^{-1}(\widetilde{U}_n(\gamma_0))\cap M(\mathbb{Z}_p) = \widetilde{C}_{d,n}(\gamma_0); \tag{4-9}$$

but this is essentially Corollary 3.4(b).

The case where $q = p$ follows from Corollary 4.6 and the second case of Lemma 4.8. □

Lemma 4.11. *On the double coset $D_{e\mu_0,p}$ we have*

$$|d\omega_G| = q^{\frac{1}{2}g(g+1)+1}\mu^{\mathrm{SO}}. \tag{4-10}$$

Proof. Let $K = G(\mathbb{Z}_p)$. First, observe that the measure μ^{SO} on $G(\mathbb{Q}_p)\cap M(\mathbb{Z}_p)$ is both left- and right-K-invariant (since multiplication by an element of $G(\mathbb{Z}_p)$ yields a bijection on mod p^n-points). Consider the decomposition of $D_{e\mu_0,p}$ into, say, left K-cosets: $D_{e\mu_0,p} = \bigsqcup_{i=1}^{s} g_iK$ (the number s of these cosets was computed by Iwahori and Matsumoto but is not needed here). It follows from left K-invariance of μ^{SO} that $\mu^{\mathrm{SO}}(g_iK)$ is the same for all i.

Second, the measure $|d\omega_G|$ is normalized so that each K-coset has volume $\#G(\mathbb{F}_p)$. Thus, in order to compare the measures μ^{SO} and $|d\omega_G|$, we need to compare the cardinality $\#\pi_n(g_iK)$ of the reduction mod p^n of any such coset g_iK that is contained in $D_{e\mu_0,p}$ with $\#G(\mathbb{F}_p)$, for sufficiently large n. (Note that $n = 1$ is insufficient, because for all such cosets the reduction mod p of any matrix in gK would be of lower rank. One needs to go to $n > e$ for the ratios $\#\pi_n(gK)/p^{n\dim(G)}$ to stabilize.) Since the answer does not depend on g_i, we can take $g_0 = e\mu_0(p) = \mathrm{diag}(q,\ldots,q,1,\ldots,1)$. In other words, we need to compute the cardinality of the fibre of the map

$$\varphi_q : G(\mathbb{Z}/p^n) \to M(\mathbb{Z}/p^n), \qquad \begin{bmatrix} A & B \\ C & D \end{bmatrix} \mapsto \begin{bmatrix} qA & qB \\ C & D \end{bmatrix}.$$

For simplicity, we would like to move the calculation to the Lie algebra. Let $n \gg e$. Observe that if $\varphi_q(\gamma_1) = \varphi_q(\gamma_2)$ for $\gamma_1, \gamma_2 \in G(\mathbb{Z}/p^n)$, then $\left[\begin{smallmatrix} qI_g & 0 \\ 0 & I_g \end{smallmatrix}\right](\gamma_1\gamma_2^{-1} - I) = 0$, where I_g is the $g\times g$-identity matrix, and I is the identity matrix in M_{2g}. This implies, in particular, that $\gamma_1\gamma_2^{-1} \equiv I \mod p^{n-e}$. Then we can write the truncated exponential approximation: $\gamma_1\gamma_2^{-1} = I + X + \frac{1}{2}X^2 + \cdots$ for some $X \in \mathfrak{g}(\mathbb{Z}_p)$; in particular, there exists $X \in \mathfrak{g}(\mathbb{Z}_p)$ such that $\gamma_1\gamma_2^{-1} \equiv I + X \mod p^{2(n-e)}$, and thus the kernel of the map φ_q is in bijection with the set of ($X \mod p^n$) for $X \in \mathfrak{g}(\mathbb{Z}_p)$ such that $\left[\begin{smallmatrix} qI_g & 0 \\ 0 & I_g \end{smallmatrix}\right] X \equiv 0 \mod p^n$.

We have $\mathfrak{g} = \mathfrak{sp}_{2g} \oplus \mathfrak{z}$, where $\mathfrak{z}$ is the 1-dimensional Lie algebra of the centre. It will be convenient to decompose it further: let $\mathfrak{h}$ be the Cartan subalgebra of $\mathfrak{sp}_{2g}$ consisting of diagonal matrices, and let V consist of matrices whose diagonal entries are all zero; then

$$\mathfrak{g} = (\mathfrak{z} \oplus \mathfrak{h}) \oplus V.$$

Consider the action of multiplication by $\left[\begin{smallmatrix} qI_g & 0 \\ 0 & I_g \end{smallmatrix}\right]$ on each term of this direct sum decomposition.

On the term $\mathfrak{z} \oplus \mathfrak{h}$ it acts by $\operatorname{diag}(a_1, \ldots, a_{2g}) \mapsto \operatorname{diag}(qa_1, \ldots, qa_g, a_{g+1}, \ldots, a_{2g})$, which in the $\mathfrak{z} \oplus \mathfrak{h}$-coordinates can be written as (recalling that $a_i + a_{g+i} = z$ is independent of i):

$$\begin{aligned} &\frac{z}{2} \oplus \left(\frac{z}{2} - a_{g+1}, \ldots, \frac{z}{2} - a_{2g}, -\frac{z}{2} + a_{g+1}, \ldots, -\frac{z}{2} + a_{2g}\right) \\ &\quad\mapsto \frac{qz}{2} \oplus \left(\frac{qz}{2} - \frac{(q+1)a_{g+1}}{2}, \ldots, \frac{qz}{2} - \frac{(q+1)a_{2g}}{2}, -\frac{qz}{2} + \frac{(q+1)a_{g+1}}{2}, \ldots, -\frac{qz}{2} + \frac{(q+1)a_{2g}}{2}\right). \end{aligned}$$

The only points $(z, a_{g+1}, \ldots, a_{2g})$ that are killed (mod p^n) by this map are of the form $(z', 0, \ldots, 0)$ with $qz' = 0$; so there are q of them.

Next consider an element $X = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] \in V$. Then A is determined by D, and B is skew-symmetric (up to a permutation of rows and columns). Multiplication by $\left[\begin{smallmatrix} qI_g & 0 \\ 0 & I_g \end{smallmatrix}\right]$ scales each entry of A and B by a factor of q, and does not change C and D. Since A is determined by D, the elements X killed by this map are in bijection with symmetric matrices B with entries in $\mathbb{Z}/p^n$ that are killed by multiplication by q. Since the space of such matrices is a $\left(\frac{1}{2}g(g+1)\right)$-dimensional linear space, the number of such matrices B is $q^{\frac{1}{2}g(g+1)}$.

Thus, we have computed that $|d\omega_G| = q^{\frac{1}{2}g(g+1)+1}\mu^{\mathrm{SO}}$ on the double coset $D_{e\mu_0,p}$. Combining this with (4-11), we get

$$\nu_p([X, \lambda]) = \frac{qp^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_p/p)} \frac{q^{-\frac{1}{2}g(g+1)-1} \operatorname{vol}_{|d\omega_G|}(\widetilde{C}_{d,n}(\gamma_0))}{\operatorname{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))} = q^{-\frac{1}{2}g(g+1)} \frac{p^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_p/p)} O_{\gamma_0}^{\mathrm{geom}}(\phi_{q,p}),$$

which completes the proof. □

Corollary 4.12. *Suppose that either X is ordinary or that $q = p$. For $\ell = p$, the Gekeler ratio (4-1) is related to the geometric orbital integral by*

$$\nu_p([X, \lambda]) = q^{-\frac{1}{2}g(g+1)} \frac{p^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_p/p)} O_{\gamma_0}^{\mathrm{geom}}(\phi_{q,p}).$$

Proof. First observe that $\operatorname{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0)) = qp^{-n\operatorname{rank}(G)}$, since we are using the invariant measure on the $\mathbb{G}_m$-factor of $\mathbb{A}_G = \mathbb{A}^{\operatorname{rank}(G)-1} \times \mathbb{G}_m$, and for γ_0 (and therefore, for all points in $\widetilde{U}_n$), that coordinate is the multiplier, with absolute value q^{-1}. Thus, by Lemma 4.5(c) and the same argument as in Corollary 4.7, we have that for $d > d(\gamma_0)$,

$$\nu_p([X,\lambda]) = \lim_{n\to\infty} \frac{\#C_{d,n}(\gamma_0)}{\#G(\mathbb{Z}/p^n\mathbb{Z})/\#\mathbb{A}_G(\mathbb{Z}/p^n\mathbb{Z})} = \frac{qp^{\dim(G)-1}}{\#G^{\mathrm{der}}(\mathbb{Z}/p\mathbb{Z})} \lim_{n\to\infty} \frac{\mathrm{vol}_{\mu^{\mathrm{SO}}}(\widetilde{C}_{d,n}(\gamma_0))}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}. \tag{4-11}$$

The ratio inside the limit on the right-hand side is the same as the ratio in Lemma 4.10, except that the measure in the numerator is the Serre–Oesterlé measure μ^{SO} rather than the measure $|d\omega_G|$. (Both measures are defined on $G(\mathbb{Q}_p)\cap M(\mathbb{Z}_p)$.) Thus, to prove the corollary, we just need to compute the conversion factor between the restrictions of the measures μ^{SO} and $|d\omega_G|$ to the support of $\phi_{q,p}$, which is the content of Lemma 4.11. □

5. The product formula

Now that the relationship between the ratios ν_ℓ and orbital integrals (with respect to the geometric measure) is established, we can translate the formula of Langlands and Kottwitz (2-1) into a Siegel-style product formula for the ratios, thus obtaining our main theorem. Recall the notation of Section 2, in particular, the element $\gamma_{[X,\lambda]} \in G(\mathbb{A}_f)$ associated with the isogeny class of $[X,\lambda]$, and its centralizer $T = T_{[X,\lambda]}$. Here in order to ease the notation we drop all the subscripts $[X,\lambda]$. Note that there is some flexibility in the choice of the measures in the Langlands–Kottwitz formula, but the measures need to be normalized by normalizing the measures on $G(\mathbb{Q}_\ell)$ and on $T(\mathbb{Q}_\ell)$ separately. We will use the canonical measure $d\mu_G^{\mathrm{can}}$ on $G(\mathbb{Q}_\ell)$ for every prime ℓ, and the Tamagawa measure μ_T^{Tama} (defined in detail below) on $T(\mathbb{Q}_\ell)$ for all ℓ. This gives a convergent product measure globally, since the local orbital integrals equal 1 at almost all places with respect to this measure. Since Gekeler-style ratios are expressed in terms of the geometric measure on orbits, we need to calculate the conversion factor between the geometric measure and the quotient $\mu_G^{\mathrm{can}}/\mu_T^{\mathrm{Tama}}$. We start with a quick review of the definition of μ_T^{Tama} in order to introduce all the relevant notation.

5A. *Tamagawa measure.* Let S be an algebraic torus; here we only discuss the setting where S is defined over $\mathbb{Q}$. The character group of S is the free $\mathbb{Z}$-module $X^*(S) = \widehat{S}$ in Ono's notation (we emphasize that this is the lattice of the characters defined over $\overline{\mathbb{Q}}$). If F is any field containing $\mathbb{Q}$, we let $(\widehat{S})_F$ be the subgroup of characters of S which are defined over F.

As usual, we have $S(\mathbb{A})$ and $S(\mathbb{A}_f)$, the points of S with values in, respectively, the ring of adeles and the ring of finite adeles. The (finite) adeles come equipped with the product absolute value $|\cdot|_{\mathbb{A}}$, and we set

$$S(\mathbb{A})^1 = \{s \in S(\mathbb{A}) : |\chi(s)|_{\mathbb{A}} = 1 \text{ for all } \chi \in (\widehat{S})_{\mathbb{Q}}\}.$$

Let F be a Galois extension which splits S. Then the character lattice $X^*(S)$ can be viewed as a $\mathrm{Gal}(F/\mathbb{Q})$-module, and this module uniquely determines S up to isomorphism. We denote this representation by σ_S, and let $L(s,\sigma_S) = \prod_\ell L_\ell(s,\sigma_S)$ be the corresponding Artin L-function (see [Bitan 2011] for a modern treatment). Let r be the multiplicity of the trivial representation in σ_S. By definition,

$$\rho_S := \lim_{s\to 1}(s-1)^r L(s,\sigma_S).$$

Let ω be an invariant gauge form on S. (In particular, ω is defined over $\mathbb{Q}$.) Set

$$\omega^{\text{Tama}} = \omega_\infty \prod_\ell L_\ell(1, \sigma_S)\omega_\ell,$$

where ω_ℓ is the invariant volume form on $S(\mathbb{Q}_\ell)$ induced by ω.

By the product formula, as long as ω is defined over $\mathbb{Q}$, none of the global invariants depend on the normalization of ω.

Let $\chi_1, \dots, \chi_r$ be a basis for $(\widehat{S})_{\mathbb{Q}}$, and define a map Λ by

$$S(\mathbb{A}) \xrightarrow{\Lambda} (\mathbb{R}_+^\times)^r, \quad x \mapsto (|\chi_1(x)|_{\mathbb{A}}, \dots, |\chi_r(x)|_{\mathbb{A}}).$$

(In the cases of interest, when $S = T^{\text{der}}$ or $S = T$, we have $r = 0$ or $r = 1$, respectively.) Then Λ induces an isomorphism

$$\widetilde{\Lambda} : S(\mathbb{A})/S(\mathbb{A})^1 \xrightarrow{\sim} (\mathbb{R}_+^\times)^r.$$

(Of course, both sides are trivial if S is anisotropic.)

Define $d\tilde{t}$ by

$$d\tilde{t} := \widetilde{\Lambda}^*\biggl(\prod_{k=1}^r \frac{dt}{t}\biggr).$$

Let $dS_{\mathbb{Q}}$ be the counting measure on $S(\mathbb{Q})$. The Tamagawa measure on $S(\mathbb{A})^1$ defined by Ono [1961, (3.5.2)] (taking into account that in our case the base field denoted by k in [Ono 1961] is $\mathbb{Q}$) is the measure μ^{Tama} that makes the following equality true:

$$\rho_S^{-1}\omega^{\text{Tama}} = d\tilde{t}\,\mu^{\text{Tama}}\,dS_{\mathbb{Q}}. \tag{5-1}$$

The Tamagawa number is defined by

$$\tau_S = \int_{S^1(\mathbb{A})/S(\mathbb{Q})} \mu^{\text{Tama}}.$$

We will also make use of the differential form on S that we denote by ω_S (this notation agrees with that of [Frenkel et al. 2010]). We define

$$\omega_S := \frac{d\chi_1}{\chi_1} \wedge \cdots \wedge \frac{d\chi_d}{\chi_d}, \tag{5-2}$$

where d is the rank of $X^*(S)$. This form is, a priori, defined over $\overline{\mathbb{Q}}$. However, in fact there exists $D \in \mathbb{Q}$ such that $\omega_S/\sqrt{D}$ is defined over $\mathbb{Q}$. (see [Gross and Gan 1999, Corollary 3.7]). Since the product formula $\sqrt{|D|}\prod_\ell \sqrt{|D|_\ell} = 1$ still holds, we can, in fact, use the form ω_S instead of ω in the definition of the Tamagawa measure, even though it is not quite defined over $\mathbb{Q}$. Specifically, we will from now on work with the form

$$\omega_S^{\text{Tama}} = (\omega_S)_\infty \prod_\ell L_\ell(1, \sigma_S)(\omega_S)_\ell \quad \text{on } S(\mathbb{A}). \tag{5-3}$$

We denote the product over the finite primes by $\omega_{S,f}$, i.e., write $\omega_S^{\text{Tama}} = (\omega_S)_\infty \omega_{S,f}$; the form $\omega_{S,f}$ defines a measure on $S(\mathbb{A}_f)$, the set of points of S over the finite adeles.

5B. ***The measure μ^{Tama} versus geometric measure.*** This section is based on [Frenkel et al. 2010]. We recall that the measure on orbits that we call μ^{geom} is constructed as a quotient of the measure $|\omega_G|$ by the measure $|\omega_A|$ on our space $\mathbb{A}_G$, which is a "naïve version" of the Steinberg–Hitchin base (see [Gordon 2022, §3.7] for a detailed comparison of the space $\mathbb{A}_G$, and the measure on it, with the actual Steinberg–Hitchin base that is used in [Frenkel et al. 2010]).

Consider the measure on T defined by the form ω_T^{Tama} at every place (its only difference from the Tamagawa measure on T is in the global factor ρ_T). For every finite prime ℓ, let $\mu_{\gamma_\ell}^{\mathrm{Tama}}$ be the measure on the orbit of γ_ℓ in $G(\mathbb{Q}_\ell)$ obtained as the quotient of the measure ω^{can} on G that gives the maximal compact subgroup $G(\mathbb{Z}_\ell)$ volume 1, by the measure $|\omega_T^{\mathrm{Tama}}|_\ell$ on T.

The following proposition is an adaptation of the equality (3.31) of [Frenkel et al. 2010] to our setting.

Proposition 5.1 [Frenkel et al. 2010, Proposition 3.29; (3.31)]. *We have*

$$\mu_{\gamma,\ell}^{\mathrm{geom}} = |\eta(\gamma)|_\ell^{-\frac{1}{4}g(g+1)}\sqrt{|D(\gamma)|_\ell}\,\mathrm{vol}_{\omega_G}(G(\mathbb{Z}_\ell))L_\ell(1,\sigma_T)\mu_\ell^{\mathrm{Tama}},$$

where $\eta(\gamma)$ is the multiplier of γ.

Proof. For $\gamma \in G^{\mathrm{der}}$, this is equivalent to the relation (3.31) of [Frenkel et al. 2010]; the additional factor $\mathrm{vol}_{\omega_G}(G(\mathbb{Z}_\ell))$ on the right appears here because in (3.31) of [Frenkel et al. 2010], the same measure on G needs to be used on both sides of the equation; here we are using the measure $|\omega_G|$ on the left, and the measure $|\omega_G^{\mathrm{can}}| = |\omega_G|/\mathrm{vol}(G(\mathbb{Z}_\ell))$ on the right; so this correction factor is needed. More precisely, the relation (3.30) in [Frenkel et al. 2010] (which we also reproved in Section 4.2.1 of [Achter and Gordon 2017]) asserts that for a semisimple group G,

$$\mu_{\gamma,\ell}^{\mathrm{geom}} = \sqrt{|D(\gamma)|_\ell}\,|\omega_{T\backslash G}|,$$

where $\omega_{T\backslash G}$ is the quotient of the measure ω_G by the measure ω_T on T defined above in Section 5A, which is the same as the measure ω_T in [Frenkel et al. 2010]. Since by definition (and the remark at the end of Section 5A which allows us to use the form ω_T in the definition of the Tamagawa measure), $\omega_{T,\ell}^{\mathrm{Tama}} = L_\ell(1,\sigma_T)\omega_{T,\ell}$, this proves the proposition for $\gamma \in G^{\mathrm{der}}$. For general γ, the factor $|\eta(\gamma)|^{-\frac{1}{4}g(g+1)}$ appears on the right-hand side because we are using the space $\mathbb{A}_G$ instead of the Steinberg–Hitchin base of GSp_{2g}. This factor is calculated by considering the action of the centre of G on all the measure spaces involved. This is explained in detail in Section 3.7 of [Gordon 2022]. □

5C. ***Proof of Theorem A.*** For convenience, we list here the results we proved above about the ratios ν_ℓ and the relevant orbital integrals:

(1) Corollary 4.7 ($\ell \neq p$) $$\nu_\ell([X,\lambda]) = \frac{\ell^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_\ell/\ell)}O_{\gamma_0}^{\mathrm{geom}}(\phi_0).$$

(2) Corollary 4.12 $$\nu_p([X,\lambda]) = q^{-\frac{1}{2}g(g+1)}\frac{p^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{Z}_p/p)}O_{\gamma_0}^{\mathrm{geom}}(\phi_{q,p}).$$

Combining these with Proposition 5.1, and observing that for $\ell \neq p$, we have $|\eta(\gamma)|_\ell = |\det(\gamma)|_\ell = 1$, and at p, we have $|\eta(\gamma)|_p = q^{-1}$, we obtain

$$\begin{aligned}
\nu_p([X,\lambda]) &= q^{-\frac{1}{2}g(g+1)} \frac{p^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{F}_p)} O^{\mathrm{geom}}_{\gamma_0}(\phi_{q,p}) \\
&= q^{-\frac{1}{2}g(g+1)} q^{\frac{1}{4}g(g+1)} \sqrt{|D(\gamma)|_p}\, \mathrm{vol}_{\omega_G}(G(\mathbb{Z}_p)) L_p(1,\sigma_T) \frac{p^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{F}_p)} O^{\mathrm{Tama}}_{\gamma_0}(\phi_{q,p}) \\
&= q^{-\frac{1}{4}g(g+1)} \sqrt{|D(\gamma)|_p} L_p(1,\sigma_{T/G}) O^{\mathrm{Tama}}_{\gamma_0}(\phi_{q,p}), \\
\nu_\ell([X,\lambda]) &= \frac{\ell^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{F}_\ell)} O^{\mathrm{geom}}_{\gamma_0}(\phi_\ell) \\
&= \sqrt{|D(\gamma)|_\ell}\, \mathrm{vol}_{\omega_G}(G(\mathbb{Z}_\ell)) L_\ell(1,\sigma_T) \frac{\ell^{\dim(G^{\mathrm{der}})}}{\#G^{\mathrm{der}}(\mathbb{F}_\ell)} O^{\mathrm{Tama}}_{\gamma_0}(\phi_\ell) \\
&= \sqrt{|D(\gamma)|_\ell} L_\ell(1,\sigma_{T/G}) O^{\mathrm{Tama}}_{\gamma_0}(\phi_\ell).
\end{aligned} \tag{5-4}$$

Here the notation $L_\ell(1,\sigma_{T/G})$ stands for $L_\ell(1,\sigma_T)(1-1/\ell)$ (including the case $\ell = p$), which agrees with the use of this notation in [Frenkel et al. 2010], and the last equality in both cases follows from (2-5).

Taking a product of these over all primes ℓ, and recalling the product formula for absolute values, we obtain

$$\prod_\ell \nu_\ell = q^{-\frac{1}{4}g(g+1)} |D(\gamma)|^{-\frac{1}{2}} L(1,\sigma_{T/G}) O^{\mathrm{Tama}}_\gamma, \tag{5-5}$$

where O^{Tama}_γ stands for the product of orbital integrals in the Langlands–Kottwitz formula, with the measure on each factor given by $\omega^{\mathrm{Tama}}_{T,\ell}$.

Lemma 5.2. $$\mathrm{vol}_{\omega^{\mathrm{Tama}}_{T,f}}(T(\mathbb{Q})\backslash T(\mathbb{A}_f)) = \rho_T \frac{1}{(2\pi)^g} \tau(T).$$

Proof. We start by emphasizing that while in the definition of the Tamagawa number, one starts with an arbitrary differential form defined over $\mathbb{Q}$, since here we have split off the infinite places, the specific choice of the differential form matters. This choice (dictated by our calculations above, specifically, Proposition 5.1) is the form ω_T of (5-2) (which is not defined over $\mathbb{Q}$ but can still be used, as discussed at the end of Section 5A). From this form, we make the form $\omega^{\mathrm{Tama}}_T = (\omega_T)_\infty \omega_{T,f}$ as in (5-3). We need to identify explicitly the component at the infinite place of the form ω^{Tama}_T, which is the same as the component $(\omega_T)_\infty$ of the differential form ω_T. We have a convenient basis for the character lattice of T; see Section 2B. Define the coordinates $(z_1,\dots,z_g,\lambda)$ on $T(\mathbb{C})$ such that $T(\mathbb{C}) = \{z_1,\dots,z_g,\lambda z_1^{-1},\dots,\lambda z_g^{-1})\}$, and define the characters $\chi_i(z_1,\dots,z_g,z_1^{-1},\dots,\lambda z_g^{-1}) = z_i$, for $i = 1,\dots,g$, and the multiplier $\eta(z_1,\dots,z_g,z_1^{-1},\dots,\lambda z_g^{-1}) := \lambda$. (Note that in the Appendix, the character lattice of T is described as a quotient of $\mathbb{Z}^{2g}$ instead, which is convenient for the cohomology computations; we do not use this description here.) We write every element of $T(\mathbb{A})$ as $a = a_f a_\infty$, where a_f has the infinity component 1 and $a_\infty = (1, z_1,\dots,z_g,\lambda)$ has all the components at the finite places equal to 1. In this notation, $T(\mathbb{A})^1$ is defined by the condition $|z_i| = |\lambda| = \|a_f\|^{-1/g}$. We note that the character η coincides with the map $\widetilde{\Lambda}$ from $T(\mathbb{A})/T(\mathbb{A})^1$ to $\mathbb{R}_+$ defined in Section 5A. Then, by the definition of μ^{Tama}, it is the volume form on $T(\mathbb{A})^1$ given by

$$\mu^{\mathrm{Tama}} \wedge \frac{d\eta}{\eta} = \frac{d\chi_1(a_f a_\infty)}{\chi_1(a_f a_\infty)} \wedge \cdots \wedge \frac{d\chi_g(a_f a_\infty)}{\chi_g(a_f a_\infty)} \wedge \frac{d\eta}{\eta},$$

and thus its component at ∞ is the form $\mu_\infty^{\mathrm{Tama}} = (dz_1/z_1) \wedge \cdots \wedge (dz_g/z_g)$ on $T(\mathbb{A})^1_\infty \simeq (S^1)^g$. It is an easy exercise (see [Gordon 2022, Example 2.4]) that the form dz/z gives precisely the arc-length measure on the unit circle. Thus, we get

$$\tau_T = \mathrm{vol}_{\mu^{\mathrm{Tama}}}(T(\mathbb{Q})\backslash T(\mathbb{A})^1) = \rho_T^{-1}\,\mathrm{vol}_{\omega_T^{\mathrm{Tama}}}(T(\mathbb{Q})\backslash T(\mathbb{A})^1) = \rho_T^{-1}\,\mathrm{vol}_{\omega_T^{\mathrm{Tama}}}(T(\mathbb{Q})\backslash T(\mathbb{A}_f))(2\pi)^g,$$

which completes the proof. □

Now we can complete the proof of the theorem. By the Langlands–Kottwitz formula (in which we choose the Tamagawa measure on T), we have

$$\widetilde{\#}I([X,\lambda]) = \mathrm{vol}_{\omega_T^{\mathrm{Tama}}}(T(\mathbb{Q})\backslash T(\mathbb{A}_f))\,O_\gamma^{\mathrm{Tama}} = \rho_T \frac{1}{(2\pi)^g}\tau_T\, O_\gamma^{\mathrm{Tama}}.$$

On the other hand, by (5-5), we have

$$\prod_\ell \nu_\ell = q^{-\frac{1}{4}g(g+1)}|D(\gamma)|^{-\frac{1}{2}} L(1,\sigma_{T/G})\,O_\gamma^{\mathrm{Tama}}.$$

It remains to recall that we have defined $\nu_\infty = |D(\gamma)|^{\frac{1}{2}}/(2\pi)^g$, and note that the Euler product for $L(1,\sigma_{T/G})$ is conditionally convergent and equals ρ_T (see [Frenkel et al. 2010, (3.25)]), since $\rho_{\mathbb{G}_m}$ is equal to 1, the residue of the Riemann zeta-function at $s=1$. Thus,

$$q^{-\frac{1}{4}g(g+1)}\tau_T\,\nu_\infty \prod_\ell \nu_\ell = \frac{\tau_T}{(2\pi)^g} L(1,\sigma_{T/G})\,O_\gamma^{\mathrm{Tama}} = \mathrm{vol}_{\omega_T^{\mathrm{Tama}}}(T(\mathbb{Q})\backslash T(\mathbb{A}_f))\,O_\gamma^{\mathrm{Tama}},$$

which completes the proof.

6. Complements

For the convenience of a hypothetical reader interested in explicit calculations, we collect here some reminders concerning the terms which arise in (1-3).

6A. ν_∞. Recall that we have defined (4-2) $\nu_\infty([X,\lambda])$ as $\sqrt{|D(\gamma_0)|}/(2\pi)^g$, where $D(\gamma_0)$ is the Weyl discriminant $D(\gamma_0) = \prod_{\alpha\in\Phi}(1-\alpha(\gamma_0))$, the product being over all roots of G. We may relate this to the (polynomial) discriminant of $f_{X/\mathbb{F}_q}(T)$, the characteristic polynomial of Frobenius, as follows.

6A1. *Weyl discriminants.* Explicitly, γ_0 has multiplier $\lambda_0 := \eta(\gamma_0) = q$. Write the (complex) eigenvalues of (a $\mathbb{Q}$-representative of) γ_0 — equivalently, the roots of $f_{X/\mathbb{F}_q}(T)$ — as $(\lambda_1,\ldots,\lambda_g,\lambda_0/\lambda_1,\ldots,\lambda_0/\lambda_g)$. Then

$$D(\gamma) = \prod_{1\le i<j\le g}\delta_{ij}\cdot\prod_{1\le i\le g}\delta_i,$$

where

$$\delta_{ij} = (1-\lambda_i/\lambda_j)(1-\lambda_j/\lambda_i)(1-\lambda_i\lambda_j/\lambda_0)(1-\lambda_0/(\lambda_i\lambda_j)) \quad\text{and}\quad \delta_i = (1-\lambda_i^2/\lambda_0)(1-\lambda_0/\lambda_i^2).$$

Possibly after reordering the conjugate pairs $\{\lambda_i, \lambda_0/\lambda_i\}$, we may and do assume that $\lambda_j = \sqrt{q}\exp(i\theta_j)$ with $0 \leq \theta_j < \pi$. Then

$$\delta_{ij} = (2\cos(\theta_i) - 2\cos(\theta_j))^2 \quad \text{and} \quad \delta_j = 4\sin^2(\theta_j).$$

6A2. *Elliptic curves.* Suppose that $[X, \lambda]$ is an elliptic curve with its canonical principal polarization, say with characteristic polynomial of Frobenius $T^2 - aT + q$. Then $a = 2\sqrt{q}\cos(\theta)$, and $D(\gamma_0) = 4\sin^2(\theta) = 4 - a^2/q$, and

$$\nu_\infty([X,\lambda]/\mathbb{F}_q) = \frac{1}{2\pi}\sqrt{|D(\gamma_0)|} = \frac{1}{\pi}\sqrt{1 - \frac{a^2}{4q}}.$$

Note that this term is *half* the archimedean term introduced in [Gekeler 2003, (3.3)] (when $q = p$) and [Achter and Gordon 2017, (2-7)]. For purposes of comparison, we summarize this relationship by writing

$$\nu_\infty([X,\lambda]) = \tfrac{1}{2}\nu_\infty^{\mathrm{Gek}}([X,\lambda]) = \tfrac{1}{2}\nu_\infty^{\mathrm{AG}}([X,\lambda]).$$

6A3. *Polynomial discriminants.* To facilitate comparison with [Achter and Williams 2015; Gekeler 2003; Gerhard and Williams 2019], we express $D(\gamma_0)$ in terms of polynomial discriminants. Let $f(T) = f_{X/\mathbb{F}_q}(T)$, and let $f^+(T) = f^+_{X/\mathbb{F}_q}(T)$ be the minimal polynomial of the sum of γ_0 and its adjoint, so that

$$f^+_{X/\mathbb{F}_q}(T) = \prod_{1\leq j\leq g}(T - (\lambda_j + q/\lambda_j)).$$

Note that $\mathbb{Q}[T]/f^+(T) \cong K^+$, the maximal totally real subalgebra of the endomorphism algebra of X.

Lemma 6.1.
$$\frac{\operatorname{disc}(f(T))}{\operatorname{disc}(f^+(T))} = (-1)^g q^{\frac{1}{2}g(3g-1)} D(\gamma_0).$$

Proof. On one hand,

$$\operatorname{disc}(f(T)) = \prod_{1\leq i<j\leq g}\alpha_{ij}^2 \prod_{1\leq i\leq g}\alpha_i^2,$$

where

$$\alpha_{ij} = (\lambda_i - \lambda_j)(\lambda_i - \lambda_0/\lambda_j)(\lambda_0/\lambda_i - \lambda_j)(\lambda_0/\lambda_i - \lambda_0/\lambda_j) \quad \text{and} \quad \alpha_i = (\lambda_i - \lambda_0/\lambda_i).$$

On the other hand,

$$\operatorname{disc}(f^+(T)) = \prod_{1\leq i<j\leq g}\beta_{ij}^2,$$

where

$$\beta_{ij} = (\lambda_i + \lambda_0/\lambda_i - (\lambda_j + \lambda_0/\lambda_j)).$$

Now use this to evaluate $\operatorname{disc}(f(T))/\operatorname{disc}(f^+(T))$, while bearing in mind that

$$\frac{\alpha_{ij}}{\beta_{ij}^2} = \lambda_0, \quad \frac{\alpha_{ij}}{\delta_{ij}} = \lambda_0^2 \quad \text{and} \quad \frac{\alpha_i^2}{\delta_i} = -\lambda_0. \qquad \square$$

6B. ν_ℓ**.** Gekeler [2003] observed that, for elliptic curves, his product formula essentially computes an L-function; a similar phenomenon has been observed in other contexts as well [Achter and Williams 2015; Gerhard and Williams 2019]. We briefly explain how this relates to (1-3). This detour also has the modest benefit of showing that the right-hand side of (1-3) converges, albeit conditionally.

6B1. *Zeta functions.* We express the zeta function of a number field M as $\zeta_M(s) = \prod_\ell \zeta_{M,\ell}(s)$, where $\zeta_{M,\ell}(s) = \prod_{\lambda|\ell}(1 - \mathrm{N}_{M/\mathbb{Q}}(\lambda)^{-s})^{-1}$. For a direct sum $M = \bigoplus_{i=1}^t M_i$ of such fields we write $\zeta_{M,\ell}(s) = \prod_i \zeta_{M_i,\ell}(s)$; the product over all primes yields $\zeta_M(s) = \prod_i \zeta_{M_i}(s)$.

Recall that to a torus $S/\mathbb{Q}$ one associates an Artin L-function $L(s, \sigma_S) = \prod_\ell L_\ell(s, \sigma_S)$. This construction is multiplicative for exact sequences of tori, and for a finite direct sum M of number fields one has $L(s, \sigma_{\boldsymbol{R}_{M/\mathbb{Q}}\mathbb{G}_m}) = \zeta_M(s)$. (It may be worth recalling that $\boldsymbol{R}_{M/\mathbb{Q}}\mathbb{G}_m \cong \bigoplus \boldsymbol{R}_{M_i/\mathbb{Q}}\mathbb{G}_m$.)

If ℓ is unramified in some splitting field for S, then (see [Bitan 2011, 2.8; Voskresenskiĭ 1998, 14.3])

$$\#S(\mathbb{F}_\ell) = \ell^{\dim S} L_\ell(1, \sigma_S)^{-1}.$$

Lemma 6.2. *Suppose that* $\ell \nmid 2p \operatorname{disc}(f_{X/\mathbb{F}_q}(T))$. *Then*

$$\nu_\ell([X, \lambda]) = \frac{\zeta_{K,\ell}(1)}{\zeta_{K^+,\ell}(1)}.$$

Proof. By Lemma 4.4,

$$\nu_\ell([X, \lambda]) = \frac{\#\{\gamma \in G(\mathbb{F}_\ell) : \gamma \sim \pi_1(\gamma_0)\}}{\#G(\mathbb{F}_\ell)/\#\mathbb{A}_G(\mathbb{F}_\ell)} = \frac{\#G(\mathbb{F}_\ell)/\#T(\mathbb{F}_\ell)}{\#G(\mathbb{F}_\ell)/(\ell^g \#\mathbb{G}_m(\mathbb{F}_\ell))} = \ell^g \frac{\#\mathbb{G}_m(\mathbb{F}_\ell)}{\#T(\mathbb{F}_\ell)} = \frac{L_\ell(1, \sigma_T)}{L_\ell(1, \sigma_{\mathbb{G}_m})},$$

since $\dim T = g + 1$. Using (2-3), first to see that $L(s, \sigma_T) = L(s, \sigma_{T^{\mathrm{der}}}) L(s, \sigma_{\mathbb{G}_m})$ and second to compute $L(s, \sigma_{T^{\mathrm{der}}})$, we recognize the final expression in the above equation to be

$$\frac{\zeta_{K,\ell}(1)}{\zeta_{K^+,\ell}(1)}. \qquad \square$$

Since $\zeta_K(s)$ and $\zeta_{K^+}(s)$ both have a simple pole at $s = 1$, we immediately deduce:

Corollary 6.3. *The right-hand side of* (1-3) *converges conditionally.*

Moreover, up to a finite factor $B([X, \lambda])$, we can express $\widetilde{\#}I([X, \lambda], \mathbb{F}_q)$ in terms of familiar quantities:

Corollary 6.4. *We have*

$$\widetilde{\#}I([X, \lambda], \mathbb{F}_q) = \tau_T \frac{q^{-\frac{1}{4}g(3g-1)}}{(2\pi)^g} \sqrt{\left|\frac{\operatorname{disc}(f)}{\operatorname{disc}(f^+)}\right|} B([X, \lambda]) \lim_{s \to 1^+} \frac{\zeta_K(s)}{\zeta_{K^+}(s)},$$

where

$$B([X, \lambda]) = \prod_{\ell | 2p \operatorname{disc}(f)} \frac{\zeta_{K^+,\ell}(1)}{\zeta_{K,\ell}(1)} \nu_\ell([X, \lambda]).$$

6B2. *Elliptic curves.* If $[X, \lambda]$ is an elliptic curve, let χ be the quadratic character associated to the imaginary quadratic field K. Then $K^+ = \mathbb{Q}$, and $\zeta_K(s)/\zeta_{\mathbb{Q}}(s) = L(s, \chi)$.

Now further suppose that $q = p$ and that the Frobenius order is maximal, i.e., that $\mathbb{Z}[T]/f_{X/\mathbb{F}_q}(T) \cong \mathcal{O}_K$. Then Gekeler shows with an explicit calculation that for *each* prime ℓ, $\nu_\ell^{\mathrm{Gek}}([X,\lambda]) = L_\ell(1,\chi)$, and thus $\prod_\ell \nu_\ell^{\mathrm{Gek}}([X,\lambda]) = L(1,\chi)$.

6B3. *Abelian varieties with maximal Frobenius order.* Similarly, suppose $[X,\lambda]$ is an ordinary abelian surface with $\mathrm{End}(X)\otimes\mathbb{Q}$ a cyclic quartic extension of $\mathbb{Q}$, and further suppose that the Frobenius order is maximal. Achter and Williams [2015] define a local term $\nu_\ell^{AW}([X,\lambda])$, and show that $\prod_\ell \nu_\ell^{AW}([X,\lambda]) = \zeta_K(1)/\zeta_{K^+}(1)$. This observation has been extended to certain abelian varieties of prime dimension [Gerhard and Williams 2019, Proposition 8.1].

6C. $\boldsymbol{\nu_p}$**.** Since the multiplier $\eta(\gamma_0)$ of Frobenius is q, γ_0, while an element of $M(\mathbb{Z}_p)\subset G(\mathbb{Q}_p)$, is never an element of $G(\mathbb{Z}_p)$. Nonetheless, if the isogeny class is *ordinary*, it is possible to transfer part of the work in calculating $\nu_p([X,\lambda])$ from $M(\mathbb{Z}_p)$ to $G(\mathbb{Z}_p)$, as follows.

Suppose X is ordinary. Then its p-divisible group splits *integrally* as $X[p^\infty] = X[p^\infty]^{\mathrm{tor}} \oplus X[p^\infty]^{\mathrm{\acute{e}t}}$. (In general, the slope filtration only exists up to isogeny, as in Lemma 3.5.) Thus, there exists a decomposition $V_{\mathbb{Z}_p} = V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}} \oplus V_{\mathbb{Z}_p}^{\mathrm{tor}}$ into maximal isotropic summands stable under γ_0, where $\alpha_0 := \gamma_0|_{V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}}} \in \mathrm{End}(V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}})$ is invertible; and the polarization induces an isomorphism $V_{\mathbb{Z}_p}^{\mathrm{tor}}$ with the dual of $V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}}$, such that $\gamma_0|_{V_{\mathbb{Z}_p}^{\mathrm{tor}}} = q(\alpha_0^\top)^{-1}$. This can also be proved directly through linear algebra. Indeed, let $\beta_0 = q\gamma_0^{-1}$. Then $V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}} = \bigcap_n \gamma_0^{on}(V_{\mathbb{Z}_p})$, while $V_{\mathbb{Z}_p}^{\mathrm{tor}} = \bigcap_n \beta_0^{on}(V_{\mathbb{Z}_p})$.

Lemma 6.5. *For n and d sufficiently large and $\gamma \in M(\mathbb{Z}_p/p^n)$, the following conditions are equivalent*:

(a) $\gamma \sim_{M(\mathbb{Z}_p/p^n)_d} \gamma_0 \bmod p^n$;

(b) *there exists some $\tilde{\gamma} \in M(\mathbb{Z}_p)$ such that $\tilde{\gamma} \bmod p^n = \gamma$ and $\tilde{\gamma} \sim_{G(\mathbb{Q}_p)} \gamma_0$*;

(c) *γ stabilizes a decomposition $V_{\mathbb{Z}_p/p^n} \cong V^+_{\mathbb{Z}_p/p^n} \oplus V^-_{\mathbb{Z}_p/p^n}$ into maximal isotropic subspaces, and there exists an isomorphism $\iota : V^+_{\mathbb{Z}_p/p^n} \to V^{\mathrm{\acute{e}t}}_{\mathbb{Z}_p/p^n}$ such that $\iota^*\alpha_0 = \gamma|_{V^+_{\mathbb{Z}_p/p^n}}$.*

Proof. The equivalence of (a) and (b) is Lemma 3.2. For the equivalence of (a) and (c), use the argument above to show that any such γ induces an appropriate decomposition of $V_{\mathbb{Z}_p/p^n}$. □

Therefore, if $\alpha_0 \bmod p$ is regular, we obtain a version of Lemma 6.2 at p.

Corollary 6.6. *Suppose $[X,\lambda]$ is ordinary and* $\mathrm{ord}_p \operatorname{disc}(f_{X/\mathbb{F}_q}(T)) = e\cdot g(g-1)$. *Then*

$$\nu_p([X,\lambda]) = \frac{\zeta_{K,p}(1)}{\zeta_{K^+,p}(1)}.$$

Note that we always have $q^{g(g-1)} \mid \operatorname{disc}(f_{X/\mathbb{F}_q}(T))$. The case $g = 1$ also follows from the explicit calculation in [Gekeler 2003, Theorem 4.4].

Proof. Define $\epsilon_0 \in \mathrm{Sp}(V_{\mathbb{Z}_p})$ by $\epsilon_0|_{V_{\mathbb{Z}_p}^{\mathrm{\acute{e}t}}} = \alpha_0$ and $\epsilon_0|_{V_{\mathbb{Z}_p}^{\mathrm{tor}}} = (\alpha_0^\top)^{-1}$.

The argument of Lemma 6.5 shows that, for sufficiently large d and n, both $\#C_{(d,n)}(\gamma_0)$ and $\#C_{(d,n)}(\gamma_0)$ are given by the product of the number of decompositions $V_{\mathbb{Z}_p/p^n} = V^+_{\mathbb{Z}_p/p^n} \oplus V^-_{\mathbb{Z}_p/p^n}$ into maximal isotropic subspaces, and the number of $\alpha \in \operatorname{End}(V_{\mathbb{Z}_p/p^n})$ with $\alpha \sim_{\operatorname{End}(V_{\mathbb{Z}_p/p^n})_d} \alpha_0$. In particular, $\#C_{(d,n)}(\gamma_0) = \#C_{(d,n)}(\epsilon_0)$.

The regularity hypothesis implies that ϵ_0 mod p is regular, and the result follows from Lemma 6.2. □

6D. *Explicit examples.*

6D1. $g = 1$. Consider the elliptic curve $E/\mathbb{F}_7$ with affine equation $y^2 = x^3 + x + 1$. Its Frobenius polynomial is $f_E(T) = T^2 - 3T + 7$, which has discriminant -19, a fundamental discriminant. So the order generated by the Frobenius endomorphism is the ring of integers in $K := \mathbb{Q}(\sqrt{-19})$; using Magma, we numerically estimate $\prod_\ell \nu_\ell(E/\mathbb{F}_7) = L\big(1, \big(\frac{-19}{\bullet}\big)\big)$ as ≈ 0.72073. Continuing to work numerically, we obtain that $\nu_\infty(E/\mathbb{F}_7) = \frac{1}{2\pi}\sqrt{4 - \frac{3^2}{7}} \approx 0.2622$. Since $T = \boldsymbol{R}_{K/\mathbb{Q}}\mathbb{G}_m$, we have $\tau_T = 1$, and therefore $\tau_T\sqrt{7}\nu_\infty(E/\mathbb{F}_7)\prod_\ell \nu_\ell(E/\mathbb{F}_7) \approx 0.5000$.

This reflects the easily verified arithmetic statement that the only elliptic curve over $\mathbb{F}_7$ with trace of Frobenius 3 is E itself; and $\operatorname{Aut}(E) \cong \mathcal{O}_K^\times = \{\pm 1\}$, so that the weighted size of this isogeny class is $\widetilde{\#}I([E], \mathbb{F}_7) = \frac{1}{2}$. (In modest contrast, [Gekeler 2003] assigns weight $2/\#\operatorname{Aut}(F)$ to an elliptic curve F; this is reflected in the fact that $\nu_\infty^{\mathrm{Gek}}([E], /\mathbb{F}_7) = 2\nu_\infty([E], \mathbb{F}_7)$.)

6D2. $g = 4$. Consider the 3-Weil polynomial

$$f(T) = T^8 - 6T^7 + 13T^6 - 10T^5 + T^4 - 30T^3 + 117T^2 - 162T + 81.$$

It turns out that there is a unique principally polarized abelian fourfold (X, λ) over $\mathbb{F}_3$ with characteristic polynomial equal to $f(T)$. (This is a single data point in a census of isogeny classes which will soon be integrated into the LMFDB.)

Let $K = \mathbb{Q}[T]/f(T)$. One readily checks that $\operatorname{disc}(f(T))/\operatorname{disc}(K) = 3^{4(4-1)}$, and so $\nu_\ell([X, \lambda]) = \zeta_{K,\ell}(1)/\zeta_{K^+,\ell}(1)$ for all finite ℓ, including $\ell = p$. Again, we numerically estimate $\prod_\ell \nu_\ell([X, \lambda]) = \lim_{s\to 1^+} \zeta_K(s)/\zeta_{K^+}(s) \approx 0.871253$ and $\nu_\infty([X, \lambda]) \approx 0.000111808$. The field K is Galois over $\mathbb{Q}$, with group $\operatorname{Gal}(K/\mathbb{Q}) \cong \mathbb{Z}/4 \oplus \mathbb{Z}/2$, and the Tamagawa number of the torus T is 2 (see Section A6). Our formula numerically yields

$$\widetilde{\#}I([X, \lambda], \mathbb{F}_3) = 3^3 \tau_T \nu_\infty([X, \lambda], \mathbb{F}_3) \lim_{s\to 1} \frac{\zeta_K(s)}{\zeta_{K^+}(s)} \approx 0.050000.$$

This reflects the fact that the torsion group of $\mathcal{O}_K^\times$, and thus $\operatorname{Aut}([X, \lambda])$, has order 20.

The referee points out that it is possible to take advantage of other recent work on isogeny classes, which appeared essentially simultaneously with or since the first preprint of this work was made available, to independently verify these assertions without recourse to the LMFDB. For example, our characterization of K shows that in fact $K = \mathbb{Q}(\zeta_{20})$; and the discriminant calculation shows that $\operatorname{End}(X) = \mathbb{Z}[\zeta_{20}]$. Since $h(K) = h(K)/h(K^+) = 1$, one can use [Howe 2022, Corollary 1.2] or [Guo et al. 2022, Theorem 1.1] to conclude that $\#I([X, \lambda], \mathbb{F}_3) = 1$.

The state of the art on calculation of Tamagawa numbers has also improved; see [Liang et al. 2021; Rüd 2022b].

6E. ***Level structure.*** The Langlands–Kottwitz formula (2-1) is actually written for abelian varieties with arbitrary level structure, and thus a version of our main formula is available in the context of abelian varieties with level structure, too.

6E1. *Product formula.* Let $\Gamma \subset G(\hat{\mathbb{Z}}_f^p)$ be an open compact subgroup. There is a notion of principally polarized abelian variety with level Γ structure; let $\mathcal{A}_{g,\Gamma}$ be the corresponding Shimura variety. If $(X, \lambda, \alpha) \in \mathcal{A}_{g,\Gamma}(\mathbb{F}_q)$ is a principally polarized abelian variety with level Γ-structure, then the size of its isogeny class in this category is given by the Kottwitz formula, except that the integrand in the adelic orbital integral is replaced with $\mathbb{1}_\Gamma$.

We make the definition

$$\nu_\ell([X, \lambda, \alpha]) = \lim_{d\to\infty} \lim_{n\to\infty} \frac{\#(C_{(d,n)}(\gamma_0) \cap \pi_n(\Gamma_\ell))}{\#G(\mathbb{Z}_\ell/\ell^n)/\#\mathbb{A}_G(\mathbb{Z}_\ell/\ell^n)}.$$

The analogue of Corollary 4.6 holds, and states that there exists $d(\gamma_0)$ such that

$$O_{\gamma_0}^{\mathrm{geom}}(\mathbb{1}_{\Gamma_\ell}) = \lim_{n\to\infty} \frac{\mathrm{vol}_{|d\omega_G|}(\widetilde{C}_{(d(\gamma_0),n)}(\gamma_0) \cap \Gamma_\ell)}{\mathrm{vol}_{|d\omega_A|}(\widetilde{U}_n(\gamma_0))}.$$

The calculations at p and ∞, as well as the global volume term, are unchanged, and we find that

$$\widetilde{\#}I([X, \lambda, \alpha], \mathbb{F}_q) = q^{\frac{1}{4}g(g+1)} \tau_T \nu_\infty([X, \lambda]) \prod_\ell \nu_\ell([X, \lambda, \alpha]). \tag{6-1}$$

6E2. *Principal level structure.* Fix a prime ℓ_0, and define $\Gamma(\ell_0) = \prod_\ell \Gamma(\ell_0)_\ell$ by

$$\Gamma(\ell_0)_\ell = \begin{cases} G(\mathbb{Z}_\ell) & \ell \neq \ell_0, \\ \ker(G(\mathbb{Z}_{\ell_0}) \to G(\mathbb{Z}_{\ell_0}/\ell_0)) & \ell = \ell_0. \end{cases}$$

Then $\mathcal{A}_{g,\Gamma(\ell_0)}$ is the moduli space of abelian varieties equipped with a full principal level ℓ_0-structure.

For example, to fix ideas, suppose that $g = 1$ and that $\ell_0 \neq p$, and let a satisfy $|a| \le 2\sqrt{q}$, $p \nmid a$ and $\ell_0 \parallel (a^2 - 4q)$; we consider the set of elliptic curves with characteristic polynomial of Frobenius $f(T) = T^2 - aT + q$. Then some, but not all, elements of the corresponding isogeny class admit a principal level ℓ_0-structure (see, e.g., [Achter and Wong 2013]).

Let (X, λ, α) be an elliptic curve over $\mathbb{F}_q$ with trace of Frobenius a and full level ℓ_0-structure α. We may explicitly compute $\nu_{\ell_0}([X_0, \lambda, \alpha])$ as follows. Let $\chi_{\ell_0} = (\,\cdot\,/\ell_0)$ be the quadratic character modulo ℓ_0.

Lemma 6.7.
$$\nu_{\ell_0}([X, \lambda, \alpha]) = \frac{1}{\ell_0^2} \frac{1}{1 - \chi_{\ell_0}(\mathrm{disc}(f)/\ell_0^2)/\ell_0}.$$

Proof. Let $\gamma_0 = \gamma_{X,\mathbb{F}_q,\ell_0}$ be a Frobenius element for X at ℓ_0. By hypothesis, $\gamma_0 = 1 + \ell_0\beta_0$ for some $\beta_0 \in \mathrm{Mat}_2(\mathbb{Z}_{\ell_0})$. Since ℓ_0^2 is the highest power of ℓ_0 dividing $\mathrm{disc}(f)$, we in fact have $\beta_0 \in \mathrm{GL}_2(\mathbb{Z}_{\ell_0})$, and β_0 is regular mod ℓ_0, i.e., $\pi_1(\beta_0)$ is regular.

Suppose that $\gamma \in \Gamma_{\ell_0}$ satisfies $\gamma \sim_{G(\mathbb{Q}_{\ell_0})} \gamma_0$. Then $\gamma = 1 + \ell_0\beta$ for some $\beta \in \mathrm{GL}_2(\mathbb{Z}_{\ell_0})$ which is regular mod ℓ_0, and direct calculation shows $\beta \sim_{G(\mathbb{Q}_{\ell_0})} \beta_0$. Lemma 3.1 then shows that $\beta \sim_{G(\mathbb{Z}_\ell)} \beta_0$.

Consequently, for any $d \geq 0$ and any $n \geq 2$, we have bijections between the following sets:

$$\{\gamma \in G(\mathbb{Z}_{\ell_0}/\ell_0^n) : \gamma \sim_{M(\mathbb{Z}_{\ell_0}/\ell_0^n)_d} \pi_n(\gamma_0)\};$$
$$\{\beta \in G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1}) : \beta \sim_{M(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})_d} \pi_{n-1}(\beta_0)\};$$
$$\{\beta \in G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1}) : \beta \sim_{G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})} \pi_n(\beta_0)\}.$$

Therefore,

$$\begin{aligned}\nu_{\ell_0}([X,\lambda,\alpha]) &= \frac{\#G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})/\#G_{\beta_0}(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})}{\#G(\mathbb{Z}_{\ell_0}/\ell_0^{n})/\#\mathbb{A}_G(\mathbb{Z}_{\ell_0}/\ell_0^{n})} = \frac{\#G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})}{\#G(\mathbb{Z}_{\ell_0}/\ell_0^{n})}\frac{\#\mathbb{A}_G(\mathbb{Z}_{\ell_0}/\ell_0^{n})}{\#G_{\beta_0}(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})} \\ &= \frac{1}{\ell_0^4}\frac{\ell_0^2\#\mathbb{A}_G(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})}{\#G_{\beta_0}(\mathbb{Z}_{\ell_0}/\ell_0^{n-1})} = \frac{1}{\ell_0^2}\frac{\#\mathbb{A}_G(\mathbb{Z}_{\ell_0}/\ell_0)}{\#G_{\beta_0}(\mathbb{Z}_{\ell_0}/\ell_0)} = \frac{1}{\ell_0^2}\frac{1}{1-\chi_{\ell_0}(\mathrm{disc}(f)/\ell_0^2)/\ell_0}. \qquad \square\end{aligned}$$

7. GL_2 reconsidered

In [Achter and Gordon 2017], we essentially treated the $g = 1$ case of the present paper. Unfortunately, a simple algebra error — $\nu_\infty^{\mathrm{AG}}([X,\lambda]) = \frac{2}{\pi}\sqrt{|D(\gamma_0)|}$ (Section 6A2), in spite of the claims of the penultimate displayed equation [Achter and Gordon 2017, p. 20] — masked certain mistakes involving the calculations at p. We take the opportunity to correct these mistakes. The reader pleasantly unaware of these issues with [Achter and Gordon 2017] may simply view the present section as an explication of our technique in the special case where $g = 1$, and thus $G = \mathrm{GL}_2$.

Note that the definition [Achter and Gordon 2017, (2-6)] could have been replaced with a criterion involving characteristic polynomials, e.g.,

$$\nu_p(a,q) = \lim_{n\to\infty}\frac{\#\{\gamma \in \mathrm{Mat}_2(\mathbb{Z}_p/p^n) : f_\gamma \equiv f_{\gamma_0} \bmod p^n\}}{\#G(\mathbb{Z}_p/p^n)/\#A(\mathbb{Z}_p/p^n)}.$$

7A. *Assertions at p.* There are two problematic claims in [Achter and Gordon 2017]:

(1) For the test function $\mathbb{1}_{G(\mathbb{Z}_\ell)}$ at $\ell \neq p$, we have

$$\nu_\ell(a,q) = \frac{\ell^3}{\#\mathrm{SL}_2(\mathbb{F}_\ell)}\mathcal{O}_{\gamma_0}^{\mathrm{geom}}(\mathbb{1}_{G(\mathbb{Z}_\ell)}).$$

It is claimed in [Achter and Gordon 2017, Lemma 3.7] that the same is true for $\ell = p$, where the test function ϕ_q is the characteristic function of $G(\mathbb{Z}_p)\left(\begin{smallmatrix} q & 0 \\ 0 & 1\end{smallmatrix}\right)G(\mathbb{Z}_p)$.

(2) In [Achter and Gordon 2017, Appendix], revisiting the calculation [Frenkel et al. 2010, (3.30)], we assert that

$$\mu_{\gamma,\ell}^{\mathrm{geom}} = \sqrt{|D(\gamma)|_\ell}\,\frac{\mathrm{vol}_{|\omega_G|_\ell}(G(\mathbb{Z}_\ell))}{\mathrm{vol}_{|\omega_T|_\ell}(T^\circ(\mathbb{Z}_\ell))}\bar{\mu}_{T\backslash G,\ell}. \tag{7-1}$$

This is valid for $\ell \neq p$, but requires correction at p, because our Chevalley–Steinberg map is not exactly the same as the map in [Frenkel et al. 2010].

7B. *From point counts to measure.* In (1), one exploits the fundamental fact that point counts mod ℓ^n converge to volume with respect to the Serre–Oesterlé measure μ^{SO}. In the case where $\ell = p$, however, the ambient space is $\mathrm{Mat}_2(\mathbb{Z}_p)$, rather than (its open subset) $G(\mathbb{Z}_p)$. Thus, the volume of the set V_n of [Achter and Gordon 2017] should be computed with respect to $\mu^{\mathrm{SO}}_{\mathrm{Mat}_2}$, which is the usual measure on the 4-dimensional affine space. We recall that for $\gamma \in \mathrm{GL}_2(\mathbb{Q}_p)$, $d\mu^{\mathrm{SO}}_{\mathrm{GL}_2}(\gamma) = |\det(\gamma)|^{-2} d\mu^{\mathrm{SO}}_{\mathrm{Mat}_2}(\gamma)$.

Moreover, one should be using the invariant measure $dx \wedge (dy/|y|)$ on the Steinberg base $\mathbb{A}_{\mathrm{GL}_2} \cong \mathbb{A}^1 \times \mathbb{G}_m$, rather than the measure pulled back from $\mathbb{A}^1 \times \mathbb{A}^1$. Then the measure of a radius p^{-n} neighbourhood U_n of (a, q) in A is $p^{-2n}/|\det(\gamma_0)| = p^{-2n}/q^{-1}$, and we find

$$\begin{aligned} \nu_{p,n}(a,q) = \frac{p^3}{\#\mathrm{SL}_2(\mathbb{F}_p)} \frac{|\det(\gamma_0)|^2 \operatorname{vol}_{\mu_{GL_2}}(V_n(\gamma_0))}{|\det(\gamma_0)| \operatorname{vol}_{\mathbb{A}^1\times\mathbb{G}_m}(U_n)} &= |\det(\gamma_0)| \frac{p^3}{\#\mathrm{SL}_2(\mathbb{F}_p)} O^{\mathrm{geom}}_{\gamma_0}(\phi_q) \\ &= q^{-1} \frac{p^3}{\#\mathrm{SL}_2(\mathbb{F}_p)} O^{\mathrm{geom}}_{\gamma_0}(\phi_q). \end{aligned}$$

(This differs from the assertion of [Achter and Gordon 2017, Lemma 3.7] by a factor of q.)

7C. *From geometric measure to canonical measure.* Since orbital integrals of rational-valued functions with respect to the canonical measure are rational, while $\sqrt{|D(\gamma_0)|_p} = \sqrt{q}$, the assertion of (2) cannot hold at $\ell = p$.

While the relation between the geometric measure and the canonical measure that we rely on is correct for a semisimple group, it needs a correction factor for a reductive group. This part is completely general for all reductive G, and is discussed in detail in [Gordon 2022]; the correct formula is stated in Proposition 5.1. In particular, the correct calculation at p is

$$\mu^{\mathrm{geom}}_{\gamma,p} = |\eta(\gamma)|_p^{-\frac{1}{4}g(g+1)} \sqrt{|D(\gamma)|_p} \frac{\operatorname{vol}_{|\omega_G|_p}(G(\mathbb{Z}_p))}{\operatorname{vol}_{|\omega_T|_p}(T^\circ(\mathbb{Z}_p))} \bar{\mu}_{T\backslash G,p},$$

where $\eta(\gamma)$ is the multiplier of γ.

Appendix

by Wen-Wei Li and Thomas Rüd

We compute the Tamagawa numbers of some anisotropic tori in GSp_{2g} and Sp_{2g} associated with a single Galois field extension (see Section 2B), and present a partial result towards the general case that illustrates the difficulties.

Recall the setup of Section 2B in the case of a single Galois extension. Let $K \supset K^+ \supset \mathbb{Q}$ be a tower of field extensions with K Galois, such that $[K : K^+] = 2$ and $[K^+ : \mathbb{Q}] = g$. We define

$$T^{\mathrm{der}} = \operatorname{Ker}\big(\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m) \xrightarrow{N_{K/K^+}} \boldsymbol{R}_{K^+/\mathbb{Q}}(\mathbb{G}_m)\big) = \boldsymbol{R}_{K^+/\mathbb{Q}} \boldsymbol{R}^{(1)}_{K/K^+}(\mathbb{G}_m) \subset \mathrm{Sp}_{2g},$$

and

$$T = \operatorname{Ker}\big(\mathbb{G}_m \times_{\operatorname{Spec}(\mathbb{Q})} \boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m) \xrightarrow{(x,y)\mapsto x^{-1}N_{K/K^+}(y)} \boldsymbol{R}_{K^+/\mathbb{Q}}(\mathbb{G}_m)\big) \subset \mathrm{GSp}_{2g}.$$

These fit in the short exact sequence

$$1 \to T^{\mathrm{der}} \to T \to \mathbb{G}_m \to 1.$$

Also, recall that we have been using τ_T (resp. $\tau_{T^{\mathrm{der}}}$) to denote the Tamagawa numbers $\tau_{\mathbb{Q}}(T)$ and $\tau_{\mathbb{Q}}(T^{\mathrm{der}})$. We will show the following.

Proposition A.1. $$\tau_{T^{\mathrm{der}}} = \tau_{K^+}(\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) = 2.$$

For the case of T, the result varies with the extension.

Proposition A.2. *Assume that K is a Galois CM-field and K^+ is its maximal totally real subfield. Then we have $\tau_T \le 2$ and*:

- *If g is odd, then $\tau_T = 1$.*
- *If $K/\mathbb{Q}$ is cyclic, then $\tau_T = 1$.*
- *If $g = 2$, then $\tau_T = 1$ when $\mathrm{Gal}(K/\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z}$ and $\tau_T = 2$ when $\mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^2$.*

Additional results and details appear in the second author's thesis [Rüd 2022a]; see also [Rüd 2022b].

We base our approach on the following formula of Ono.

Theorem A.3 [Ono 1963]. *Let $\boldsymbol{T}$ be an algebraic torus defined over a number field F and split over some Galois extension L. Then its Tamagawa number can be computed as*

$$\tau_F(\boldsymbol{T}) = \frac{|H^1(L/F, \boldsymbol{X}^\star(\boldsymbol{T}))|}{|\text{Ш}^1(\boldsymbol{T})|}.$$

Here $\boldsymbol{X}^\star(\boldsymbol{T})$ denotes the character lattice of $\boldsymbol{T}$. The symbol $\text{Ш}^1(\boldsymbol{T})$ denotes the corresponding Tate–Shafarevich group defined by

$$\text{Ш}^i(\boldsymbol{T}) = \mathrm{Ker}\Big(H^i(L/F, \boldsymbol{T}) \to \prod_v H^i(L_w/F_v, \boldsymbol{T})\Big), \tag{A-1}$$

where v runs over the primes of F and w is a prime of L with $w|v$.

Our approach is to do the computation on the level of character lattices. A very important consequence of the Tate–Nakayama duality theorem (see [Platonov and Rapinchuk 1994, Theorem 6.10]) is that for a torus $\boldsymbol{T}$ as in the previous theorem, the Pontryagin dual of $\text{Ш}^1(\boldsymbol{T})$ is isomorphic to $\text{Ш}^2(\boldsymbol{X}^\star(\boldsymbol{T}))$, so it suffices to compute $|\text{Ш}^2(\boldsymbol{X}^\star(\boldsymbol{T}))|$.

The proof of Proposition A.1 is given in the next section. The proof of Proposition A.2 occupies Sections A2–A5. In Section A6 we present an example not covered by Proposition A.2. In Section A7 we present a computation for the numerator that illustrates the difficulties that arise for a general torus (not assuming that T is constructed from a single field).

A1. *Computation of $\tau_{T^{\mathrm{der}}}$*. We write the proof of Proposition A.1 using Theorem A.3. Since the Tamagawa number is preserved by restriction of scalars we have $\tau_{T^{\mathrm{der}}} = \tau_{\mathbb{Q}}(\boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) = \tau_{K^+}(\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m)$. The cohomology of the characters of a norm 1 torus is obtained by a classic computation that one can see, for instance, in the proof of the Hasse norm theorem in [Platonov and Rapinchuk 1994, Theorem 6.11]. We have

$$\widehat{H}^i(K/\mathbb{Q}, \boldsymbol{X}^\star(T^{\mathrm{der}})) = \widehat{H}^i(K/K^+, \boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) = \widehat{H}^{i+1}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } i \text{ is odd,} \\ \{0\} & \text{if } i \text{ is even.} \end{cases}$$

In particular, $|\text{Ш}^1(T^{\mathrm{der}})| = |\text{Ш}^2(\boldsymbol{X}^\star(T^{\mathrm{der}}))| \le |H^2(\boldsymbol{X}^\star(T^{\mathrm{der}}))| = 1$. We conclude $\tau_{T^{\mathrm{der}}} = \frac{2}{1} = 2$.

This proves Proposition A.1.

A2. ***Computation of the first cohomology group of the character lattice.*** From now on, we focus on the proof of Proposition A.2, and therefore we will assume that K is a CM-field with K^+ its maximal totally real subfield. Let ι be the nontrivial element of $\mathrm{Gal}(K/K^+)$, and let Γ and Γ^+ denote respectively the Galois groups of $K/\mathbb{Q}$ and $K^+/\mathbb{Q}$. Note that K^+ is indeed Galois over $\mathbb{Q}$ by virtue of K being a CM-field. The torus T arises as the subtorus of $\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m)$ with the set of $\mathbb{Q}$-points consisting of elements $x \in K^\times$ such that $x\iota(x) \in \mathbb{Q}$, and $T^{\mathrm{der}}(\mathbb{Q})$ is the set of elements $x \in K^\times$ such that $x\iota(x) = 1$. We have the following exact sequence of finite groups:

$$1 \to \langle\iota\rangle \cong \mathbb{Z}/2\mathbb{Z} \to \Gamma \to \Gamma^+ \to 1. \tag{A-2}$$

For each $\sigma \in \Gamma^+$ fix a preimage $\hat{\sigma} \in \Gamma$. We get the description of $\boldsymbol{X}^\star(\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m)) = \mathbb{Z}[\Gamma]$ as the set of $\mathbb{Z}$-linear combinations of $\hat{\sigma}$ and $\hat{\sigma}\iota$ with $\sigma \in \Gamma^+$.

The embedding of T in $\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m)$ gives us a surjective map $\boldsymbol{X}^\star(\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m)) \to \boldsymbol{X}^\star(T)$. For $\chi = \sum_{\sigma\in\Gamma^+} a_\sigma\hat{\sigma} + \sum_{\sigma\in\Gamma^+} b_\sigma\hat{\sigma}\iota \in \boldsymbol{X}^\star(\boldsymbol{R}_{K/\mathbb{Q}}(\mathbb{G}_m))$ and $t \in T(\mathbb{Q})$, we have

$$\begin{aligned}
\chi(t) &= \prod_{\sigma\in\Gamma^+} \hat{\sigma}(t)^{a_\sigma} \prod_{\sigma\in\Gamma^+} \hat{\sigma}(\iota(t))^{b_\sigma} \\
&= \prod_{\sigma\in\Gamma^+} \hat{\sigma}(t)^{a_\sigma} \prod_{\sigma\in\Gamma^+} \hat{\sigma}(\lambda t^{-1})^{b_\sigma} \quad (\text{where } t\iota(t) = \lambda \in \mathbb{Q}^\times) \\
&= \lambda^{\sum_{\sigma\in\Gamma^+} b_\sigma} \prod_{\sigma\in\Gamma^+} \hat{\sigma}(t)^{a_\sigma - b_\sigma}.
\end{aligned}$$

We get the descriptions

$$\begin{aligned}
\boldsymbol{X}^\star(T) &= \mathbb{Z}[\Gamma]/\Big\{\sum_{\sigma\in\Gamma^+} a_\sigma\hat{\sigma} + \sum_{\sigma\in\Gamma^+} b_\sigma\hat{\sigma}\iota : a_\sigma = b_\sigma \text{ for } \sigma\in\Gamma^+ \text{ and } \sum_{\sigma\in\Gamma^+} b_\sigma = 0\Big\} && \text{(A-3)} \\
&= \mathbb{Z}[\Gamma]/L, && \text{(A-4)}
\end{aligned}$$

where $L = \big\{\sum_{\sigma\in\Gamma^+} a_\sigma\hat{\sigma}(1+\iota) : \sum_{\sigma\in\Gamma^+} a_\sigma = 0\big\}$. For T^{der}, we have $\lambda = 1$, so we recover

$$\boldsymbol{X}^\star(T^{\mathrm{der}}) = \mathbb{Z}[\Gamma]/\{x = \iota(x)\} = \mathbb{Z}[\widehat{\Gamma^+}] \otimes \mathbb{Z}[\iota]/\langle 1+\iota\rangle = \boldsymbol{X}^\star(\boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}(\mathbb{G}_m)). \tag{A-5}$$

In order to compute $H^1(K/\mathbb{Q}, \boldsymbol{X}^\star(T))$ we use the inflation-restriction exact sequence, which one can find in [Gille and Szamuely 2006, Proposition 3.3.14, p. 65]. To simplify notation, let $\Lambda = \boldsymbol{X}^\star(T) = \mathbb{Z}[\Gamma]/L$ as in (A-3).

The inflation-restriction exact sequence associated with the short exact sequence (A-2) takes the form

$$0 \to H^1(\Gamma^+, \Lambda^{\mathbb{Z}/2\mathbb{Z}}) \to H^1(\Gamma, \Lambda) \to H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda)^{\Gamma^+} \to H^2(\Gamma^+, \Lambda^{\mathbb{Z}/2\mathbb{Z}}) \to H^2(\Gamma, \Lambda). \tag{A-6}$$

Lemma A.4. *The sequence (A-6) can be rewritten as*

$$0 \to 0 \to H^1(\Gamma, \Lambda) \to (\mathbb{Z}/2\mathbb{Z})^{\frac{1}{2}(1+(-1)^g)} \to \Gamma^{+\mathrm{ab}} \to H^2(\Gamma, \Lambda). \tag{A-7}$$

In particular, $\tau_T \le |H^1(\Gamma, \boldsymbol{X}^\star(T))| \le 2$.

Proof. Let $x \in \mathbb{Z}[\Gamma]$, and let $[x]$ denote its class in Λ. Clearly $[x]$ is fixed by ι if and only if $x - \iota x \in L$, and since every element of L is fixed by ι, then so must $x - \iota x$ be, which forces $x = \iota x$. Therefore,

$$\Lambda^{\mathbb{Z}/2\mathbb{Z}} = \left\{ \sum_{\sigma\in\Gamma^+} a_\sigma \hat{\sigma}(1+\iota) \right\} \Big/ \left\langle \sum_{\sigma\in\Gamma^+} a_\sigma \hat{\sigma}(1+\iota) : \sum_{\sigma\in\Gamma^+} a_\sigma = 0 \right\rangle \cong \mathbb{Z}[\Gamma^+]/I,$$

where I is the augmentation ideal of $\mathbb{Z}[\Gamma^+]$, i.e., the subspace of sum-zero vectors. Further, observe that $\mathbb{Z}[\Gamma^+]/I \cong \mathbb{Z}$ as Γ^+-modules, where $\mathbb{Z}$ has trivial Γ^+-action (by definition of I). We get that

$$H^1(\Gamma^+, \Lambda^{\mathbb{Z}/2\mathbb{Z}}) \cong H^1(\Gamma^+, \mathbb{Z}) = \mathrm{Hom}(\Gamma^+, \mathbb{Z}) = \{0\}.$$

Also, using the sequence

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0, \tag{A-8}$$

since the middle term is uniquely divisible hence cohomologically trivial, one has

$$H^2(\Gamma^+, \mathbb{Z}) \cong H^1(\Gamma^+, \mathbb{Q}/\mathbb{Z}) = \mathrm{Hom}(\Gamma^+, \mathbb{Q}/\mathbb{Z});$$

the last term is noncanonically isomorphic to Γ^{+ab}.

The only term left to compute is $H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda)^{\Gamma^+}$. As a $\mathbb{Z}/2\mathbb{Z}$-module, we can write $\Lambda = \mathbb{Z}^g \oplus \mathbb{Z}^g/L$, where $L = \{(a, a) : a = (a_1, \dots, a_g) \text{ such that } \sum a_i = 0\}$, and $\mathbb{Z}/2\mathbb{Z}$ acts as $(a, b) \mapsto (b, a)$. Therefore,

$$0 \to L \to \mathbb{Z}^g \oplus \mathbb{Z}^g = \mathbb{Z}[\mathbb{Z}/2\mathbb{Z}]^g \to \Lambda \to 0.$$

Since the middle term is cohomologically trivial as a $\mathbb{Z}/2\mathbb{Z}$-module, and $\mathbb{Z}/2/Z$ acts trivially on L,

$$H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda) \cong H^2(\mathbb{Z}/2\mathbb{Z}, L) \cong \widehat{H}^0(\mathbb{Z}/2\mathbb{Z}, L) \cong L/2L.$$

To compute $L/2L$, we view L as a submodule of $\mathbb{Z}^g$ of zero-sum elements. Recall that by construction of L as a group algebra, Γ^+ acts transitively on L.

Let $\boldsymbol{a} = (a_1, \dots, a_g) \in L$. We want to compute $(L/2L)^{\Gamma^+}$, and for that, we reason on the parity of a_i's.

- If all a_i are even, then $\boldsymbol{a} = 2\boldsymbol{a}'$ and $\boldsymbol{a}' \in L$, and hence $\boldsymbol{a} \in 2L$.
- If $\boldsymbol{a}$ has a_i even and a_j odd, considering a permutation $\sigma \in \Gamma^+$ sending the i-th coordinate to the j-th, we have that the j-th coordinate of $\boldsymbol{a} - \sigma\boldsymbol{a}$ is $a_j - a_i$, which is odd, and hence $\boldsymbol{a} - \sigma\boldsymbol{a} \notin 2L$.
- The last case to consider is when all a_i are odd. In that case, $\sum_i a_i$ has the same parity as g, so $\boldsymbol{a} \in L$ can only happen if g is even. One can prove that every element of $L/2L$ has a representative of the form $\boldsymbol{a} = (a_1, \dots, a_g) \in \mathbb{Z}^g$ with $\sum_i a_i = 0$ and $|a_i| \le 1$. When g is even and all a_i are odd, the only

possible such elements are vectors with half the coordinates being -1 and the other half 1. Moreover all such vectors are in the same coset of $2L$ (one can permute the ± 1 coordinates by adding ± 2).

This shows that $(L/2L)^{\Gamma^+}$ contains no nontrivial element when g is odd, and only one when g is even, and concludes the proof. $\square$

Corollary A.5. *When g is odd, one has $H^1(K/\mathbb{Q}, \boldsymbol{X}^\star(T)) = \{0\}$ and*

$$H^2(K/\mathbb{Q}, \boldsymbol{X}^\star(T)) \cong \Gamma^{+\mathrm{ab}}.$$

In particular, $H^2(K/\mathbb{Q}, \boldsymbol{X}^\star(T))$ has odd order, and so does $Ш^1(T)$, since it is dual to $Ш^2(\boldsymbol{X}^\star(T)) \subset H^2(K/\mathbb{Q}, \boldsymbol{X}^\star(T))$.

Proof. The first equality comes directly from the sequence (A-7).

For the second equality, taking duals of the exact sequence (2-3), one gets

$$0 \to \mathbb{Z} \to \boldsymbol{X}^\star(T) \to \boldsymbol{X}^\star(T^{\mathrm{der}}) \to 0.$$

The cohomology of this sequence gives us

$$H^1(\Gamma, \boldsymbol{X}^\star(T)) \to H^1(\Gamma, \boldsymbol{X}^\star(T^{\mathrm{der}})) \to H^2(\Gamma, \mathbb{Z}) \to H^2(\Gamma, \boldsymbol{X}^\star(T)) \to H^2(\Gamma, \boldsymbol{X}^\star(T^{\mathrm{der}})).$$

We computed the cohomology $H^i(\Gamma, \boldsymbol{X}^\star(T^{\mathrm{der}})) = H^i(\mathbb{Z}/2\mathbb{Z}, \boldsymbol{X}^\star(\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m))$ in Section A1.

We can substitute $H^1(\Gamma, \boldsymbol{X}^\star(T)) = \{0\} = H^2(\Gamma, \boldsymbol{X}^\star(T^{\mathrm{der}}))$, $H^2(\Gamma, \mathbb{Z}) \cong H^1(\Gamma, \mathbb{Q}/\mathbb{Z}) \cong \Gamma^{ab}$, and $H^2(\Gamma, \boldsymbol{X}^\star(T^{\mathrm{der}})) \cong \mathbb{Z}/2\mathbb{Z}$, which gives us

$$0 \to \mathbb{Z}/2\mathbb{Z} \to \Gamma^{ab} \to H^2(\Gamma, \boldsymbol{X}^\star(T)) \to 0,$$

as desired. $\square$

Proposition A.6. *When the sequence (A-2) splits,*

$$H^1(K/\mathbb{Q}, \boldsymbol{X}^\star(T)) = \begin{cases} \{0\} & \text{if } g \text{ is odd}, \\ \mathbb{Z}/2\mathbb{Z} & \text{if } g \text{ is even}. \end{cases}$$

In particular, this gives an alternate proof of the triviality of $H^1(K/\mathbb{Q}, \boldsymbol{X}^\star(T))$ whenever g is odd.

Proof. Since (A-2) splits, one can write the inflation-restriction exact sequence associated with the short exact sequence

$$1 \to \Gamma^+ \to \Gamma \to \mathbb{Z}/2\mathbb{Z} \to 1.$$

This gives us

$$0 \to H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda^{\Gamma^+}) \to H^1(\Gamma, \Lambda) \to H^1(\Gamma^+, \Lambda)^{\mathbb{Z}/2\mathbb{Z}} \to H^2(\mathbb{Z}/2\mathbb{Z}, \Lambda^{\Gamma^+}) \to H^2(\Gamma, \Lambda). \tag{A-9}$$

Since the sequence (A-2) splits, we have $\mathbb{Z}[\Gamma] = \mathbb{Z}[\Gamma^+] \otimes \mathbb{Z}[\iota]$ as a Γ-module. We have $\Lambda = \mathbb{Z}[\Gamma^+] \otimes \mathbb{Z}[\iota]/I \otimes (1+\iota)$, where I is the augmentation ideal of $\mathbb{Z}[\Gamma^+]$. Because $\mathbb{Z}[\Gamma^+]$ is an induced module, and $I \otimes (1+\iota) \cong I$ as a Γ^+-module, we get $\widehat{H}^i(\Gamma^+, \Lambda) = \widehat{H}^{i+1}(\Gamma^+, I)$. Now since $\mathbb{Z} = \mathbb{Z}[\Gamma^+]/I$ with $\mathbb{Z}$ seen as a trivial module, the same argument yields $\widehat{H}^i(\Gamma^+, \Lambda) \cong \widehat{H}^i(\Gamma^+, \mathbb{Z})$. In particular, $H^1(\Gamma^+, \Lambda) = \{0\}$, so the sequence (A-9) gives an isomorphism $H^1(\Gamma, \Lambda) \cong H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda^{\Gamma^+})$.

Direct computations using that $\overset{+}{\sigma} := \sum_{\sigma\in\Gamma^+}\sigma$ spans the set of Γ^+-fixed elements of $\mathbb{Z}[\Gamma^+]$ give us that $\{1\otimes(1+\iota), \overset{+}{\sigma}\otimes\iota\}$ is a $\mathbb{Z}$-basis for Λ^{Γ^+}, on which ι acts via $\left(\begin{smallmatrix}1 & g\\ 0 & -1\end{smallmatrix}\right)$. Identifying the space with $\mathbb{Z}^2$ we can compute cocycles and coboundaries. Coboundaries are of the form $a_\iota = (-gb, 2b)$ for $b\in\mathbb{Z}$. Cocycles are of the form $a_\iota = (a, b)$ with $2a+gb = 0$. Thus, if b is even then it is a coboundary, if b is odd then g cannot be odd, and so we only get a nontrivial cocycle with g even and b odd. The difference of two nontrivial cocycles has an even second entry, so it is a coboundary. This proves

$$H^1(\mathbb{Z}/2\mathbb{Z}, \Lambda^{\Gamma^+}) = \begin{cases}\{0\} & \text{if } g \text{ is odd},\\ \mathbb{Z}/2\mathbb{Z} & \text{if } g \text{ is even},\end{cases}$$

as desired.

For the last assertion in the proposition, when g is odd, by the Schur–Zassenhaus theorem (see [Rotman 1995, Theorem 7.41]) the sequence (A-2) splits and we get our result immediately. □

A3. ***Case g odd.*** Now we are ready to show:

Lemma A.7. *If g is odd, then $\tau_T = \frac{1}{1} = 1$.*

Proof. Using Corollary A.5 we have that $|H^1(K/\mathbb{Q}, \boldsymbol{X}^\star(T))| = 1$, and $Ш^1(T)$ has odd order. To show that $Ш^1(T)$ is trivial, it suffices to show it is 2-torsion. The cohomology of the sequence (2-3) yields

$$H^1(K/\mathbb{Q}, T^{\mathrm{der}}) \to H^1(K/\mathbb{Q}, T) \to H^1(K/\mathbb{Q}, \mathbb{G}_m) = 1,$$

where the right equality holds by Hilbert 90. So we have a surjection of $H^1(K/\mathbb{Q}, T^{\mathrm{der}})$ onto $H^1(K/\mathbb{Q}, T)$. We claim that $H^1(K/\mathbb{Q}, T)$ is 2-torsion; it suffices so show that $H^1(K/\mathbb{Q}, T^{\mathrm{der}})$ is.

Taking the cohomology of the sequence

$$1 \to \boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m \to \boldsymbol{R}_{K/K^+}\mathbb{G}_m \xrightarrow{N_{K/K^+}} \mathbb{G}_m \to 1,$$

where the middle term is cohomologically trivial, we have $H^1(K/K^+, \boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) \cong \widehat{H}^0(K/K^+, \mathbb{G}_m) = (K^+)^\times/N_{K/K^+}K^\times$.

This gives us

$$H^1(K/\mathbb{Q}, T^{\mathrm{der}}) = H^1(K/\mathbb{Q}, \boldsymbol{R}_{K^+/\mathbb{Q}}\boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) = H^1(K/K^+, \boldsymbol{R}^{(1)}_{K/K^+}\mathbb{G}_m) \cong (K^+)^\times/N_{K/K^+}K^\times.$$

This group is 2-torsion, hence so is $H^1(K/\mathbb{Q}, T)$ and $Ш^1(T)$ is a subgroup of the latter. We can conclude that $Ш(T)$ is a 2-torsion group of odd order, hence it is trivial.

We can conclude using Theorem A.3 that $\tau_T = \frac{1}{1} = 1$.

Note that one need not use Corollary A.5 to know that $Ш^1(T)$ has odd order and hence is trivial. Indeed, given that the extension K/K^+ is quadratic, by the Chebotarev density theorem, we know that there is a prime $\mathfrak{p}\in K^+$ inert in the extension K/K^+. Since $\mathfrak{p}$ is stable under ι, which is of order 2, then its decomposition group $\Gamma(\mathfrak{p})$ has even order, and therefore odd index in Γ. Now it suffices to look at the restriction-corestriction sequence $H^1(\Gamma, T) \to H^1(\Gamma(\mathfrak{p}), T) \to H^1(\Gamma, T)$. The composition of the two maps is just multiplication by $n = [\Gamma : \Gamma(\mathfrak{p})]$. By definition of $Ш^1(T)$, this subgroup of $H^1(\Gamma, T)$ is killed by the restriction map, hence it is n-torsion, and we know n is odd, as desired. □

A4. ***The case $K/\mathbb{Q}$ cyclic.***

Lemma A.8. *When K is cyclic, we have $\tau(T) = \frac{1}{1} = 1$. In particular, this holds when $K^+/\mathbb{Q}$ is cyclic of odd order.*

Proof. Write $X^\star(T) = \mathbb{Z}[\Gamma]/L$ as in (A-3).

By virtue of K being cyclic, the Tate cohomology is 2-periodic, so

$$H^1(K/\mathbb{Q}, X^\star(T)) \cong H^2(K/\mathbb{Q}, L) \cong \widehat{H}^0(K/\mathbb{Q}, L).$$

Since L has trivial ι action, we can see it as the augmentation ideal of $\mathbb{Z}[\Gamma^+]$, which has no Γ^+-fixed point, as any augmentation ideal. In particular it has no Γ-fixed point and so $\widehat{H}^0(K/\mathbb{Q}, L) = \{0\}$.

Again using the fact that K is cyclic, we get that $Ш^1(T)$ is trivial. Indeed, by the Chebotarev density theorem, every cyclic extension has a prime $p \in \mathbb{Z}$ that will stay inert, and therefore $\Gamma(p) = \Gamma$, where $\Gamma(p)$ is the corresponding decomposition group. Therefore, the map in the definition of $Ш^1(T)$ is injective.

We can conclude by Theorem A.3 that $\tau_T = \frac{1}{1} = 1$. □

A5. ***Case $g = 2$.***

Lemma A.9. *When $g = 2$ we have $\tau_T = 1$ if Γ is cyclic and $\tau_T = 2$ otherwise.*

Proof. The first case is a consequence of Lemma A.8. If Γ isn't cyclic, the only possibility is $\Gamma \cong (\mathbb{Z}/2\mathbb{Z})^2$.

In that case, Proposition A.6 gives $H^1(\Gamma, X^\star(T)) \cong \mathbb{Z}/2\mathbb{Z}$. Concerning the Tate–Shafarevich group, Cortella [1997] showed that $Ш^1(T) = \{0\}$ if $g < 4$.

Alternatively, since Γ is abelian, every proper cyclic subgroup appears as decomposition group. In this specific case, it is a consequence of the Chinese remainder theorem and quadratic reciprocity. Using SageMath computations (see below), we obtain that the map

$$H^2(\Gamma, X^\star(T)) \to H^2(\mathbb{Z}/2\mathbb{Z} \times \{0\}, X^\star(T)) \oplus H^2(\{0\} \times \mathbb{Z}/2\mathbb{Z}, X^\star(T))$$

is injective, and therefore $Ш^1(T) = 0$. □

A6. ***Computing the Tamagawa numbers with SageMath.*** The second author implemented methods in SageMath to deal with algebraic tori through their character lattices. Those methods should eventually be added to SageMath in a future release.

Here we briefly describe the computation of the Tamagawa number which arises in Section 6D2, where $K = \mathbb{Q}[T]/f(T)$ for

$$f(T) = T^8 - 6T^7 + 13T^6 - 10T^5 + T^4 - 30T^3 + 117T^2 - 162T + 81.$$

We have $\Gamma = \mathrm{Gal}(K/\mathbb{Q}) = \mathbb{Z}/4 \oplus \langle \iota \rangle$, where ι denotes the complex involution.

Nakayama duality lets us compute the Tamagawa number as a function of the character lattice:

$$\tau_{\mathbb{Q}}(T) = \frac{|H^1(\mathbb{Q}, X^\star(T))|}{|Ш^2(X^\star(T))|}.$$

Let Λ denote $\boldsymbol{X}^\star(T)$. We build Λ in SageMath by inducing the trivial lattice $\mathbb{Z} = \boldsymbol{X}^\star(\mathbb{G}_m)$ to Γ, build the sublattice of zero-sum elements of ι-fixed points, and take the quotient of the former by the latter.

We can compute the first cohomology group by computing cocycles as solutions of linear equations in $\Lambda^{|\Gamma|}$. However, for this example, Proposition A.6 gives us $H^1(\Gamma, \Lambda) = 2$. For the denominator, we build a method that, given $\mathcal{H}$ a collection of subgroups of Γ, and a Γ-lattice L, computes

$$Ш^1_{\mathcal{H}}(L) = \operatorname{Ker}\left(H^1(\Gamma, L) \to \bigoplus_{\Delta \in \mathcal{H}} H^1(\Delta, L) \right)$$

by checking which cocycles restrict to coboundaries on all $\Delta \in \mathcal{H}$.

Consider the embedding $\varphi : \Lambda \to \mathbb{Z}[\Gamma] \otimes_{\mathbb{Z}} \Lambda$ (with Γ-action on the left component) via $a \mapsto \sum_{g \in \Gamma} g \otimes g^{-1} a$. We build

$$\Lambda' = \mathbb{Z}[\Gamma] \otimes_{\mathbb{Z}} \Lambda / \varphi(\Lambda).$$

Since $\mathbb{Z}[\Gamma] \otimes_{\mathbb{Z}} \Lambda$ is induced, it is cohomologically trivial, so $\widehat{H}^i(\Gamma, \Lambda) = \widehat{H}^{i-1}(\Gamma, \Lambda')$ for all $i \in \mathbb{Z}$. Consequently, if every element of $\mathcal{H}$ arises as a decomposition group, we have

$$Ш^1_{\mathcal{H}}(\Lambda') \supset Ш^1(\Lambda') = Ш^2(\Lambda).$$

We take $\mathcal{H}$ to be the list of cyclic subgroups of Γ. They all arise as decomposition groups. Using our new SageMath package, we get $|Ш^1_{\mathcal{H}}(\Lambda')| = 1 \geq Ш^2(\Lambda) \geq 1$, and hence

$$\tau(T) = \frac{|H^1(\Gamma, \Lambda)|}{|Ш^2(\Lambda)|} = \frac{2}{1} = 2.$$

A7. ***The numerator in Ono's formula: general case.*** We give some indications for the general case in which T is described by a CM-algebra $K = \bigoplus_{i=1}^t K_i$, each K_i being a CM-field. The Rosati involution on K is still denoted as ι, with fixed subalgebra $K^+ = \bigoplus_{i=1}^t K_i^+$. In this section we denote by Γ the absolute Galois group of $\mathbb{Q}$. Our modest aim is to understand $|H^1(\Gamma, \boldsymbol{X}^\star(T))|$ through Kottwitz's isomorphism (see [Kottwitz 1984b, (2.4.1) and §2.4.3])

$$H^1(\Gamma, \boldsymbol{X}^\star(T)) \cong \pi_0(\widehat{T}^\Gamma),$$

where $\widehat{T}$ is the dual $\mathbb{C}$-torus. This isomorphism is valid for all tori.

To describe $\boldsymbol{X}^\star(T)$, we first write T as

$$T = (\mathbb{G}_m \times T^{\mathrm{der}})/\{(z, z) : z \in \mu_2\}, \qquad \mu_2 := \{\pm 1\}.$$

Choose a subset $\Phi = \bigsqcup_{i=1}^t \Phi_i$ of $\operatorname{Hom}_{\mathbb{Q}\text{-alg}}(K, \overline{\mathbb{Q}})$, such that $\Phi_i \subset \operatorname{Hom}_{\mathbb{Q}\text{-alg}}(K_i, \overline{\mathbb{Q}})$ and

$$\operatorname{Hom}_{\mathbb{Q}\text{-alg}}(K_i, \overline{\mathbb{Q}}) = \Phi_i \sqcup \Phi_i \iota$$

for all $i = 1, \ldots, t$. Note that $|\Phi| = g$. It is well known that

$$\boldsymbol{X}^\star(T^{\mathrm{der}}) = \bigoplus_{\phi \in \Phi} \mathbb{Z}\epsilon_\phi$$

for some basis $\{\epsilon_\phi\}_{\phi\in\Phi}$. Then Γ permutes $\{\pm\epsilon_\phi\}_{\phi\in\Phi}$ by

$$\sigma\epsilon_\phi = \begin{cases} \epsilon_\psi & \text{if } \sigma\phi = \psi \in \Phi, \\ -\epsilon_\psi & \text{if } \sigma\phi = \psi\iota \in \Phi\iota, \end{cases} \qquad \phi, \psi \in \Phi. \tag{A-10}$$

The inclusion $\mu_2 \hookrightarrow T^{\mathrm{der}}$ corresponds to the map

$$X^\star(T^{\mathrm{der}}) \twoheadrightarrow X^\star(\mu_2) = \mathbb{Z}/2\mathbb{Z}, \qquad \sum_{\phi\in\Phi} x_\phi\epsilon_\phi \mapsto \sum_{\phi\in\Phi} x_\phi \mod 2.$$

Write $X^\star(\mathbb{G}_m) = \mathbb{Z}\eta$, where η is the standard generator. Applying Cartier duality to the exact sequence

$$1 \to \mu_2 \to \mathbb{G}_m \times T^{\mathrm{der}} \to T \to 1,$$

we obtain

$$X^\star(T) = \left\{ t\eta + \sum_{\phi\in\Phi} x_\phi\epsilon_\phi : t + \sum_{\phi} x_\phi \in 2\mathbb{Z} \right\} \subset X^\star(\mathbb{G}_m) \oplus X^\star(T^{\mathrm{der}})$$

and a basis of $X^\star(T)$,

$$\{2\eta\} \sqcup \{\eta + \epsilon_\phi : \phi \in \Phi\}.$$

The element 2η is surely Γ-invariant. On the other hand, for $\phi, \psi \in \Phi$ and $\sigma \in \Gamma$, we derive from (A-10) that

$$\sigma(\eta + \epsilon_\phi) = \begin{cases} \eta + \epsilon_\phi & \text{if } \sigma\phi = \psi \in \Phi, \\ 2\eta - (\eta + \epsilon_\psi) & \text{if } \sigma\phi = \psi\iota \in \Phi\iota. \end{cases} \tag{A-11}$$

The Γ-action on $\widehat{T} := X^\star(T) \otimes \mathbb{C}^\times \cong \mathbb{C}^\times \times (\mathbb{C}^\times)^\Phi$ is thus

$$\sigma \cdot (z, \overbrace{1, \dots, \underbrace{w}_{\phi}, \dots, 1}^{\Phi}) = \begin{cases} (z, 1, \dots, \underbrace{w}_{\psi}, \dots, 1) & \text{if } \sigma\phi = \psi \in \Phi, \\ (zw, 1, \dots, \underbrace{w^{-1}}_{\psi}, \dots, 1) & \text{if } \sigma\phi = \psi\iota \in \Phi\iota. \end{cases}$$

Recall from the description of $X^\star(T^{\mathrm{der}})$ that $\widehat{T^{\mathrm{der}}}$ can be identified with $(\mathbb{C}^\times)^\Phi$. Note that $(\widehat{T^{\mathrm{der}}})^\Gamma$ is finite since T^{der} is anisotropic.

Lemma A.10. *There is a canonical isomorphism $(\widehat{T^{\mathrm{der}}})^\Gamma \cong \mu_2^t$ characterized as follows. For $(a_i)_{i=1}^t \in \mu_2^t$, the corresponding $\hat{t} = (\hat{t}_\phi)_{\phi\in\Phi} \in (\widehat{T^{\mathrm{der}}})^\Gamma$ is specified by*

$$\phi \in \Phi_i \implies \hat{t}_\phi = a_i \quad \textit{for all } 1 \le i \le t.$$

Proof. Shapiro's lemma reduces the computation of $(\widehat{T^{\mathrm{der}}})^\Gamma$ or $H^1(\Gamma, X^\star(T))$ to the easy case $t = 1$ over the base field K^+. □

By dualizing $1 \to T^{\mathrm{der}} \to T \to \mathbb{G}_m \to 1$ into $1 \to \mathbb{C}^\times \to \widehat{T} \to \widehat{T_0} \to 1$, then taking Γ-invariants, we obtain the exact sequence

$$1 \to \mathbb{C}^\times \to \widehat{T}^\Gamma \to (\widehat{T^{\mathrm{der}}})^\Gamma.$$

It induces

$$\pi_0(\widehat{T}^\Gamma) = \widehat{T}^\Gamma/\mathbb{C}^\times \hookrightarrow (\widehat{T^{\mathrm{der}}})^\Gamma = \pi_0((\widehat{T^{\mathrm{der}}})^\Gamma).$$

For each $\sigma \in \Gamma$, set $\Phi(\sigma) := \{\phi \in \Phi : \sigma\phi \notin \Phi\}$. For each i, set $\Phi_i(\sigma) := \Phi(\sigma) \cap \Phi_i$.

Proposition A.11. *For any $(a_i)_i \in \mu_2^t$, the corresponding element $\hat{t} \in (\widehat{T^{\mathrm{der}}})^\Gamma$ belongs to the image of $\widehat{T}^\Gamma/\mathbb{C}^\times$ if and only if*

$$A(a_1, \dots, a_t; \sigma) := \sum_{\substack{1 \le i \le t \\ a_i = -1}} |\Phi_i(\sigma)| \in 2\mathbb{Z} \quad \textit{for all } \sigma \in \Gamma.$$

Proof. Identify $\widehat{T^{\mathrm{der}}}$ with $(\mathbb{C}^\times)^\Phi$. Identify $\widehat{T}$ with $\mathbb{C}^\times \times (\mathbb{C}^\times)^\Phi$ using the basis $\{2\eta\} \sqcup \{\eta + \epsilon_\phi : \phi \in \Phi\}$ of $X^\star(T)$; the homomorphism $\widehat{T} \to \widehat{T^{\mathrm{der}}}$ is simply the projection.

Note that $\hat{t}$ is the image of $(1, \hat{t}) \in \widehat{T}$. It comes from $\widehat{T}^\Gamma/\mathbb{C}^\times$ if and only if $(1, \hat{t})$ (or any other preimage) is Γ-invariant. For all $\sigma \in \Gamma$, Lemma A.10 and the description (A-11) lead to

$$\sigma \cdot (1, \hat{t}) = ((-1)^{A(a_1, \dots, a_t; \sigma)}, \hat{t}).$$

The assertion follows at once. □

To illustrate the use of Proposition A.11, we prove the following

Proposition A.12. *If $t = 1$ and g is odd, then $H^1(\Gamma, X^\star(T)) \cong \pi_0(\widehat{T}^\Gamma)$ is trivial.*

Proof. It suffices to show $A(-1, \dots, -1; c) = |\Phi(c)| \notin 2\mathbb{Z}$, where $c \in \Gamma$ is the complex conjugation. Indeed, $c\phi = \phi\iota$ for all $\phi \in \Phi$ by generalities on CM-fields, so $\Phi(c) = \Phi$ has g elements, which is odd. □

Note that K is not assumed to be Galois over $\mathbb{Q}$.

Kottwitz's theory also relates Tate–Shafarevich groups to similar objects attached to dual tori; see Section 4 of [Kottwitz 1984b]. Nevertheless, we are not yet able to determine the Tate–Shafarevich group of T by this approach in the non-Galois case.

Acknowledgements

Achter would like to thank Edgar Costa, Stefano Marseglia, David Roe and Christelle Vincent for sharing preliminary data on isogeny classes. Altuğ would like to thank Sungmun Cho and Jacob Tsimerman for various discussions. Gordon is grateful to William Casselman, Sug-Woo Shin, and Nicolas Templier for several helpful discussions, and to Cornell University, the Association for Women in Mathematics, and the Michler family, as part of her work was carried out during a semester at Cornell University enabled by the Ruth I. Michler memorial prize. Our collaboration also benefited from NSERC funding, and our paper benefited from the referee's careful reading. Altuğ was partially supported by the NSF (DMS1702176) and a startup grant from Boston University, and Achter was partially supported by grants from the NSA (H98230-14-1-0161, H98230-15-1-0247 and H98230-16-1-0046).

References

[Achter and Gordon 2017] J. D. Achter and J. Gordon, "Elliptic curves, random matrices and orbital integrals", *Pacific J. Math.* **286**:1 (2017), 1–24. With an appendix by S. A. Altuğ. MR Zbl

[Achter and Williams 2015] J. Achter and C. Williams, "Local heuristics and an exact formula for abelian surfaces over finite fields", *Canad. Math. Bull.* **58**:4 (2015), 673–691. MR Zbl

[Achter and Wong 2013] J. D. Achter and S. Wong, "Quotients of elliptic curves over finite fields", *Int. J. Number Theory* **9**:6 (2013), 1395–1412. MR Zbl

[Bitan 2011] R. A. Bitan, "The discriminant of an algebraic torus", *J. Number Theory* **131**:9 (2011), 1657–1671. MR Zbl

[Bourbaki 1985] N. Bourbaki, *Éléments de mathématique*: *Algèbre commutative*, *Chapitres* 5 *à* 7, Masson, Paris, 1985. MR Zbl

[Clozel 1990] L. Clozel, "The fundamental lemma for stable base change", *Duke Math. J.* **61**:1 (1990), 255–302. MR Zbl

[Conrad 2014] B. Conrad, "Reductive group schemes", pp. 93–444 in *Autour des schémas en groupes*, *I*, Panor. Synthèses **42/43**, Soc. Math. France, Paris, 2014. MR Zbl

[Cortella 1997] A. Cortella, "The Hasse principle for the similarities of bilinear forms", *Algebra i Analiz* **9**:4 (1997), 98–118. MR Zbl

[Deuring 1941] M. Deuring, "Die Typen der Multiplikatorenringe elliptischer Funktionenkörper", *Abh. Math. Sem. Hansischen Univ.* **14** (1941), 197–272. MR Zbl

[Frenkel et al. 2010] E. Frenkel, R. Langlands, and B. C. Ngô, "Formule des traces et fonctorialité: le début d'un programme", *Ann. Sci. Math. Québec* **34**:2 (2010), 199–243. MR Zbl

[Gan and Yu 2000] W. T. Gan and J.-K. Yu, "Group schemes and local densities", *Duke Math. J.* **105**:3 (2000), 497–524. MR Zbl

[Gekeler 2003] E.-U. Gekeler, "Frobenius distributions of elliptic curves over finite prime fields", *Int. Math. Res. Not.* **2003**:37 (2003), 1999–2018. MR Zbl

[Gerhard and Williams 2019] J. Gerhard and C. Williams, "Local heuristics and an exact formula for abelian varieties of odd prime dimension over finite fields", *New York J. Math.* **25** (2019), 123–144. MR Zbl

[Gille and Szamuely 2006] P. Gille and T. Szamuely, *Central simple algebras and Galois cohomology*, Cambridge Studies in Advanced Mathematics **101**, Cambridge Univ. Press, 2006. MR Zbl

[Gordon 2022] J. Gordon, "Orbital integrals and normalization of measures", preprint, 2022. To appear in proceedings of *On the Langlands program*: *endoscopy and beyond*. arXiv 2205.02391

[Gross 1997] B. H. Gross, "On the motive of a reductive group", *Invent. Math.* **130**:2 (1997), 287–313. MR Zbl

[Gross 1998] B. H. Gross, "On the Satake isomorphism", pp. 223–237 in *Galois representations in arithmetic algebraic geometry* (Durham, 1996), edited by A. J. Scholl and R. L. Taylor, London Math. Soc. Lecture Note Ser. **254**, Cambridge Univ. Press, 1998. MR Zbl

[Gross and Gan 1999] B. H. Gross and W. T. Gan, "Haar measure and the Artin conductor", *Trans. Amer. Math. Soc.* **351**:4 (1999), 1691–1704. MR Zbl

[Guo et al. 2022] J.-W. Guo, N.-H. Sheu, and C.-F. Yu, "Class numbers of CM algebraic tori, CM abelian varieties and components of unitary Shimura varieties", *Nagoya Math. J.* **245** (2022), 74–99. MR Zbl

[Howe 2022] E. W. Howe, "Variations in the distribution of principally polarized abelian varieties among isogeny classes", *Ann. H. Lebesgue* **5** (2022), 677–702. MR Zbl

[Kisin 2017] M. Kisin, "mod p points on Shimura varieties of abelian type", *J. Amer. Math. Soc.* **30**:3 (2017), 819–914. MR Zbl

[Kottwitz 1982] R. E. Kottwitz, "Rational conjugacy classes in reductive groups", *Duke Math. J.* **49**:4 (1982), 785–806. MR Zbl

[Kottwitz 1984a] R. E. Kottwitz, "Shimura varieties and twisted orbital integrals", *Math. Ann.* **269**:3 (1984), 287–300. MR Zbl

[Kottwitz 1984b] R. E. Kottwitz, "Stable trace formula: cuspidal tempered terms", *Duke Math. J.* **51**:3 (1984), 611–650. MR Zbl

[Kottwitz 1990] R. E. Kottwitz, "Shimura varieties and λ-adic representations", pp. 161–209 in *Automorphic forms*, *Shimura varieties*, *and L-functions*, *I* (Ann Arbor, MI, 1988), edited by L. Clozel and J. S. Milne, Perspect. Math. **10**, Academic Press, Boston, 1990. MR Zbl

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

[Kottwitz 2005] R. E. Kottwitz, "Harmonic analysis on reductive p-adic groups and Lie algebras", pp. 393–522 in *Harmonic analysis*, *the trace formula*, *and Shimura varieties*, edited by L. Clozel and J. S. Milne, Clay Math. Proc. **4**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Langlands 1973] R. P. Langlands, "Modular forms and ℓ-adic representations", pp. 361–500 in *Modular functions of one variable*, *II* (Proc. Internat. Summer School, Univ. Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, 1973. MR Zbl

[Liang et al. 2021] P.-X. Liang, Y. Oki, H.-Y. Yang, and C.-F. Yu, "On Tamagawa numbers of CM tori", preprint, 2021. arXiv 2109.04121

[Marseglia 2021] S. Marseglia, "Computing square-free polarized abelian varieties over finite fields", *Math. Comp.* **90**:328 (2021), 953–971. MR Zbl

[Oesterlé 1982] J. Oesterlé, "Réduction modulo p^n des sous-ensembles analytiques fermés de $\mathbb{Z}_p^N$", *Invent. Math.* **66**:2 (1982), 325–341. MR Zbl

[Ono 1961] T. Ono, "Arithmetic of algebraic tori", *Ann. of Math.* (2) **74** (1961), 101–139. MR Zbl

[Ono 1963] T. Ono, "On the Tamagawa number of algebraic tori", *Ann. of Math.* (2) **78** (1963), 47–73. MR Zbl

[Platonov and Rapinchuk 1994] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics **139**, Academic Press, Boston, 1994. MR Zbl

[Rotman 1995] J. J. Rotman, *An introduction to the theory of groups*, 4th ed., Graduate Texts in Mathematics **148**, Springer, 1995. MR Zbl

[Rüd 2022a] T. Rüd, *Tamagawa numbers of symplectic algebraic tori, orbital integrals, and mass formulae for isogeny class of abelian varieties over finite fields*, Ph.D. thesis, University of British Columbia, 2022, available at https://open.library.ubc.ca/collections/ubctheses/24/items/1.0416617.

[Rüd 2022b] T. Rüd, "Explicit Tamagawa numbers for certain algebraic tori over number fields", *Math. Comp.* **91**:338 (2022), 2867–2904. MR Zbl

[Serre 1981] J.-P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Études Sci. Publ. Math.* **54** (1981), 323–401. MR Zbl

[Tate 1966] J. Tate, "Endomorphisms of abelian varieties over finite fields", *Invent. Math.* **2** (1966), 134–144. MR Zbl

[Veys 1992] W. Veys, "Reduction modulo p^n of p-adic subanalytic sets", *Math. Proc. Cambridge Philos. Soc.* **112**:3 (1992), 483–486. MR Zbl

[Voskresenskiĭ 1998] V. E. Voskresenskiĭ, *Algebraic groups and their birational invariants*, Translations of Mathematical Monographs **179**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Waterhouse 1969] W. C. Waterhouse, "Abelian varieties over finite fields", *Ann. Sci. École Norm. Sup.* (4) **2** (1969), 521–560. MR Zbl

[Weil 1982] A. Weil, *Adeles and algebraic groups*, Progress in Mathematics **23**, Birkhäuser, Boston, 1982. MR Zbl

[Xue and Yu 2021] J. W. Xue and C. F. Yu, "On counting certain abelian varieties over finite fields", *Acta Math. Sin.* (*Engl. Ser.*) **37**:1 (2021), 205–228. MR Zbl

[Yu 2012] C.-F. Yu, "Superspecial abelian varieties over finite prime fields", *J. Pure Appl. Algebra* **216**:6 (2012), 1418–1427. MR Zbl

Communicated by Kiran S. Kedlaya
Received 2019-07-31 Revised 2022-05-02 Accepted 2022-06-10

j.achter@colostate.edu — *Department of Mathematics, Colorado State University, Fort Collins, CO, United States*

ali.altug@gmail.com — *New York, NY, United States*

l.e.garcia@ucl.ac.uk — *Department of Mathematics, University College London, United Kingdom*

gor@math.ubc.ca — *Department of Mathematics, University of British Columbia, Vancouver, BC, Canada*

wwli@pku.edu.cn — *Beijing International Center for Mathematical Research, Peking University, Beijing, China*

tompa.rud@gmail.com — *Massachusetts Institute of Technology, Cambridge, MA, United States*

ALGEBRA AND NUMBER THEORY 17:7 (2023)
https://doi.org/10.2140/ant.2023.17.1281

The log product formula

Leo Herr

Let V, W be a pair of smooth varieties. We want to compare curve counts on $V \times W$ with those on V and W. The product formula in Gromov–Witten theory compares the virtual fundamental classes of stable maps to a product $\overline{M}_{g,n}(V \times W)$ to the product of stable maps $\overline{M}_{g,n}(V) \times \overline{M}_{g,n}(W)$. We prove the analogous theorem for *log stable maps* to log smooth varieties V, W.

This extends results of Y.P. Lee and F. Qu, who introduced this formula after K. Behrend. We introduce "log normal cones" and "log virtual fundamental classes," as well as modified versions of standard intersection-theoretic machinery adapted to log geometry.

0. Introduction

The log product formula. The purpose of the present paper is to prove the "product formula" for log Gromov–Witten invariants. We assume the reader is familiar with log geometry at the level of [Ogus 2018].

Let V, W be log smooth, quasiprojective log schemes. The moduli stack of log stable maps [Gross and Siebert 2013; Chen 2014; Abramovich and Chen 2014] $\mathscr{M}^{\ell}_{g,n}(V)$ parametrizes families of fs log smooth curves $C \to S$ with a stable map $C \to V$ of log schemes. These coincide with ordinary stable maps if V has trivial log structure $M_V \simeq \mathcal{O}^*_V$.

Let Q be the fiber product

$$\mathscr{M}^{\ell}_{g,n}(V) \times^{\ell}_{\overline{M}_{g,n}} \mathscr{M}^{\ell}_{g,n}(W)$$

in the category of *fs* (fine and saturated) log algebraic stacks [Ogus 2018, Corollary III.2.1.6], with maps

$$\mathscr{M}^{\ell}_{g,n}(V \times W) \xrightarrow{h} Q \xrightarrow{\widetilde{\Delta}} \mathscr{M}^{\ell}_{g,n}(V) \times \mathscr{M}^{\ell}_{g,n}(W).$$

We define log virtual fundamental class in Definition 3.1. One can endow Q with a log virtual fundamental class in two ways: pushing forward that of $\mathscr{M}^{\ell}_{g,n}(V \times W)$ or pulling back that of $\mathscr{M}^{\ell}_{g,n}(V) \times \mathscr{M}^{\ell}_{g,n}(W)$. The product formula equates these.

Theorem 0.1 (the "log Gromov–Witten product formula"). *The two log virtual fundamental classes are equal in the Chow group A_*Q:*

$$h_*[\mathscr{M}^{\ell}_{g,n}(V \times W)]^{\ell vir} = \Delta^![\mathscr{M}^{\ell}_{g,n}(V) \times \mathscr{M}^{\ell}_{g,n}(W)]^{\ell vir}.$$

MSC2020: 14A21, 14C17, 14N35.
Keywords: logarithmic geometry, intersection theory, algebraic geometry, algebraic stacks.

The symbol $\Delta^!$ refers to the log Gysin map of Definition 3.1.

The original, nonlog product formula was established by M. Kontsevich and Yu. Manin in genus zero [Kontsevich and Manin 1996]. It was extended to arbitrary genus by Behrend [1999].

Theorem 0.1 was formulated using ordinary virtual fundamental classes by Lee and Qu [2018] and proved under the assumption that one of V or W has trivial log structure. Like their work and the work of Behrend [1999] before it, our proof centers on this cartesian diagram (Situation 5.5):

$$\begin{tikzcd}
\mathscr{M}^{\ell}_{g,n}(V \times W) \arrow[r, "h"] \arrow[d, "c"] \arrow[dr, phantom, "\ulcorner\ell"] & Q \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \mathscr{M}^{\ell}_{g,n}(V) \times \mathscr{M}^{\ell}_{g,n}(W) \arrow[d, "a"] \\
\mathfrak{D} \arrow[r, "l"] & Q' \arrow[r, "\phi"] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \mathfrak{M}_{g,n} \times \mathfrak{M}_{g,n} \arrow[d, "s\times s"] \\
& \overline{M}_{g,n} \arrow[r, "\Delta"] & \overline{M}_{g,n} \times \overline{M}_{g,n}
\end{tikzcd}$$

One applies Costello's formula [2006, Theorem 5.0.1] and commutativity of the Gysin map to this diagram to compare virtual fundamental classes.

In the log setting, one requires this diagram to be cartesian in the 2-category of *fs log algebraic stacks* in order to preserve modular interpretations. The assumption of [Lee and Qu 2018] that V or W have trivial log structure ensures that these squares are *also* cartesian as underlying algebraic stacks.

These fs pullback squares in question likely aren't cartesian on underlying algebraic stacks. Therefore, none of the standard machinery of ordinary Gysin maps and normal cones is valid. This quandary forced us to prove the log analogues of Costello's formula and commutativity for our "log Gysin map." With these modifications, the original proof of K. Behrend essentially still works. We pause to comment on the new technology.

Log normal cones. The *log normal cone* $C^{\ell}_{X/Y} = C_{X/\mathcal{L}Y}$ of a map $f : X \to Y$ of log algebraic stacks is the central object of the present paper. Every log map factors as the composition of a strict and an étale map $X \to \mathcal{L}Y \to Y$, so the cone is determined by two properties:

- It agrees with the ordinary normal cone for strict maps.
- If one can factor f as $X \to Y' \to Y$ with $Y' \to Y$ log étale, the cones are canonically isomorphic:

$$C^{\ell}_{X/Y} \simeq C^{\ell}_{X/Y'}.$$

This object becomes simpler in the presence of charts. Locally, we may assume the map $X \to Y$ has a chart given by a map of Artin cones $\mathscr{A}_P \to \mathscr{A}_Q$. The map $\mathscr{A}_P \to \mathscr{A}_Q$ is log étale, so we can base change across it to get a strict map without altering the log normal cone.

Because this method can lead to radical alterations of the target Y, we recall another strategy that we learned from [Ito et al. 2020, Proposition 2.3.12]. For ordinary schemes, one locally factors a map as a closed immersion composed with a smooth map to get a presentation for the normal cone [Behrend and

Fantechi 1997]. We obtain a similar local factorization (Construction 1.1) into a *strict* closed immersion composed with a log smooth map, and the same presentation exists for the log normal cone.

The above is made more precise in Remark 2.7. The charts and factorizations these techniques require are only locally possible, so we need to know how log normal cones change after étale localization. We encounter a well-known subtlety noticed by W. Bauer [Olsson 2005, Section 7]: The log normal cone isn't invariant under base-changes by log étale maps (Remark 2.13). Our workaround is somewhat different from that of Olsson. These results are at the service of log intersection theory, and we outline a standard package of log virtual fundamental classes and log Gysin maps.

Pushforward and Gysin pullback. The proof of the product formula needs two ingredients: commutativity of Gysin maps and compatibility of pushforward with Gysin maps. The commutativity of Gysin maps readily generalizes to the log setting in Theorem 3.12; on the other hand, compatibility with pushforward simply fails!

Nevertheless, the original proof of the product formula depends on a weak form of this compatibility first introduced by K. Costello [2006, Theorem 5.0.1]. This theorem is false as stated due to a missing properness hypothesis. We fix and generalize the statement in [Herr and Wise 2022] and offer a log generalization in Section 4.

We obtain another partial result towards compatibility of pushforward and Gysin pullback. For a log blowup $p : \widehat{X} \to X$ with a log smoothness assumption, we show $p_*[\widehat{X}]^{\ell vir} = [X]^{\ell vir}$ in Theorem 3.10. The alternative approach of Barrott [2018] may extend our results by modifying the notions of dimension, degree, pushforward, Chow groups, etc. in the log setting. See also [Ranganathan 2022] for an insightful approach to log Chow groups.

D. Ranganathan obtained a version of the log product formula contemporaneously using an explicit blowup instead of our log virtual fundamental class machinery [Ranganathan 2019]. We hope the technology and the strategy of reducing statements about log normal cones to the strict, ordinary case will be of interest.

Context and motivation. A pair (X, D) of a smooth divisor on a smooth variety is an example of a log smooth target V. Jun Li [2001] defined relative stable maps to such a pair (X, D). We instead use the log stable maps $\mathscr{M}^{\ell}_{g,n}(V)$ of Gross and Siebert [2013], and Chen [2014] and Abramovich and Chen [2014].[1] Even if one starts with V and W smooth pairs, their product $V \times W$ is a log smooth log scheme and likely not a smooth pair.

Gross and Siebert define the virtual fundamental class of $\mathscr{M}^{\ell}_{g,n}(V)$ relative to a log variant $\mathcal{L}\mathfrak{M}_{g,n}$ of the stack of prestable curves $\mathfrak{M}_{g,n}$. We take this as a definition of the log virtual fundamental class and related log normal cone. The log virtual fundamental class is then *invariant* under log modifications of the target V. The classes defined in this paper live in ordinary Chow groups $A_*(\cdot)$ but can be refined [Herr et al. 2023] to both large and small log Chow groups [Holmes et al. 2019] or to K theory [Chou et al. 2020].

[1] The forthcoming [Herr et al. 2023] compares relative and log stable maps.

One might try to prove Theorem 0.1 using the ordinary Gysin map $\Delta^!$ instead of the log version. This works if V or W has trivial log structure [Lee and Qu 2018] but is false in general. See [Chou et al. 2020] for counterexamples to the ordinary Gysin map $\Delta^!$ version. Our log Gysin map is necessary to prove Theorem 0.1 partially because it produces classes on the fs fiber product Q instead of the fiber product of underlying schemes.

These log Gysin maps may be of interest wherever the fs fiber product arises in enumerative geometry. For example, fs pullback squares abound in the punctured log stable maps of [Abramovich et al. 2020]. The fs fiber products in [Nabijou and Ranganathan 2022, Section 2.1] are reduced to ordinary fiber products by weak semistable reduction to use ordinary intersection theory. The multiplicativity found in [Holmes et al. 2019] is an example of a log intersection product.

Log Chow groups are still under construction. The log Chow group of X can be defined as a limit or colimit over the Chow groups of log modifications of X [Holmes et al. 2019; Barrott 2018]. The log virtual fundamental class can be refined to lie in log Chow [Chou et al. 2020; Herr et al. 2023]. Rather than making sense of pairing with ψ classes, we simply prove the expected equality of virtual fundamental classes in ordinary Chow here.

See [Molcho and Ranganathan 2021, Section 1.2] for a down-to-earth log intersection product related to our $\Delta^!$.

Our log product formula may compute the log Gromov–Witten invariants of toric varieties, as shown to the author independently by J. Wise and D. Ranganathan. Any pair of toric varieties Y_1, Y_2 of the same dimension are related by a third which is a log blowup of each: $Y_1 \leftarrow \tilde{Y} \to Y_2$. Log virtual fundamental classes and Gromov–Witten invariants are invariant under log blowups [Abramovich and Wise 2018], Theorem 3.10. The log Gromov–Witten invariants of *all* toric varieties of dimension n are essentially the same, and one can compute just one example $(\mathbb{P}^1)^n$.

Conventions. • We *only consider fs log structures*. We therefore use $\mathcal{L}$, $\mathcal{L}Y$ to refer to Olsson's stacks $\mathscr{T}or$, $\mathscr{T}orY$.

• We work over the complex numbers $\mathbb{C}$.

• We adhere to the convention of [Olsson 2003] regarding the use of the term "algebraic stack": we mean a stack in the sense of [Laumon and Moret-Bailly 2000, 3.1] such that

- the diagonal is representable and of finite presentation, and
- there exists a surjective, smooth morphism to it from a scheme.

We do not require the diagonal morphism to be separated.

• By "log algebraic stack," we mean an algebraic stack with a map to $\mathcal{L}$. Maps between them need not lie over $\mathcal{L}$.

• The name "DM stack" means Deligne–Mumford stack and a morphism $f : X \to Y$ of algebraic stacks is (of) "DM-type" or simply "DM" if every Y-scheme $T \to Y$ pulls back to a DM stack $T \times_{f,Y} X$ [Manolache 2012].

- The word "cone" in "log normal cone" refers to a cone stack in the sense of [Behrend and Fantechi 1997].
- Let P be a sharp fs monoid. Write

$$\mathscr{A}_P = [\operatorname{Spec} \mathbb{C}[P] / \operatorname{Spec} \mathbb{C}[P^{gp}]]$$

for the stack quotient in the étale topology endowed with its natural log structure [Abramovich et al. 2017; Cavalieri et al. 2020; Olsson 2003]. Beware that some of these sources first take the dual monoid. This log stack has a notable functor of points for fs log schemes

$$\operatorname{Hom}_{fs}(T, \mathscr{A}_P) = \operatorname{Hom}_{mon}(P, \Gamma(\overline{M}_T)).$$

In particular,

$$\operatorname{Hom}_{fs}(\mathscr{A}_P, \mathscr{A}_Q) = \operatorname{Hom}_{mon}(Q, P).$$

We write $\mathscr{A}$ for $\mathscr{A}_{\mathbb{N}} = [\mathbb{A}^1/\mathbb{G}_m]$. Log algebraic stacks of this form are called "Artin cones." "Artin fans" are log algebraic stacks which admit a strict étale cover by Artin cones. The 2-category of Artin fans is equivalent to a category of "cone stacks" [Cavalieri et al. 2020, Theorem 6.11].

- The present paper concerns analogues of normal cones and pullbacks in the logarithmic category. We use the notation $\ulcorner$, $\times$, C for pullbacks and normal cones of ordinary stacks, and write $\ulcorner\ell$, $\times^{\ell}$, C^{ℓ} to distinguish the fs pullbacks and log normal cones. When they happen to coincide, we write $\not\ell$, $\ulcorner\not\ell$, $\times^{\not\ell}$, $C^{\not\ell}$ to emphasize this coincidence.
- Many of our citations could be made to original sources, often written by K. Kato, but we have opted for the book [Ogus 2018]. We have doubled references to Costello's formula [Costello 2006, Theorem 5.0.1; Herr and Wise 2022] where appropriate because the original formulation is incorrect.

1. Preliminaries and the log normal sheaf

The present paper originated with one central construction, which we learned from [Ito et al. 2020, Lemma 2.3.12].

Construction 1.1. The normal cone of a morphism $f : B \to A$ of finite type is constructed by choosing a factorization $B \to B[x_1, \dots, x_r] \twoheadrightarrow A$ inducing a closed immersion into affine r-space

$$\operatorname{Spec} A \hookrightarrow \mathbb{A}^r_B \to \operatorname{Spec} B.$$

The normal cone of f may then be expressed as the quotient of the ordinary normal cone of the closed immersion by the action of the tangent bundle of $\mathbb{A}^r_B \to \operatorname{Spec} B$.

Let $P \to A$ and $Q \to B$ be morphisms from fs monoids to the multiplicative monoids of rings ("prelog rings"). A commutative square

$$\begin{tikzcd} B \arrow[r, "f"] & A \\ Q \arrow[u] \arrow[r, "\theta"] & P \arrow[u] \end{tikzcd}$$

is a chart of a map between affine log schemes. Assume f is of finite type; θ automatically is by the fs assumption. We will obtain a factorization of the induced log schemes into a *strict closed immersion* followed by a *log smooth map*.

Start with a similar factorization

$$\begin{tikzcd}
B \arrow[r] & B[x_1, \dots, x_r, y_1, \dots, y_s] \arrow[r, two heads] & A \\
Q \arrow[u] \arrow[r] & Q_s \arrow[u] \arrow[r, two heads] & P \arrow[u]
\end{tikzcd}$$

with $Q_s = Q \oplus \mathbb{N}^s$ mapping to $B[x_1, \dots x_r, y_1, \dots y_s]$ by sending the generators of $\mathbb{N}^s$ to the algebra generators $y_1, \dots y_s$ and $Q_s \to P$ surjective. Define Q_s^θ via the cartesian product:

$$\begin{tikzcd}
Q_s \arrow[r, hook] \arrow[d, hook] & Q_s^\theta \arrow[r, two heads] \arrow[d, hook] \arrow[dr, phantom, "\ulcorner"] & P \arrow[d, hook] \\
Q_s^{gp} \arrow[r, equal] & Q_s^{gp} \arrow[r] & P^{gp}
\end{tikzcd}$$

By definition, $Q_s^\theta \to P$ is exact, and $Q_s \to Q_s^\theta$ is a "log modification:" an isomorphism on groupifications. Witness also that $Q_s^\theta \to P$ is surjective, so the characteristic monoid map $\overline{Q_s^\theta} \xrightarrow{\sim} \overline{P}$ is an isomorphism [Ogus 2018, Proposition I.4.2.1(5)] and $\mathscr{A}_P \to \mathscr{A}_{Q_s^\theta}$ is strict. Take Spec and Artin cones of monoids to obtain a diagram with strict vertical arrows:

$$\begin{tikzcd}
X \arrow[r, hook] \arrow[d] & X_\theta \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \mathbb{A}_Y^{r+s} \arrow[r] \arrow[d] & Y \arrow[d] \\
\mathscr{A}_P \arrow[r] & \mathscr{A}_{Q_s^\theta} \arrow[r] & \mathscr{A}_{Q_s} \arrow[r] & \mathscr{A}_Q
\end{tikzcd}$$

We've written $Y = \operatorname{Spec} B$, $X = \operatorname{Spec} A$ and introduced the fs pullback X_θ in the diagram. The top row expresses our original map Spec f as the composition of a strict closed immersion, a log modification, and a smooth and log smooth morphism. The log modification $\mathscr{A}_{Q_s^\theta} \to \mathscr{A}_{Q_s}$ and hence $X_\theta \to \mathbb{A}_Y^{r+s}$ may be expressed as a (strict) open immersion into a log blowup as in [Ogus 2018, Lemma II.1.8.2, Remark II.1.8.5]. Hence $X \subseteq X_\theta$ is a strict closed immersion and $X_\theta \to Y$ is log smooth.

Remark 1.2. Continue in the notation of Construction 1.1. If we began with a morphism of fs log rings with f and θ both surjective, we could omit $Q_s \to B[x_1, \dots, x_r, y_1, \dots y_s]$. In that case, we obtain a factorization

$$X \subseteq X_\theta \to Y$$

where $X_\theta \to Y$ is not only log smooth but log étale.

As in [Behrend and Fantechi 1997], we will present the log normal cone locally as $C^\ell_{X/Y} = [C_{X/X_\theta}/T^\ell_{X_\theta/Y}]$ using these factorizations. The difficulty is then piecing together the local descriptions and checking compatibility. In this sense, the heavy lifting has already been done for us by [Manolache 2012]. We spend the rest of this section collecting relevant properties of the log normal sheaf $N^{\ell}_{X/Y}$. When we define the log normal cone $C^{\ell}_{X/Y} \subseteq N^{\ell}_{X/Y}$, its important properties will be locally deduced from such factorizations.

Remark 1.3. An algebraic stack X is DM if and only if the map $X \to \operatorname{Spec} k$ to the base field is of DM-type. If $X \to Y$ is a morphism of DM type and Y admits a stratification by global quotients, then so does X [Manolache 2012, Remark 3.2]. A morphism $f : X \to Y$ of algebraic stacks is of DM type if and only if its diagonal $\Delta_{X/Y} : X \to X \times_Y X$ is unramified [Stacks 2005–, 06N3].

Lemma 1.4. *Let $f : X \to Y$ be a morphism of log algebraic stacks. If the map on underlying stacks is of DM-type, then the induced maps $\mathcal{L}X \to \mathcal{L}Y$ and $X \to \mathcal{L}Y$ are DM-type.*

Proof. The inclusion $X \subseteq \mathcal{L}X$ representing strict maps is open, so it suffices to show that $\mathcal{L}X \to \mathcal{L}Y$ is DM-type.

We will argue that the diagonal of $\mathcal{L}X \to \mathcal{L}Y$ is unramified [Stacks 2005–, 04YW]. The isomorphism $\mathcal{L}X \times_{\mathcal{L}Y} \mathcal{L}X \simeq \mathcal{L}(X \times^{\ell}_{Y} X)$ identifies the diagonal $\Delta_{\mathcal{L}X/\mathcal{L}Y}$ with the result of $\mathcal{L}$ applied to the fs diagonal

$$\Delta^{\ell}_{X/Y} : X \to X \times^{\ell}_{Y} X.$$

Any diagram

$$\begin{tikzcd} S_0 \arrow[r] \arrow[d, hook] & \mathcal{L}X \arrow[d, "\mathcal{L}\Delta^{\ell}_{X/Y}"] \\ S_0' \arrow[r] \arrow[ur, dashed, shift left] \arrow[ur, dashed, shift right] & \mathcal{L}(X \times^{\ell}_{Y} X) \end{tikzcd}$$

with $S_0 \subseteq S_0'$ a squarezero closed immersion of schemes is equivalent to a diagram

$$\begin{tikzcd} S \arrow[r] \arrow[d, hook] & X \arrow[d, "\Delta^{\ell}_{X/Y}"] \\ S' \arrow[r] \arrow[ur, dashed, shift left] \arrow[ur, dashed, shift right] & X \times^{\ell}_{Y} X \end{tikzcd}$$

with $S \subseteq S'$ an exact closed immersion of log schemes. Composing with the fsification map $X \times^{\ell}_{Y} X \to X \times_Y X$ sends this square to

$$\begin{tikzcd} S \arrow[r] \arrow[d, hook] & X \arrow[d, "\Delta^{\ell}_{X/Y}"] \\ S' \arrow[r] \arrow[ur, dashed, shift left] \arrow[ur, dashed, shift right] & X \times_{Y} X \end{tikzcd}$$

in which case the two dashed arrows have the same underlying scheme map because $X \to X \times_Y X$ is unramified by hypothesis. Then the maps on log structure must be the same as well, because

$$(M_X \oplus^{\ell}_{M_Y} M_X)|_{S'} \to (M_X)|_{S'}$$

is an epimorphism. $\square$

Recall the normal sheaf. If $X \subseteq Y$ is a closed embedding with ideal I, the normal sheaf is simply $N_{X/Y} = \operatorname{Spec}\operatorname{Sym}^* I/I^2$. If $X \to Y$ is smooth, then $N_{X/Y} = BT_{X/Y}$ is the classifying space for the tangent bundle. Behrend and Fantechi obtain a normal sheaf more generally by locally factoring $X \to Y$ into a closed immersion into affine space $X \subseteq \mathbb{A}^n_Y \to Y$. If $X \subseteq \mathbb{A}^n_Y$ can be chosen to be a regular closed embedding, the map $X \to Y$ is l.c.i. and the cotangent complex $\mathbb{L}_{X/Y}$ is perfect of amplitude in $[-1, 0]$.

Definition 1.5 (normal sheaf). Let $f : X \to Y$ be a DM type qcqs morphism of algebraic stacks with cotangent complex $\mathbb{L}_{X/Y}$ given by the system of truncations $\{\tau_{\geq -n}\pi^*\mathbb{L}_{X/Y}\}_n$ [Olsson 2007, Theorem 8.1]. The *normal sheaf* is the associated Picard stack [SGA 4_1 1972, XVIII.1.4]

$$h^1/h^0((\mathbb{L}_{X/Y,fl})^\vee)$$

as in [Behrend and Fantechi 1997, Section 2]. An *obstruction theory* for f is a fully faithful functor $N_{X/Y} \subseteq E$ into a vector bundle stack E over X.

Remark 1.6. The Picard stack h^1/h^0 only depends on the truncation $\tau_{\geq -1}$, so we don't need the entire system $\{\tau_{\geq -n}\pi^*\mathbb{L}_{X/Y}\}_n$. Moreover, we can bypass π and pull back directly to the big fppf site

$$X_{fl} \to X_{lis-\text{ét}} \xrightarrow{\pi} X_{\text{ét}}.$$

Factor the map f locally as $X \subseteq M \to Y$ as a closed immersion composed with a smooth map $M \to Y$. Writing $I = I_{X/M}$ for the ideal sheaf, we obtain a map

$$\mathbb{L}_{X/Y} \to [I/I^2 \to \Omega_{M/Y}]$$

which induces isomorphisms on the first two cohomology groups h^{-1}, h^0. This identifies their Picard stacks

$$N_{X/Y} \simeq [N_{X/M}/T_{M/Y}|_X].$$

The functor of points of the normal sheaf on a X-scheme S is given by the category of algebra extensions

$$N_{X/Y}(S) = \underline{\operatorname{Ext}}(\mathbb{L}_{X/Y}|_S, \mathcal{O}_S) = \left\{ \begin{tikzcd} & \mathcal{O}_Y|_S \arrow[d] \arrow[dr, bend left] & & \\ 0 \arrow[r] & \mathcal{O}_S \arrow[r] & \mathcal{A} \arrow[r] & \mathcal{O}_X|_S \arrow[r] & 0 \end{tikzcd} \quad \begin{array}{c} \text{a squarezero algebra} \\ \text{extension on ét}(T) \end{array} \right\}.$$

Definition 1.7 (log normal sheaf). Let $f : X \to Y$ be a DM type qcqs morphism between log algebraic stacks. Define the log normal sheaf $N^\ell_{X/Y} := N_{X/\mathcal{L}Y} = h^1/h^0(\mathbb{L}^{\ell}_{X/Y,fl}{}^\vee)$.

We will also have cause to consider $N_{\mathcal{L}X/\mathcal{L}Y}$, and $N_{\mathcal{L}X/\mathcal{L}Y}|_X = N^\ell_{X/Y}$.

Remark 1.8. Locally in X and Y, f factors as $X \subseteq M \to Y$ with $X \subseteq M$ a *strict* closed immersion and $M \to Y$ log smooth by Construction 1.1. We have a similar presentation

$$N^\ell_{X/Y} = [N_{X/M}/T^\ell_{M/Y}|_X].$$

One may alternately define $N^\ell_{X/Y}$ by gluing together these local presentations as in [Behrend and Fantechi 1997, Corollary 3.9]. One checks for $Y \to Z$ étale that the map on normal sheaves $N^\ell_{X/Y} \simeq N^\ell_{X/Z}$ is an isomorphism.

One may alternately use Gabber's notion of log cotangent complex $\mathbb{L}^G_{X/Y}$ because the two agree on truncations $\tau_{\geq -2}\mathbb{L}^\ell_{X/Y} \simeq \tau_{\geq -2}\mathbb{L}^\ell_{X/Y}$ [Olsson 2005, Theorem 8.27]. Gabber's version has the advantage that a distinguished triangle exists for *all* composable pairs of arrows $X \to Y \to Z$.

We claim the functor of points of $N^\ell_{X/Y}$ on an X-scheme S is given by squarezero extensions of algebras

$$\begin{tikzcd} & & \mathcal{O}_Y|_S \arrow[d] \arrow[dr, bend left] & & \\ 0 \arrow[r] & \mathcal{O}_S \arrow[r] & \mathcal{A} \arrow[r] & \mathcal{O}_X|_S \arrow[r] & 0 \end{tikzcd}$$

together with a "log structure" $M_{\mathcal{A}} \to \mathcal{A}$ making

$$\begin{tikzcd} \mathcal{A}^* \arrow[r] \arrow[d] & \mathcal{O}_X^*|_S \arrow[d] \\ M_{\mathcal{A}} \arrow[r] & M_X|_S \end{tikzcd}$$

a pushout. We avoid this perspective because it requires squarezero extensions and log structures on étale-locally ringed topoi $(\text{ét}(T), \mathcal{O}_X|_T)$, $(\text{ét}(T), \mathcal{A})$. The reader may notice a resemblance to "deformations of log structures" in [Illusie 1997] and to the classical notion of squarezero extensions along a map $X \to Y$ reprised in [Olsson 2005, Section 5].

Remark 1.9. The central object of this paper is a subcone stack $C^\ell_{X/Y} \subseteq N^\ell_{X/Y}$ introduced in the next section. This substack has no functor of points, as it is defined by a blowup; see [Khan and Rydh 2018] for a derived workaround. Most of our arguments about $C^\ell_{X/Y}$ go by way of the functor of points of $N^\ell_{X/Y}$, together with a pointwise argument to compare these substacks.

Functoriality of $N^\ell_{X/Y}$. To write down the functoriality of the log normal sheaf, we need to recall some of the machinery of log stacks found in [Olsson 2005].

We denote $\mathcal{L}^i := \mathcal{L}^{[i]}$, the stack of i-simplices of fs log structures. The j-th face map d_j sends

$$(M_0 \to M_1 \to \cdots \to M_{i+1}) \mapsto \begin{cases} (M_1 \to M_2 \to \cdots \to M_{i+1}) & \text{if } j = 0, \\ (M_0 \to \cdots \to M_{j-1} \to M_{j+1} \cdots \to M_{i+1}) & \text{if } j \neq 0, i+1, \\ (M_0 \to \cdots \to M_i) & \text{if } j = i+1. \end{cases}$$

We write $s, t : \mathcal{L}^1 \to \mathcal{L}^0 = \mathcal{L}$ for the "source" d_1 and "target" d_0 maps, respectively. We have an isomorphism $\mathcal{L}^i = \mathcal{L}^1 \times_{t,\mathcal{L},s} \mathcal{L}^1 \times_{t,\mathcal{L},s} \cdots \times_{t,\mathcal{L},s} \mathcal{L}^1$ (i factors).

Endow $\mathcal{L}^i$ with the final tautological log structure, M_{i+1} in the above. All the face maps d_j are strict except $j = i+1$.

We continue [Olsson 2005] to use "$\square$" to denote the category with these objects, arrows, and relations:

$$\begin{tikzcd} 0 \arrow[r] \arrow[d] \arrow[dr] & 1 \arrow[d] \arrow[dl, phantom, "\circ"] \\ 2 \arrow[r] \arrow[ur, phantom, "\circ"] & 3 \end{tikzcd}$$

We adopt pictorial mnemonics for fully faithful morphisms of these finite diagrams: $\boxbslash$ means the functor $[2] \subseteq \square$ avoiding 2, etc.

A commutative square

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ Y' \arrow[r] & Y \end{tikzcd}$$

should lead to a map $N^{\ell}_{X'/Y'} \to N^{\ell}_{X/Y}$. This square does *not* induce a commutative diagram

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\times"] & X \arrow[d] \\ \mathcal{L}Y' \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

so the naive strategy to get a map $N^{\ell}_{X'/Y'} \to N^{\ell}_{X/Y}$ doesn't work. To get around this, Olsson introduces another stack.

Definition 1.10 (compare [Olsson 2005, Lemma 3.12]). Define $\mathcal{V} := \mathcal{L}^1 \times^{\ell}_{t,\mathcal{L},t} \mathcal{L}^1$. Given a scheme T, the points of this stack are cocartesian squares of *fs* log structures:

$$\mathcal{V}(T) := \left\{ \begin{tikzcd} M_0 \arrow[r] \arrow[d] & M_1 \arrow[d] \\ M_2 \arrow[r] \arrow[ur, phantom, "\ell\lrcorner"] & M_3 \end{tikzcd} \right\}$$

This is the "fsification" of the ordinary pullback $\mathcal{L}^1 \times_{t,\mathcal{L},t} \mathcal{L}^1$, endowed with the non-fs pushout $M_1 \oplus^{mon}_{M_0} M_2$ of the universal log structures.

The natural embedding $\mathcal{V} \to \mathcal{L}^{\square}$ exhibits the squares which are cocartesian as an open substack, as we'll record in Lemma 1.12.

For a morphism $q : Y' \to Y$ of log algebraic stacks, we obtain relative variants

$$\mathcal{V}_q := \mathcal{V} \times_{\overline{\square}, \mathcal{L}^1} Y', \quad \mathcal{L}^{\square}_q := \mathcal{L}^{\square} \times_{\overline{\square}, \mathcal{L}^1} Y'.$$

The fs pullback here agrees with the ordinary one because $Y' \to \mathcal{L}^1$ is strict. The points of these stacks over some scheme T are squares

$$\begin{tikzcd} M_Y|_T \arrow[r] \arrow[d] & M_{Y'}|_T \arrow[d] \\ M_0 \arrow[r] & M_1 \end{tikzcd}$$

with those of $\mathcal{V}_q$ required to be cocartesian.

Lemma 1.11. *Let $\mathcal{L}^{arbfine}$ denote the stack of log structures which are fine but not necessarily saturated. The natural monomorphism*

$$\mathcal{L} \hookrightarrow \mathcal{L}^{arbfine}$$

is an open immersion.

Proof. Consider some scheme X and pullback diagram:

$$\begin{tikzcd} X^{fs} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{L} \arrow[d] \\ X \arrow[r] & \mathcal{L}^{arbfine} \end{tikzcd}$$

Then $X^{fs} \hookrightarrow X$ is a monomorphism, the locus where the stalks of M_X are saturated. After passing to an open cover of X, [Ogus 2018, Theorem II.2.5.4] provides us with a locally finite stratification $X = \bigsqcup_{\sigma \in \Sigma} X_\sigma$ where:

- For each $\sigma \in \Sigma$, $\overline{M}_X|_\sigma$ is constant.
- The cospecialization maps for $x \in \overline{\{\xi\}} \subseteq X$

  $$\overline{M}_x \to \overline{M}_\xi$$

  are localizations at faces.

The localization of a saturated monoid remains saturated [Ogus 2018, Remark I.1.4.5] and a monoid is saturated if and only if its characteristic monoid is [loc. cit., Proposition I.1.3.5]. We then have that $X^{fs} \subseteq X$ is locally a constructible subset which is closed under generalization, and hence open [Stacks 2005–, Tag 0542]. $\square$

We collect several results of [Olsson 2005] adapted to the fs setting.

Lemma 1.12 [Olsson 2005, Theorem 2.4, Proposition 2.11, Lemma 3.12]. *These statements remain true in the **fs** context*:

(1) *For any finite category Γ, the fibered category $\mathcal{L}^\Gamma$ of diagrams of fs log structures indexed by Γ is an algebraic stack.*

(2) *The simplicial face maps $d_j : \mathcal{L}^{i+1} \to \mathcal{L}^i$ are strict, étale, and DM-type for $j \leq i$.*

(3) *If $[1] \to \square$ avoids the initial object 0 (or), it induces a strict étale, DM-type morphism*

$$\mathcal{L}^\square \to \mathcal{L}^1.$$

(4) *If $[2] \to \square$ omits either 1 or 2 (or), it induces an étale, DM-type morphism*

$$\mathcal{L}^\square \to \mathcal{L}^2.$$

(5) *The map $\mathcal{V} \subseteq \mathcal{L}^\square$ is an open embedding.*

(6) *Given an fs pullback square*

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell", very near start] & X \arrow[d] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

the associated square of stacks

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & X \arrow[d] \\ \mathcal{V}_q \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

is a pullback.

Proof. Facts (1) through (4) are immediate by Lemma 1.11 and the analogous facts in [Olsson 2005]. The last two follow by the same arguments applied in the fs category. $\square$

Remark 1.13. Apply $\mathcal{L}$ once more to the map $\mathcal{L}Y \to Y$, one gets

$$d_1 : \mathcal{L}^2 Y \to \mathcal{L}Y, \quad (M_Y \to M_0 \to M_1) \mapsto (M_Y \to M_1).$$

The result is étale, so the original $d_1 : \mathcal{L}Y \to Y$ is log étale [Olsson 2003, Theorem 4.6(ii)]. The same reasoning implies $d_{i+1} : \mathcal{L}^{i+1} Y \to \mathcal{L}^i Y$ is log étale in general. In summary, all the face maps are log étale and all but $j = i+1$ are furthermore strict étale.

Remark 1.14. Given $q : Y' \to Y$ DM, the natural maps

$$\mathcal{V}_q \subseteq \mathcal{L}_q^{\square} \to \mathcal{L}Y'$$

are étale. The second map is the product of the étale map

$$\boxbslash^* : \mathcal{L}^{\square} \to \mathcal{L}^2$$

over $\mathcal{L}^1$ (via $\square$) with Y'.

Definition 1.15. Use Lemma 1.12(6) to turn one commutative square of DM maps into another:

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ Y' \arrow[r, "q"] & Y. \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ \mathcal{L}_q^{\square} \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

Maps of normal sheaves

$$\varphi : N^{\ell}_{X'/Y'} \simeq N_{X/\mathcal{L}_q^{\square}} \to N^{\ell}_{X/Y}$$

arise from Remark 1.14 and the second square. This coincides with the "natural map"

$$\mathbb{L}^{\ell}_{X/Y}|_{X'} \to \mathbb{L}^{\ell}_{X'/Y'}$$

of [Olsson 2005, (1.1.2)]. We call the composite φ *Olsson's morphism*.

Remark 1.16. In Definition 1.15, if the first square was an fs pullback square, the second factors:

$$\begin{tikzcd} X' \arrow[rr] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & & X \arrow[d] \\ \mathcal{V}_q \arrow[r, hook] & \mathcal{L}_q^{\square} \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

Since this square is a pullback, Olsson's morphism

$$\varphi : N^{\ell}_{X'/Y'} \simeq N_{X'/\mathcal{L}_q^{\square}} \simeq N_{X'/\mathcal{V}_q} \hookrightarrow N^{\ell}_{X/Y}|_{X'}$$

is then a closed immersion. This may be checked locally in X, Y, so we assume there is a factorization $X \subseteq X_\theta \to Y$ as in Construction 1.1. Take the fs pullback

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & X \arrow[d] \\ X'_\theta \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & X_\theta \arrow[d] \\ Y' \arrow[r] & Y \end{tikzcd}$$

along q to obtain a factorization $X' \subseteq X'_\theta \to Y'$. The strict case gives $N_{X'/X'_\theta} \subseteq N_{X/X_\theta}|_{X'}$ and pullback identifies the tangent spaces $T^{\ell}_{X'_\theta/Y'} \simeq T^{\ell}_{X_\theta/Y}|_{X'_\theta}$.

If $X \to Y$ is also log flat, φ is an isomorphism [Olsson 2005, (1.1(iv))]. This is *not* true if q is log flat, as seen in Remark 2.13. See Lemmas 2.14, 2.15 for the strict case.

Remark 1.17. A commutative square of DM maps may be factored:

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ \mathcal{L}_q^{\square} \arrow[r] \arrow[d] & \mathcal{L}Y \arrow[d] \\ Y' \arrow[r] & Y \end{tikzcd} \tag{1}$$

This induces a commutative square of normal sheaves:

$$\begin{tikzcd} N^{\ell}_{X'/Y'} \simeq N_{X'/\mathcal{L}_q^{\square}} \arrow[r] \arrow[d] \arrow[dr, phantom, "\circ"] & N^{\ell}_{X/Y} \arrow[d] \\ N_{X'/Y'} \arrow[r] & N_{X/Y} \end{tikzcd} \tag{2}$$

The Olsson morphisms are thereby seen to be compatible with the ordinary functoriality of the normal sheaf via the forgetful maps $N^{\ell}_{X/Y} \to N_{X/Y}$.

Now suppose the original square (1) is an fs pullback:

- If q is strict, then (2) is cartesian. This is local in X, Y, so we assume we have a factorization as in Construction 1.1

$$X \subseteq X_\theta \to \mathbb{A}^{r+s}_Y \to Y$$

and take the *ordinary scheme-theoretic* pull back to obtain a similar factorization of $X' \to Y'$ using strictness of q. Since $X \subseteq \mathbb{A}^{r+s}_Y$ is a closed immersion, the statement reduces to the cartesian square of ordinary normal sheaves

$$\begin{tikzcd} N_{X'/X'_\theta} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & N_{X/X_\theta} \arrow[d] \\ N_{X'/\mathbb{A}^{r+s}_{Y'}} \arrow[r] & N_{X/\mathbb{A}^{r+s}_Y} \end{tikzcd}$$

which may be checked using the functor of points, for example.

- If instead f is strict, then $X' \to \mathcal{V}_q$ factors through Y', and the factorization

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & X \arrow[d] \\ Y' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & Y \arrow[d, hook] \\ \mathcal{V}_q \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

shows that the vertical arrows of (2) are isomorphisms and the Olsson morphism is the same as the ordinary functoriality of the normal sheaf.

Remark 1.18. Given a commutative square

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

of DM maps we can form two other commutative squares out of it:

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ \mathcal{L}^\square_q \arrow[r] & \mathcal{L}Y \end{tikzcd} \qquad \begin{tikzcd} X' \arrow[r] \arrow[d] & \mathcal{L}X \arrow[d] \\ \mathcal{L}Y' \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

They induce morphisms

$$N^\ell_{X'/Y'} \simeq N^\ell_{X'/\mathcal{L}^\square_q} \to N^\ell_{X/Y}|_{X'} \quad \text{and} \quad N^\ell_{X'/Y'} \to N_{\mathcal{L}X/\mathcal{L}Y}|_{X'}.$$

Form the diagram

$$\begin{tikzcd}
X' \arrow[r] \arrow[d] & \mathcal{L}X \arrow[r, "s"] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & X \arrow[d] \\
\mathcal{L}^{\square}_q \arrow[r] \arrow[d] \arrow[rr, bend right] & \mathcal{L}^2Y \arrow[r, "d_0"] & \mathcal{L}Y \\
\mathcal{L}Y' & &
\end{tikzcd}$$

to see that the two morphisms of normal sheaves are compatible:

$$N^{\ell}_{X'/Y'} \simeq N_{X'/\mathcal{L}^{\square}_q} \to N_{\mathcal{L}X/\mathcal{L}^2Y}|_{X'} \subseteq N^{\ell}_{X/Y}|_{X'}.$$

L. Barrott pointed out to the author that “Condition (T)” as phrased in [Olsson 2005, (1.5.1)] ensures étale locally that the square

$$\begin{tikzcd}
\mathcal{L}X \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & X \arrow[d] \\
\mathcal{L}^2Y \arrow[r] & \mathcal{L}Y
\end{tikzcd}$$

above is Tor-independent.

Lemma 1.19. *Suppose given a pair of commutative squares*

$$\begin{tikzcd}
X' \arrow[r] \arrow[d] & Y' \arrow[r] \arrow[d] & Z' \arrow[d] \\
X \arrow[r, "f"] & Y \arrow[r, "g"] & Z
\end{tikzcd}$$

of DM-type maps. The diagram

$$\begin{tikzcd}
& N^{\ell}_{Y'/Y} \arrow[dr] & \\
N^{\ell}_{X'/X} \arrow[ur] \arrow[rr] & & N^{\ell}_{Z'/Z}
\end{tikzcd}$$

commutes, where all the arrows are Olsson’s morphisms.

Proof. Introduce an algebraic X-stack $\mathcal{W}$, with functor of points:

$$\left\{ \begin{tikzcd} & \mathcal{W} \arrow[d] \\ T \arrow[ur, dashed] \arrow[r] & X \end{tikzcd} \right\} := \left\{ \begin{tikzcd} M_2 & M_1 \arrow[l] & M_0 \arrow[l] \\ M_X|_T \arrow[u] & M_Y|_T \arrow[l] \arrow[u] & M_Z|_T \arrow[l] \arrow[u] \end{tikzcd} \quad \begin{array}{c}\text{commutative diagrams of} \\ \text{fs log structures on } T\end{array} \right\}$$

In other words, $\mathcal{W} := (\mathcal{L}^{\square} \times_{\mathcal{L}^1} \mathcal{L}^{\square}) \times_{\mathcal{L}^2} X$.

All the triangles in this diagram commute by inspection:

$$\begin{tikzcd}
& N_{X/\mathcal{L}^{\square}_{g\circ f}} \arrow[rrd, bend left] & & \\
N^{\ell}_{X'/X} \arrow[ur, dash, bend left, "\sim"] \arrow[r] \arrow[d, dash, "\sim"] & N_{X/\mathcal{W}} \arrow[u] \arrow[rr] \arrow[dl] \arrow[dd] & & N^{\ell}_{Z'/Z} \\
N_{X'/\mathcal{L}^{\square}_{f}} \arrow[dr, bend right] & & & \\
& N_{Y/\mathcal{L}^{\square}_{g}} \arrow[uurr] \arrow[r, dash, "\sim"'] & N^{\ell}_{Y'/Y} \arrow[uur, bend right] &
\end{tikzcd}$$

Restricting the diagram to $N^{\ell}_{X'/X}$, $N^{\ell}_{Y'/Y}$, and $N^{\ell}_{Z'/Z}$, we get the result. $\square$

Proposition 1.20. *Given $X \xrightarrow{f} Y \xrightarrow{g} Z$ DM-type maps of log algebraic stacks, the Olsson morphisms yield a complex of stacks*

$$N^{\ell}_{X/Y} \to N^{\ell}_{X/Z} \to N^{\ell}_{Y/Z}|_X,$$

in that the composite factors through the vertex.

If h is smooth, $N^{\ell}_{Y/Z} = BT^{\ell}_{Y/Z}$ and rotating the triangle in the derived category yields an exact sequence of cone stacks:

$$T^{\ell}_{Y/Z}|_X \to N^{\ell}_{X/Y} \to N^{\ell}_{X/Z}.$$

Proof. The Olsson morphisms come about from the commutative diagram:

$$\begin{tikzcd}
X \arrow[r, equal] \arrow[d] & X \arrow[r] \arrow[d] & Y \arrow[d] \arrow[r, phantom, "\rightsquigarrow"] & N^{\ell}_{X/Y} \arrow[r] \arrow[d] & N^{\ell}_{Y/Y}|_X = X \arrow[d, "0"] \\
Y \arrow[r] & Z \arrow[r, equal] & Z & N^{\ell}_{X/Z} \arrow[r] & N^{\ell}_{Y/Z}|_X
\end{tikzcd}$$

Use Gabber's log cotangent complex as in Remark 1.8 and rotate to get a distinguished triangle

$$\mathbb{L}^{G}_{X/Z} \to \mathbb{L}^{G}_{X/Y} \to \mathbb{L}^{G}_{Y/Z}|_X[1] \xrightarrow{+1}.$$

Then $\mathbb{L}^{\ell}_{Y/Z} = \Omega^{\ell}_{Y/Z}[0]$ and $h^1/h^0(\mathbb{L}^{G}_{Y/Z}[1])|_X = h^1/h^0(\mathbb{L}^{\ell}_{Y/Z}[1])|_X = T^{\ell}_{Y/Z}|_X$. $\square$

Remark 1.21. Suppose given a (not necessarily commutative) finite diagram of cones. If the diagram induced by taking abelian hulls is commutative, so was the original.

2. Properties of the log normal cone

We are ready to define the log normal cone. We recall the essential properties of the ordinary normal cone; the rest of the section establishes analogous properties in the log context.

Remark 2.1. Consider a DM-type morphism $f : X \to Y$ of algebraic stacks. K. Behrend and B. Fantechi defined the (intrinsic) normal cone [Behrend and Fantechi 1997, Definition 3.10]

$$C_f = C_{X/Y} \subseteq N_{X/Y};$$

C. Manolache [2012, Definition 2.30] removed their assumptions of smooth Y and DM X. This cone has the following basic properties:

(1) A commutative diagram

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

yields a morphism of cones $\varphi : C_{X'/Y'} \to C_{X/Y} \times_X X'$:

- If the square was cartesian, φ is a closed embedding.
- If the square was cartesian and also f or q was flat, φ is an isomorphism.

(2) For a composite

$$X \xrightarrow{f} Y \xrightarrow{g} Z :$$

- If g is l.c.i., $C_{X/Y} = N_{X/Y}$ and the sequence

$$N_{X/Y} \to C_{X/Z} \to C_{Y/Z}|_X$$

of cone stacks is exact.
- If h is smooth, the sequence

$$T_{Y/Z}|_X \to C_{X/Y} \to C_{X/Z}$$

is exact.

(3) Obstruction theories and Gysin pullbacks are obtained by placing the cone in a vector bundle stack $C_{X/Y} \subseteq E$ via an isomorphism $A_*E \simeq A_*X$ called "intersecting with the zero section;" see [Manolache 2012, Section 3; Wise 2011, Proposition 3.6; Kresch 1999, Section 6.2].

Definition 2.2 (log intrinsic normal cone, Olsson morphisms). Let $f : X \to Y$ be a DM-type morphism of log algebraic stacks. We define the *log* (*intrinsic*) *normal cone*

$$C^{\ell}_{X/Y} := C_{X/\mathcal{L}Y} \subseteq N^{\ell}_{X/Y}$$

after [Gross and Siebert 2013]. Endow it with the log structure pulled back from X. Given a commutative square of log algebraic stacks and its partner:

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ Y' \arrow[r, "q"] & Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} X' \arrow[r] \arrow[d] & X \arrow[d] \\ \mathcal{L}_q^{\square} \arrow[r] & \mathcal{L}Y \end{tikzcd}$$

The latter induces

$$\varphi : C^{\ell}_{X'/Y'} \simeq C_{X'/\mathcal{L}_q^{\square}} \to C^{\ell}_{X/Y}.$$

This is again called the *Olsson morphism*.

Remark 2.3. The map $\mathcal{L}Y \to Y$ has a section $Y \subseteq \mathcal{L}Y$ which is an open immersion. This open immersion represents strict log maps to Y.

As a result, if $X \to Y$ is DM and strict, $C^{\ell}_{X/Y} = C_{X/Y}$ and $N^{\ell}_{X/Y} = N_{X/Y}$. In addition, the Olsson morphisms are the same as the ordinary functoriality of the normal cone (Remarks 1.21 and 1.17).

The Olsson morphism of any fs pullback square is a closed immersion, because it fits into a commutative square of closed immersions from Remark 1.16:

$$\begin{tikzcd} C^{\ell}_{X'/Y'} \arrow[r] \arrow[d, hook] & C^{\ell}_{X/Y}|_{X'} \arrow[d, hook] \\ N^{\ell}_{X'/Y'} \arrow[r, hook] & N^{\ell}_{X/Y}|_{X'} \end{tikzcd}$$

Remark 2.4 (short exact sequences of cone stacks). Recall [Behrend and Fantechi 1997, Definition 1.12]. Let E be a vector bundle stack and C, D cone stacks all on some base algebraic stack X. A composable pair of morphisms of cone stacks

$$E \to C \to D$$

is called a *short exact sequence* if:

- $C \to D$ is a smooth epimorphism.
- The square

$$\begin{tikzcd} E \times C \arrow[r, "pr_2"] \arrow[d, "\sigma"] & C \arrow[d] \\ C \arrow[r] & D \end{tikzcd}$$

  where pr_2 is the projection and σ the action, is cartesian.

These are equivalent to having $C \simeq E \times_X D$ locally in X.

Note that this definition is *fpqc*-local in the base X [Stacks 2005–, 02VL]. Another reduction we will need applies in case there is a commutative diagram of cone stacks

$$\begin{tikzcd} E \arrow[r] \arrow[d] & C \arrow[r] \arrow[d, "s"] & D \arrow[d, "t"] \\ E' \arrow[r] & C' \arrow[r] & D' \end{tikzcd}$$

with E, E' vector bundles. If the top sequence is exact and the arrows labeled s, t are smooth and surjective, then the bottom is exact. To see this, push out along $E \to E'$ so as to assume $E = E'$ (s, t remain smooth and surjective). The diagram on the left is the pullback along the smooth surjection $D' \to D$ of the one on the right:

$$\begin{tikzcd} E \times C' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & C' \arrow[d] \\ C' \arrow[r] & D' \end{tikzcd} \qquad \begin{tikzcd} E \times C \arrow[r] \arrow[d] & C \arrow[d] \\ C \arrow[r] & D \end{tikzcd}$$

We can verify that $E \times C$ is the pullback after smooth-localizing.

Proposition 2.5. *Suppose $X \xrightarrow{f} Y \xrightarrow{g} Z$ are DM maps between log algebraic stacks, and g is log smooth. Then*

$$T^{\ell}_{Y/Z}|_X \to C^{\ell}_{X/Y} \to C^{\ell}_{X/Z}$$

is an exact sequence of cone stacks.

Proof. Encode the log structures on the maps via the top row of the diagram:

$$\begin{tikzcd} X \arrow[r] \arrow[dr] & \mathcal{L}Y \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{L}^2 Z \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{L}^2 \arrow[d] \\ & Y \arrow[r] & \mathcal{L}Z \arrow[r] & \mathcal{L}^1 \end{tikzcd}$$

Since $Y \to \mathcal{L}Z$ is smooth, $\mathcal{L}Y \to \mathcal{L}^2 Z$ is. Moreover, they have the same tangent bundle

$$T^{\ell}_{Y/Z}|_{\mathcal{L}Y} = T_{Y/\mathcal{L}Z}|_{\mathcal{L}Y} = T_{\mathcal{L}Y/\mathcal{L}^2 Z}$$

since the vertical maps are log étale [Ogus 2018, Corollary IV.3.2.4].

Together with the isomorphism $C^{\ell}_{X/Z} \simeq C_{X/\mathcal{L}^2 Z}$, we obtain the exact sequence. □

Remark 2.6. In the proof, the composite

$$C^{\ell}_{X/Y} \to C_{X/\mathcal{L}^2 Z} \simeq C^{\ell}_{X/Z}$$

is precisely the Olsson morphism. This is immediate from the diagram:

$$\begin{tikzcd} X \arrow[r, equal] \arrow[d] & X \arrow[d] \\ \mathcal{L}_q^{\square} \arrow[r] \arrow[d] & \mathcal{L}^2 Z \arrow[d] \\ \mathcal{L}Y \arrow[ur] & \mathcal{L}Z \end{tikzcd}$$

Remark 2.7. The introduction promised three characterizations of $C^{\ell}_{X/Y}$.

The log intrinsic normal cone is characterized by the strict case of Remark 2.3 and the log étale case of Proposition 2.5. This is because any map $X \to Y$ factors into the strict map $X \to \mathcal{L}Y$ composed with the log étale map $\mathcal{L}Y \to Y$ (Remark 1.13).

We can unpack this definition locally using charts. Suppose a morphism has a global fs chart by Artin cones:

$$\begin{tikzcd} X \arrow[r] \arrow[d] & Y \arrow[d] \\ \mathscr{A}_P \arrow[r] & \mathscr{A}_Q \end{tikzcd}$$

The morphism $\mathscr{A}_P \to \mathscr{A}_Q$ is log étale [Olsson 2003, Corollary 5.23]. Let $W = \mathscr{A}_P \times^{\ell}_{\mathscr{A}_Q} Y$ denote the fs pullback, so that $X \to Y$ factors through a strict map to W and W is log étale over Y. We immediately get

$$C^{\ell}_{X/Y} = C_{X/W}.$$

The reader may be reassured by working locally with this definition. If the reader wants instead to work with charts $\mathrm{Spec}(P \to \mathbb{C}[P])$ in the traditional sense, then log étaleness is no longer immediate and we must check Kato's criteria [Ogus 2018, Corollary IV.3.1.10].

Recall Construction 1.1 — after localizing in the étale topology, we obtain a factorization of any map $X \to Y$ as a *strict* closed immersion followed by a log smooth map

$$X \subseteq X_{\theta} \to Y.$$

Proposition 2.5 therefore locally provides a presentation of the log normal cone:

$$C^{\ell}_{X/Y} = [C_{X/X_{\theta}} / T^{\ell}_{X_{\theta}/Y}].$$

Remark 2.8. We want to work with quasicompact, quasiseparated stacks, but $\mathcal{L}$ is not quasicompact. It *is* quasiseparated in the sense of [Stacks 2005–, 04YW], but [Olsson 2003, Remark 3.17] points out it is not quasiseparated in the sense of [Laumon and Moret-Bailly 2000]; see [Chou et al. 2020, Remark 1.1]. A map $X \to Y$ between algebraic stacks with X quasicompact factors through a quasicompact open substack $U \subseteq Y$.

This ensures that any DM map $X \to Y$ of log stacks with X quasicompact and X, Y quasiseparated factors through $X \to U \to Y$ with $X \to U$ strict, U quasicompact and quasiseparated, and $U \to Y$ log étale.

Example 2.9. We provide an example of Construction 1.1 and Remark 1.2.

Consider the diagonal morphism $\mathbb{A}^1 \xrightarrow{\Delta} \mathbb{A}^2$. The addition map $\mathbb{N}^2 \xrightarrow{+} \mathbb{N}$ gives a chart for Δ.

Denote by B the log blowup of $\mathbb{A}^2$ at the ideal $I \subseteq M_{\mathbb{A}^2}$ generated by $\mathbb{N}^2 \setminus \{0\} \subseteq \mathbb{N}^2$. The pullback $\Delta^* I$ is generated by the image of the composite

$$\mathbb{N}^2 \setminus \{0\} \subseteq \mathbb{N}^2 \xrightarrow{+} \mathbb{N}.$$

The pullback is generated globally by a single element and so Δ factors through the log blowup B.

Name the generators $\mathbb{N}^2 = \mathbb{N}e \oplus \mathbb{N}f$. The log blowup B is covered by two affine opens $D_+(e)$ and $D_+(f)$, on which e and f are invertible.

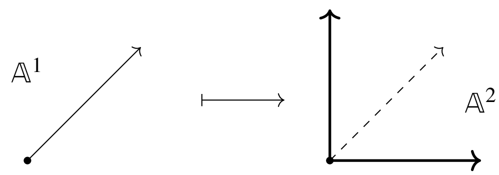
On the chart $D_+(e)$, the morphism $\mathbb{A}^1 \to B$ looks like:

$$\begin{tikzcd}
\mathbb{N} \arrow[d] & \mathbb{N}e \oplus \mathbb{N}(f-e) \arrow[l] \arrow[d] \\
\mathbb{C}[t] & \mathbb{C}\left[x, \frac{y}{x}\right] \arrow[l]
\end{tikzcd}$$

The horizontal morphisms send $f - e \mapsto 0$ and $\frac{y}{x} \mapsto 1$. Because $(f-e)$ maps to $1 \in \mathbb{C}[t]$, the composite

$$\mathbb{N}e \oplus \mathbb{N}(f-e) \to \mathbb{N} \to \mathbb{C}[t]$$

is another chart for the same log structure on $\mathbb{A}^1$. This means that $\mathbb{A}^1 \to D_+(e)$ is strict. The same discussion applies to $D_+(f)$. In the tropical picture [Cavalieri et al. 2020, Section 2], we subdivided $\mathbb{A}^2$ at the image of the ray corresponding to $\mathbb{A}^1$:



Proposition 2.10. *Consider DM-type morphisms $X \xrightarrow{f} Y \xrightarrow{g} Z$ between log algebraic stacks. If $C^\ell_{X/Y} = N^\ell_{X/Y}$, then*

$$N^\ell_{X/Y} \to C^\ell_{X/Z} \to C_{\mathcal{L}Y/\mathcal{L}Z}|_X$$

is an exact sequence of cone stacks.

Proof. Compare [Behrend and Fantechi 1997, Proposition 3.14].

By Proposition 1.20 and Remark 1.18, this sequence composes to zero. Remark 2.4 allows us to repeatedly *fpqc*-localize in X to check exactness of such a sequence. Localizing along strict smooth covers of Z and strict étale covers of X and Y ensures that the normal cones and sheaf pull back. Reduce to the case where X, Y, and Z are affine log schemes and the map $Y \to Z$ admits a global fs chart. We are therefore in the situation of Construction 1.1.

Reduction to $g : Y \to Z$ strict: Factor $Y \to Z$ into a strict closed immersion composed with a log smooth map

$$Y \subseteq W \twoheadrightarrow Z.$$

We obtain a diagram:

$$\begin{tikzcd}
 & T^{\ell}_{W/Z}|_X \arrow[r, equal] \arrow[d] & T_{\mathcal{L}W/\mathcal{L}Z}|_X \arrow[d] \\
N_{X/Y} \arrow[r] \arrow[d, equal] & C^{\ell}_{X/W} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & C_{\mathcal{L}Y/\mathcal{L}W}|_X \arrow[d] \\
N_{X/Y} \arrow[r] & C^{\ell}_{X/Z} \arrow[r] & C_{\mathcal{L}Y/\mathcal{L}Z}|_X
\end{tikzcd}$$

Observe that the diagram commutes — the morphism $T^{\ell}_{W/Z}|_X \to C^{\ell}_{X/W}$ in the proof of Proposition 2.5 factors through an identification $T^{\ell}_{W/Z}|_{\mathcal{L}W} \simeq T^{\ell}_{\mathcal{L}W/\mathcal{L}^2Z}$. Because $\mathcal{L}W \to W$ is log étale, the two tangent spaces are isomorphic [Ogus 2018, IV.3.2.4]. Thus the right square is a pullback. The vertical maps of cones are smooth surjections, so it suffices to show the middle row is exact as in Remark 2.4. We may thereby assume $W = Z$ and $g : Y \to Z$ is a strict closed immersion.

Reduction to $f : X \to Y$ strict: Use Construction 1.1 again to factor $X \to Z$ as a strict closed immersion composed with a log smooth map $X \subseteq W \twoheadrightarrow Z$. The map $X \to W' := W \times_Z Y$ is again a strict closed immersion:

$$\begin{tikzcd}
X \arrow[r, hook] \arrow[dr, bend right] & W' \arrow[r, hook] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & W \arrow[d] \\
 & Y \arrow[r, hook] & Z
\end{tikzcd} \tag{3}$$

Because the top row is strict, $X \to \mathcal{L}W'$ factors through the open subset $W' \subseteq \mathcal{L}W'$ and

$$C_{\mathcal{L}W'/\mathcal{L}W}|_X = C_{\mathcal{L}W'/\mathcal{L}W}|_{W'}|_X = C^{\ell}_{W'/W}|_X = C_{W'/W}|_X.$$

The fs pullback square in (3) also induces a cartesian square of stacks

$$\begin{tikzcd}
\mathcal{L}W' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & \mathcal{L}W \arrow[d] \\
\mathcal{L}Y \arrow[r] & \mathcal{L}Z
\end{tikzcd}$$

with $\mathcal{L}W \to \mathcal{L}Z$ smooth. This reveals that

$$C_{\mathcal{L}Y/\mathcal{L}Z}|_{\mathcal{L}W'} = C_{\mathcal{L}W'/\mathcal{L}W}.$$

Putting this together with the above, we have computed

$$C_{\mathcal{L}Y/\mathcal{L}Z}|_X = C_{W'/W}|_X.$$

The factorization (3) gives a diagram:

$$\begin{tikzcd}
T^{\ell}_{W'/Y}|_X \arrow[r, equal] \arrow[d] & T^{\ell}_{W/Z}|_X \arrow[d] & \\
N_{X/W'} \arrow[r] \arrow[d] & C^{\ell}_{X/W} \arrow[r] \arrow[d] & C_{W'/W}|_X \arrow[d, equal] \\
N_{X/Y} \arrow[r] & C^{\ell}_{X/Z} \arrow[r] & C_{\mathcal{L}Y/\mathcal{L}Z}|_X
\end{tikzcd}$$

The composable vertical arrows are the quotients of Proposition 2.5, so the bottom row will be exact if we show the middle row is. The middle row is exact by a relative form of the original [Behrend and Fantechi 1997, Proposition 3.14]. □

Remark 2.11. The exact sequences of cone stacks in Propositions 2.5, 2.10 are natural in morphisms of composable pairs of arrows.

There is a version of Proposition 2.10 for log cotangent complexes that we will use once later on. From any composable pair $X \to Y \to Z$, we get $X \to \mathcal{L}Y \to \mathcal{L}Z$ and $X \to \mathcal{L}Y \to \mathcal{L}^2 Z$. Both result in the same distinguished triangle

$$\mathbb{L}_{\mathcal{L}Y/\mathcal{L}Z}|_X \to \mathbb{L}^{\ell}_{X/Z} \to \mathbb{L}^{\ell}_{X/Y} \to$$

of [Olsson 2007, 8.10].

In the next example, the log normal cone differs from the ordinary scheme-theoretic one.

Example 2.12. In Example 2.9, we considered the log blowup B of $\mathbb{A}^2$ at the origin and the diagonal map. Pull back along the diagonal to get the identity log blowup of $\mathbb{A}^1$:

$$\begin{tikzcd}
\mathbb{A}^1 \arrow[r] \arrow[d, equal] \arrow[dr, phantom, "\ulcorner \ell"] & B \arrow[d] \\
\mathbb{A}^1 \arrow[r] & \mathbb{A}^2
\end{tikzcd}$$

Let $\bar{o}_{\mathbb{N}}, \bar{o}_{\mathbb{N}^2}$ both be $\operatorname{Spec}\mathbb{C}$, with log structures coming from $\mathbb{N}$ and $\mathbb{N}^2$, respectively. Then the inclusions of the origins $\bar{o}_{\mathbb{N}} \in \mathbb{A}^1$ and $\bar{o}_{\mathbb{N}^2} \in \mathbb{A}^2$ are strict.

Take the pullback of the above diagram along the inclusion $\bar{o}_{\mathbb{N}^2} \in \mathbb{A}^2$:

$$\begin{tikzcd}
\bar{o}_{\mathbb{N}} \arrow[r] \arrow[d, equal] \arrow[dr, phantom, "\ulcorner \ell"] & D \arrow[d] \\
\bar{o}_{\mathbb{N}} \arrow[r] & \bar{o}_{\mathbb{N}^2}
\end{tikzcd}$$

The map $D \to \bar{o}_{\mathbb{N}^2}$ is the exceptional divisor of B, which is $\mathbb{P}^1$ with log structure $\overline{M}_x = \mathbb{N}^2$ at the intersections with the axes and $\overline{M}_x = \mathbb{N}$ elsewhere.

To see the log normal cone differ from the ordinary one, compute the normal cones of the arrows in this square: $C^\ell_{\bar{o}_{\mathbb{N}}/\bar{o}_{\mathbb{N}}} = \bar{o}$, $C^\ell_{\bar{o}_{\mathbb{N}}/\bar{o}_{\mathbb{N}^2}} = C^\ell_{\bar{o}_{\mathbb{N}}/D} = \mathbb{A}^1$, and $C^\ell_{D/\bar{o}_{\mathbb{N}^2}} = \mathbb{P}^1$. Although $\bar{o}_{\mathbb{N}}$ and $\bar{o}_{\mathbb{N}^2}$ have the same underlying scheme, the log normal cones of $\bar{o}_{\mathbb{N}}$ over them are different.

Remark 2.13. A handy consequence of Proposition 2.10 is that, if $Y \to Z$ is a DM-type morphism between log algebraic stacks and $Y' \to Y$ is a *strict* étale map, then

$$C^\ell_{Y'/Z} \simeq C^\ell_{Y/Z}|_{Y'}.$$

This is *not* true without the strictness assumption. This is the observation of W. Bauer precluding the existence of a log cotangent complex with all its desiderata; see [Olsson 2005, Section 7].

In general, it need only be a closed immersion. This is because

$$C^\ell_{Y'/Z} \simeq C_{\mathcal{L}Y/\mathcal{L}Z}|_{Y'} \subseteq N_{\mathcal{L}Y/\mathcal{L}Z}|_{Y'} \subseteq N^\ell_{Y/Z}|_{Y'}$$

is a closed immersion which factors through $C^\ell_{Y/Z}|_{Y'}$, as in Remark 1.18.

For a single example, take the log blowup $B \to \mathbb{A}^2$ of the origin $\bar{o} \in \mathbb{A}^2$. The pullback defines a strict pullback square:

$$\begin{tikzcd} D \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell" very near start] & B \arrow[d] \\ \bar{o} \arrow[r] & \mathbb{A}^2 \end{tikzcd}$$

Because the horizontal morphisms are strict, their log normal cones coincide with the ordinary ones. Log blowups are log étale, so we would erroneously be led to conclude that

$$C_{D/B} \stackrel{?}{=} C_{\bar{o}/\mathbb{A}^2}|_D.$$

The inclusion $D \subseteq B$ is regular, and so is $\bar{o} \in \mathbb{A}^2$, so the normal cones and normal sheaves agree:

$$N_{D/B} = \mathcal{O}_B(D)|_D \quad \text{and} \quad N_{\bar{o}/\mathbb{A}^2}|_D = \mathbb{A}^2_D.$$

The dimensions are different, so they can't be equal.

Lemma 2.14. *Suppose given a strict pullback square*

$$\begin{tikzcd} X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell" very near start] & X \arrow[d] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

of DM-type morphisms between log algebraic stacks for which q is strict and smooth. Then the Olsson morphism

$$C^\ell_{X'/Y'} \xrightarrow{\sim} C^\ell_{X/Y}|_{X'}$$

is an isomorphism.

Proof. We first note that the Olsson morphism $N^\ell_{X'/Y'} \to N^\ell_{X/Y}|_{X'}$ on log normal sheaves is an isomorphism. This is clear from the q strict pullback part of Remark 1.17 and the fact that the ordinary normal sheaves are isomorphic.

Now we know that the morphism of cones $C^\ell_{X'/Y'} \to C^\ell_{X/Y}|_{X'}$ is a closed immersion, and it suffices to show that it is moreover smooth and surjective. We express this map as a composite

$$C^\ell_{X'/Y'} \to C^\ell_{X'/Y} \to C^\ell_{X/Y}|_{X'}.$$

Proposition 2.5 asserts that the first map is smooth and surjective and Proposition 2.10 says the same for the second. $\square$

Lemma 2.15. *Suppose given a pair of fs pullback squares*

$$\begin{tikzcd} \widetilde{X}' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \widetilde{X} \arrow[d, "z"] \\ X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & X \arrow[d] \\ Y' \arrow[r] & Y \end{tikzcd}$$

of DM-type morphisms between log algebraic stacks for which z is strict and smooth. Then the diagram of log normal cones

$$\begin{tikzcd} C^\ell_{\widetilde{X}'/Y'} \arrow[r] \arrow[d, "s'"] \arrow[dr, phantom, "\ulcorner"] & C^\ell_{\widetilde{X}/Y} \arrow[d, "s"] \\ C^\ell_{X'/Y'} \arrow[r] & C^\ell_{X/Y} \end{tikzcd}$$

is cartesian and the arrows s, s' are smooth epimorphisms.

Proof. Proposition 2.10 provides a map of short exact sequences of cone stacks:

$$\begin{tikzcd} BT^\ell_{\widetilde{X}'/X'} \arrow[r] \arrow[d, equal] & C^\ell_{\widetilde{X}'/Y'} \arrow[r, "t'"] \arrow[d, hook] \arrow[dr, phantom, "\ulcorner"] & C^\ell_{X'/Y'}|_{\widetilde{X}'} \arrow[d, hook] \\ BT^\ell_{\widetilde{X}/X}|_{\widetilde{X}'} \arrow[r] & C^\ell_{\widetilde{X}/Y}|_{\widetilde{X}'} \arrow[r, "\tilde{t}"] & C^\ell_{X/Y}|_{\widetilde{X}'} \end{tikzcd}$$

Witness that the right square is cartesian because [Olsson 2005]

$$T^\ell_{\widetilde{X}'/X'} = T^\ell_{\widetilde{X}/X}|_{\widetilde{X}'}$$

and that the arrows $t', \tilde{t}$ are clearly smooth epimorphisms. The arrow $\tilde{t}$ is pulled back from the smooth epimorphism $t : C^{\ell}_{\widetilde{X}/Y} \to C^{\ell}_{X/Y}|_{\widetilde{X}}$, so we have the top pullback square:

$$\begin{tikzcd}
C^{\ell}_{\widetilde{X}'/Y'} \arrow[r] \arrow[d, "t'"] \arrow[dd, bend right=60, "s'"'] \arrow[dr, phantom, "\ulcorner"] & C^{\ell}_{\widetilde{X}/Y} \arrow[d, "t"] \arrow[dd, bend left=60, "s"] & \\
C^{\ell}_{X'/Y'}|_{\widetilde{X}'} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & C^{\ell}_{X/Y}|_{\widetilde{X}} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & \widetilde{X} \arrow[d] \\
C^{\ell}_{X'/Y'} \arrow[r] & C^{\ell}_{X/Y} \arrow[r] & X
\end{tikzcd}$$

The composite vertical rectangle of cones is the diagram we are after, and so the fact that this square is cartesian is clear. It remains only to note the bent arrows s, s' are smooth epimorphisms because they are the composites of t, t' with pullbacks of the smooth epimorphism $\widetilde{X} \to X$. $\square$

3. Log intersection theory

We develop a log intersection theory package using log cotangent complexes and log normal cones in place of the ordinary ones, closely following [Manolache 2012, Sections 3 and 4].

Definition 3.1 (log perfect obstruction theory). Define a *log perfect obstruction theory* (hereafter "Log POT") for a DM-type morphism $f : X \to Y$ to be a closed immersion of cone stacks

$$C^{\ell}_{X/Y} \subseteq E \quad (\text{equiv. } N^{\ell}_{X/Y} \subseteq E)$$

of the log normal cone into a vector bundle stack E.

Given an fs pullback square

$$\begin{tikzcd}
X' \arrow[r] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell"] & X \arrow[d, "f"] \\
Y' \arrow[r] & Y
\end{tikzcd}$$

and a Log POT $C^{\ell}_{X/Y} \subseteq E$ for f, the Olsson morphism

$$C^{\ell}_{X'/Y'} \xhookrightarrow{\varphi} C^{\ell}_{X/Y}|_{X'} \subseteq E|_{X'}$$

defines a *pullback* Log POT.

A related notion of *pullback* Log POT arises when $X' \to X$ is log étale and $f : X \to Y$ any DM-type map. Then Remark 2.13 shows the map

$$C^{\ell}_{X'/Y} \to C^{\ell}_{X/Y}|_{X'}$$

is a closed immersion, and we can compose with an obstruction theory for f to get one for the composite $X' \to X \to Y$.

Given a Log POT $C^{\ell}_{X/Y} \subseteq E$ for some f, suppose X has a stratification by global quotient stacks and Y is log smooth and equidimensional. Then [Kresch 1999, Proposition 5.3.2] gives us a unique cycle

$$[X, E]^{\ell vir} \in A_*X$$

which pulls back to the class $[C^{\ell}_{X/Y}] \in A_*E$. This class is called the *log virtual fundamental class* (hereafter "Log VFC").

Remark 3.2. When $\mathcal{L}Y$ is equidimensional, so is $C^{\ell}_{X/Y}$. The correct definition of the Log VFC requires that the cone be equidimensional. If Y is log smooth, $Y \subseteq \mathcal{L}Y$ is dense. If Y is also equidimensional, we get that $\mathcal{L}Y$ is. This explains our assumptions in Definition 3.1. We don't include these assumptions in the definition of a Log POT only because we may have log Gysin maps more generally.

Nonequidimensional log stacks arise naturally elsewhere in log Gromov–Witten theory. For example, the stacks of punctured log curves $\breve{\mathfrak{M}}$ and punctured maps $\breve{\mathfrak{M}}(\mathscr{X}/B)$ to an Artin fan $\mathscr{X}$ are not equidimensional. They are "idealized log smooth" over the base [Abramovich et al. 2020, Proposition 3.3, Theorem 3.24], which locally entails a composite of a log smooth map and a closed embedding coming from a monoidal ideal.

Definition 3.3 (log Gysin map). Suppose a DM-type $f : X \to Y$ has a Log POT $C^{\ell}_{X/Y} \subseteq E$. Given a DM-type log map $k : V \to Y$ with V log smooth and equidimensional, form the fs pullback:

$$\begin{tikzcd} W \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & V \arrow[d, "k"] \\ X \arrow[r, "f"] & Y \end{tikzcd}$$

The embedding

$$C^{\ell}_{W/V} \subseteq C^{\ell}_{X/Y}|_W \subseteq E|_W$$

results in a class

$$[C^{\ell}_{W/V}, E] \in A_*W.$$

Mimicking [Manolache 2012], we call this "map"

$$f^! = f^!_E$$

the *log Gysin map*.

Remark 3.4. Consider a DM-type morphism $f : X \to Y$ of log algebraic stacks. The cartesian square

$$\begin{tikzcd} \mathcal{L}X \arrow[r, "s"] \arrow[d] \arrow[dr, phantom, "\ulcorner"] & X \arrow[d] \\ \mathcal{L}^2Y \arrow[r, "d_0"] & \mathcal{L}Y \end{tikzcd}$$

from Remark 1.18 results in a closed embedding

$$C_{\mathcal{L}X/\mathcal{L}Y} \simeq C_{\mathcal{L}X/\mathcal{L}^2Y} \subseteq C^{\ell}_{X/Y}|_{\mathcal{L}X}$$

which we use to canonically extend an obstruction theory $C^{\ell}_{X/Y} \subseteq E$ to a closed embedding

$$C_{\mathcal{L}X/\mathcal{L}Y} \subseteq E|_{\mathcal{L}X}.$$

Now suppose given a composable pair $X \xrightarrow{f} Y \xrightarrow{g} Z$ as above and equip f, g with Log POT's

$$C^{\ell}_{X/Y} \subseteq F, \quad C^{\ell}_{Y/Z} \subseteq G.$$

Define a *compatibility datum* or *compatible triple* for such a pair to be a traditional compatibility datum [Manolache 2012, Definition 4.5] for

$$X \xrightarrow{f} \mathcal{L}Y \xrightarrow{g} \mathcal{L}^2 Z,$$

endowing $\mathcal{L}Y \to \mathcal{L}^2 Z$ with the extended obstruction theory

$$C_{\mathcal{L}Y/\mathcal{L}^2 Z} \subseteq C^{\ell}_{Y/Z}|_{\mathcal{L}Y} \subseteq G|_{\mathcal{L}Y}.$$

This entails a commutative diagram in the derived category of X

$$\begin{tikzcd}
\mathcal{G}|_X \arrow[r] \arrow[d] & \mathcal{E} \arrow[r] \arrow[d] & \mathcal{F} \arrow[r, "+1"] \arrow[d] & {} \\
\mathbb{L}_{\mathcal{L}Y/\mathcal{L}^2 Z}|_X \arrow[r] & \mathbb{L}^{\ell}_{X/Z} \arrow[r] & \mathbb{L}^{\ell}_{X/Y} \arrow[r, "+1"] & {}
\end{tikzcd}$$

where

- the rows are distinguished triangles,
- the objects $\mathcal{E}, \mathcal{F}, \mathcal{G}$ are of perfect amplitude in $[-1, 0]$, inducing vector bundle stacks E, F, G,
- the vertical arrows are isomorphisms on h^0 and epimorphisms on h^{-1}, inducing obstruction theories $C^{\ell}_{X/Z} \subseteq E$, $C^{\ell}_{X/Y} \subseteq F$, $C_{\mathcal{L}Y/\mathcal{L}^2 Z} \subseteq G|_{\mathcal{L}Y}$.

We offer a couple of basic remarks about our definitions before the examples and theorems.

Remark 3.5. The map $f^!$ just defined takes in log smooth equidimensional stacks DM over Y and produces classes in certain Chow Groups. The operations $f^!$ are refined to log Chow in [Barrott 2018; Herr et al. 2023].

Remark 3.6. Given an fs pullback square

$$\begin{tikzcd}
X' \arrow[r] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell"] & X \arrow[d, "f"] \\
Y' \arrow[r] & Y
\end{tikzcd}$$

of DM maps where f has a Log POT: $C^{\ell}_{X/Y} \subseteq E$, endow f' with the pullback Log POT. Then

$$f^! = f'^!$$

when applied to log smooth, equidimensional log schemes over Y'.

Remark 3.7. If $C^{\ell}_{X/Y} = N^{\ell}_{X/Y}$ for a DM morphism $f : X \to Y$, then $N^{\ell}_{X/Y}$ is a vector bundle stack. One locally factors $X \subseteq X_\theta \to Y$ as in Construction 1.1 and recognizes $C_{X/X_\theta} \simeq N_{X/X_\theta}$, implying $X \subseteq X_\theta$ is regular and N_{X/X_θ} is a vector bundle. We can then take $E = N^{\ell}_{X/Y}$ as an obstruction theory. If X, Y are equidimensional and Y is log smooth, unwinding definitions shows

$$f^!(Y) = [X],$$

where $[X]$ is the fundamental class of X.

Remark 3.8. Log Gysin maps don't commute with pushforward. Let

$$\begin{tikzcd} X' \arrow[r, "p"] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell" near start] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

be an fs pullback square. Endow $f : X \to Y$ with a Log POT $C^{\ell}_{X/Y} \subseteq E$ and give f' the pullback obstruction theory. Then the usual equality [Manolache 2012, Theorem 4.1(i)] can fail:

$$f^! q_* \neq p_* f'^!.$$

Take the square of Example 2.12

$$\begin{tikzcd} \bar{o}_{\mathbb{N}} \arrow[r] \arrow[d, equal] \arrow[dr, phantom, "\ulcorner \ell" near start] & D \arrow[d] \\ \bar{o}_{\mathbb{N}} \arrow[r] & \bar{o}_{\mathbb{N}^2} \end{tikzcd}$$

and apply both operations to $[\bar{o}_{\mathbb{N}}]$ for a counterexample.

This arises in [Holmes et al. 2019] because pushing forward along various blowups fails to preserves intersection products of DR cycles. This phenomenon was also observed in [Ranganathan 2019].

Remark 3.9. Virtual fundamental classes don't push forward along log blowups: Let $X \to F$ be the morphism from a stack X to its Artin fan (the reader may take a traditional chart instead of F). Choose a finite subdivision $\widehat{F} \to F$, and form the fs pullback:

$$\begin{tikzcd} \widehat{X} \arrow[r] \arrow[d, "p"'] \arrow[dr, phantom, "\ulcorner \ell" near start] & \widehat{F} \arrow[d] \\ X \arrow[r] & F \end{tikzcd}$$

Suppose given a map $f : X \to Y$ with a Log POT $C^{\ell}_{X/Y} \subseteq E$ and equip $f \circ p : \widehat{X} \to Y$ with the pullback obstruction theory

$$C^{\ell}_{\widehat{X}/Y} \subseteq C^{\ell}_{X/Y}|_{\widehat{X}} \subseteq E|_{\widehat{X}}.$$

Then possibly

$$p_*[\widehat{X}, E]^{\ell vir} \neq [X, E]^{\ell vir}.$$

A counterexample is again given by $p : D \to \bar{o}_{\mathbb{N}^2}$, $f : \bar{o}_{\mathbb{N}^2} = \bar{o}_{\mathbb{N}^2}$ as in Example 2.12: $p_*[\mathbb{P}^1] = 0$ for dimension reasons.

This doesn't contradict the main result of [Abramovich and Wise 2018]. We offer a version of their statement in Theorem 3.10.

The rest of this section and the next should reassure the disheartened reader that commonsense formulas of ordinary intersection theory do remain true in the log setting. We regard Remarks 3.8, 3.9 as defects of the usual notion of pushforward p_* in the log setting. The morphisms $\bar{o}_{\mathbb{N}} \to D$, $D \to \bar{o}_{\mathbb{N}^2}$ of Example 2.12 are monomorphisms in the fs category, and $\bar{o}_{\mathbb{N}} \to \bar{o}_{\mathbb{N}^2}$ should be a cycle of *dimension one* in the "two dimensional" log point $\bar{o}_{\mathbb{N}^2}$.

Barrott [2018] introduced log chow groups to correct this defect, in particular via suitable notions of dimension and degree. It also contains compatibility statements between the log notion of p_* and the Gysin map $f^!$ introduced here. See also [Mochizuki 2015].

For now, we content ourselves to use the observation of [Nizioł 2006, Proposition 4.3] that log blowups are birational if the target is log smooth. We will use it to prove that weaker forms of the naïve guesses of Remarks 3.8, 3.9 do hold true, as well as straightforward commutativity of the Gysin maps.

We will need to use Costello's notion [2006, before Theorem 5.0.1] of "pure degree d" to make sense of pushforward on the level of cycles, given by cones embedded in vector bundles. The next theorem allows us to check statements about Log VFC's after a log blowup if the target is log smooth. Its statement and proof are similar to [Abramovich and Wise 2018].

Theorem 3.10. *Suppose given a DM-type map $f : X \to Y$ between locally noetherian algebraic stacks locally of finite type over $\mathbb{C}$ where Y is log smooth and equidimensional. Endow f with a Log POT E and let $X \to F$ be any DM morphism to an Artin Fan. Take the fs pullback along a proper birational map of Artin fans*:

$$\begin{tikzcd} \widehat{X} \arrow[r] \arrow[d, "p"'] \arrow[dr, phantom, "\ulcorner \ell"] & \widehat{F} \arrow[d] \\ X \arrow[r] & F \end{tikzcd} \tag{4}$$

For example, $\widehat{F} \to F$ could be a subdivision or a root stack.

Endow $f \circ p$ with the pullback Log POT

$$C^{\ell}_{\widehat{X}/Y} \subseteq C^{\ell}_{X/Y}|_{\widehat{X}} \subseteq E|_{\widehat{X}}.$$

Then

$$p_*[\widehat{X}, E]^{\ell vir} = [X, E]^{\ell vir}$$

Proof. We will actually show that the map

$$t : C^{\ell}_{\widehat{X}/Y} \to C^{\ell}_{X/Y}$$

is of pure degree one. Then the pushforward $A_*E|_{\widehat{X}} \to A_*E$ sends the class of one cone to the other, and "intersecting with the zero section" gives the equality of VFC's.

We will reduce to the case where $X \to F$ is strict. The statement "t is of pure degree one" may be verified étale-locally in X, as we now argue.

Given a strict étale cover $X' \to X$, write $\widehat{X'} := \widehat{X} \times_X X'$. We have a pullback diagram

$$\begin{tikzcd} C^\ell_{\widehat{X'}/F} \arrow[r, "t'"] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & C^\ell_{X'/F} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & X' \arrow[d] \\ C^\ell_{\widehat{X}/F} \arrow[r, "t"] & C^\ell_{X/F} \arrow[r] & X \end{tikzcd}$$

as in Remark 2.13. Since $X' \to X$ is étale, the other vertical arrows are as well. The property "pure degree one" is smooth-local in the target, so t has it if t' does.

Now étale-localize in X so that $X \to F$ factors through a chart $X \to F_X \to F$ for X. Take the fs pullback along the subdivision $\widehat{F} \to F$:

$$\begin{tikzcd} \widehat{X} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell", very near start] & \widehat{F}_X \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell", very near start] & \widehat{F} \arrow[d] \\ X \arrow[r] & F_X \arrow[r] & F \end{tikzcd}$$

We can then replace F by F_X in the proof of the theorem and assume $X \to F$ is strict.

Apply the proof of Costello's formula [2006, Theorem 5.0.1] to (4) to conclude

$$t : C^\ell_{\widehat{X}/\widehat{F}} \to C^\ell_{X/F}$$

is of pure degree one, since $\widehat{F} \to F$ is birational.

Expanding upon (4):

$$\begin{tikzcd} \widehat{X} \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell", very near start] & \widehat{F} \times Y \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell", very near start] & \widehat{F} \arrow[d] \\ X \arrow[r] & F \times Y \arrow[r] \arrow[d] & F \\ & Y & \end{tikzcd}$$

We get a map of exact sequences of cone stacks:

$$\begin{tikzcd} T^\ell_Y|_{\widehat{X}} \arrow[r] \arrow[d] & C^\ell_{\widehat{X}/\widehat{F}\times Y} \arrow[r] \arrow[d, "\widehat{t}"] \arrow[dr, phantom, "\ulcorner", very near start] & C^\ell_{\widehat{X}/\widehat{F}} \arrow[d, "t"] \\ T^\ell_Y|_X \arrow[r] & C^\ell_{X/F\times Y} \arrow[r] & C^\ell_{X/F} \end{tikzcd}$$

After pulling the bottom row back to $\widehat{X}$, we get the identity on tangent bundles and see that the right square is a pullback. Since the property"of pure degree one" pulls back along smooth maps, the quotient maps in exact sequences of cone stacks are smooth, and t is pure degree one, $\widehat{t}$ is also pure degree one. Because F, $\widehat{F}$ are log étale over a point, $C^{\ell}_{\widehat{X}/\widehat{F}\times Y} = C^{\ell}_{\widehat{X}/\widehat{Y}}$ and $C^{\ell}_{X/F\times Y} = C^{\ell}_{X/Y}$, so the claim is proven. $\square$

Example 3.11. One must be cautious, for Theorem 3.10 is false without the assumption that Y is log smooth. Recall the exceptional divisor $D \to \bar{o}$ of the blowup of $\mathbb{A}^2$ at the origin $\bar{o} = \operatorname{Spec}\mathbb{C}$ from Example 2.12 and its normal cone $C^{\ell}_{D/\bar{o}} = \mathbb{P}^1$.

For the sake of contradiction, let $\widehat{X} = \mathbb{P}^1$ and $X = Y = \bar{o}$ as in the theorem. Endow $C^{\ell}_{\bar{o}/\bar{o}} = \bar{o}$ with the initial Log POT, $E = \bar{o}$. Then

$$[\widehat{X}, E]^{\ell vir} = [D, E]^{\ell vir} = [\mathbb{P}^1] \quad \text{and} \quad [X, E]^{\ell vir} = [\bar{o}, E]^{\ell vir} = [\bar{o}],$$

but again $p_*[\mathbb{P}^1] = 0$ for dimension reasons.

Theorem 3.12 (commutativity of log Gysin map). *Given a composable pair of DM-type maps between log algebraic stacks*

$$X \xrightarrow{f} Y \xrightarrow{g} Z,$$

outfit f, g, and $g \circ f$ with log obstruction theories F, G, E and a compatibility datum (Remark 3.4). Require X to admit stratifications by global quotients.

If $k : V \to Z$ is a log smooth and equidimensional Z-stack and k is DM-type, take fs pullbacks:

$$\begin{tikzcd} T \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell" very near start] & U \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell" very near start] & V \arrow[d] \\ X \arrow[r] & Y \arrow[r] & Z \end{tikzcd}$$

Then the equality

$$[C^{\ell}_{g\circ f} \subseteq E] = [C^{\ell}_{C^{\ell}_g|_X/C^{\ell}_g} \subseteq F \oplus G|_X] \tag{5}$$

holds on X.

Proof. Pullback via k all obstruction theories and their compatibility datum to reduce to showing the theorem for $k : V = Z$. We essentially apply [Manolache 2012, Theorem 4.8] to $X \to \mathcal{L}Y \to \mathcal{L}^2 Z$, endowed with the compatible triple F, G, E by composing with an isomorphism of distinguished triangles:

$$\begin{tikzcd} G|_X \arrow[r] \arrow[d] & F \arrow[r] \arrow[d] & E \arrow[d] \\ \mathbb{L}_{\mathcal{L}Y/\mathcal{L}Z}|_X \arrow[r] \arrow[d, "\sim"] & \mathbb{L}^{\ell}_{X/Z} \arrow[r] \arrow[d, equal] & \mathbb{L}^{\ell}_{X/Y} \arrow[d, "\sim"] \\ \mathbb{L}_{\mathcal{L}Y/\mathcal{L}^2 Z}|_X \arrow[r] & \mathbb{L}_{X/\mathcal{L}Z} \arrow[r] & \mathbb{L}_{X/\mathcal{L}^2 Z} \end{tikzcd}$$

Use Remark 2.8 repeatedly to obtain a strict diagram with U, V quasicompact and étale over the stacks $\mathcal{L}Y$, $\mathcal{L}^2 Z$:

$$\begin{tikzcd} X \arrow[r] \arrow[dr, bend right] & U \arrow[r] \arrow[d] & V \arrow[d] \\ & \mathcal{L}Y \arrow[r] & \mathcal{L}^2 Z \end{tikzcd}$$

Endow the cone $C_{\mathcal{L}Y/\mathcal{L}Z}$ with the pullback log structure from $\mathcal{L}Y$ and pull it back along the part of the diagram above $\mathcal{L}Y$:

$$\begin{tikzcd} C_{\mathcal{L}Y/\mathcal{L}Z}|_X = C_{U/V}|_X \arrow[r] \arrow[dr, bend right] & C_{U/V} \arrow[d] \\ & C_{\mathcal{L}Y/\mathcal{L}Z} \end{tikzcd}$$

The triangle is strict and the map $C_{U/V} \to C_{\mathcal{L}Y/\mathcal{L}Z}$ is pulled back from the étale $U \to \mathcal{L}Y$, so

$$C^{\ell}_{C_{\mathcal{L}Y/\mathcal{L}Z}|_X / C_{\mathcal{L}Y/\mathcal{L}Z}} = C_{C_{U/V}|_X / C_{U/V}}.$$

Write $i : X \to U$ $j : U \to V$ for the maps. Then the compatibility datum pulls back and [Manolache 2012, Theorem 4.8] gives us

$$(j \circ i)^!_E([V]) = i^!_F \circ j^!_G([V]).$$

Unwinding definitions, this becomes

$$[C_{X/V} \subseteq E] = [C_{C_{U/V}|_X / C_{U/V}} \subseteq F \oplus G|_X]. \tag{6}$$

This may be rewritten as

$$[C^{\ell}_{X/Z} \subseteq E] = [C^{\ell}_{C_{\mathcal{L}Y/\mathcal{L}Z}|_X / C_{\mathcal{L}Y/\mathcal{L}Z}} \subseteq F \oplus G|_X],$$

the claimed equality of classes. □

Remark 3.13. Theorem 3.12 says that

$$(g \circ f)^! = f^! g^!$$

in the sense that any log smooth, equidimensional log stack over Z has rationally equivalent images under these two operations.

Remark 3.14. Consider an fs pullback of DM-type morphisms between log algebraic stacks:

$$\begin{tikzcd} X' \arrow[r, "p"] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell"] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

Write $r : X' \to Y$ for the composite $f \circ p = q \circ f'$. If f, q are endowed with Log POT's $C^{\ell}_{X/Y} \subseteq F$, $C^{\ell}_{Y'/Y} \subseteq E$, how should we give r a Log POT?

The fs pullback square induces a pullback of stacks, which may be reexpressed as a "magic square:"

$$\begin{tikzcd} \mathcal{L}X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{L}X \arrow[d] \\ \mathcal{L}Y' \arrow[r] & \mathcal{L}Y \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} \mathcal{L}X' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{L}X \times \mathcal{L}Y' \arrow[d] \\ \mathcal{L}Y \arrow[r] & \mathcal{L}Y \times \mathcal{L}Y. \end{tikzcd}$$

The magic square induces a closed immersion

$$C_{\mathcal{L}X'/\mathcal{L}Y} \subseteq C_{\mathcal{L}X/\mathcal{L}Y}|_{\mathcal{L}X'} \times_{\mathcal{L}X'} C_{\mathcal{L}Y'/\mathcal{L}Y}|_{\mathcal{L}X'}$$

which pulls back to a closed immersion

$$C^{\ell}_{X'/Y} \subseteq C_{\mathcal{L}X/\mathcal{L}Y}|_{X'} \times_{X'} C_{\mathcal{L}Y'/\mathcal{L}Y}|_{X'}$$

on X'. As in Remark 3.4, we have closed embeddings $C_{\mathcal{L}X/\mathcal{L}Y} \subseteq C^{\ell}_{X/Y}|_{\mathcal{L}X}$, $C_{\mathcal{L}Y'/\mathcal{L}Y} \subseteq C^{\ell}_{Y'/Y}|_{\mathcal{L}Y'}$. We endow r with the Log POT given by the composite

$$C^{\ell}_{X'/Y} \subseteq C_{\mathcal{L}X/\mathcal{L}Y}|_{X'} \times_{X'} C_{\mathcal{L}Y'/\mathcal{L}Y}|_{X'} \subseteq C^{\ell}_{X/Y}|_{X'} \times_{X'} C^{\ell}_{Y'/Y}|_{X'} \subseteq F|_{X'} \times_{X'} E|_{X'}.$$

We now construct a compatibility datum for the triangle $r = q \circ f'$, leaving the reader to apply the same argument to the other triangle $r = f \circ p$. By the definitions of the Log POT's, we have a commutative diagram:

$$\begin{tikzcd} C^{\ell}_{X'/Y'} \arrow[r] \arrow[d, hook] & C^{\ell}_{X'/Y} \arrow[r] \arrow[d, hook] & C_{\mathcal{L}Y'/\mathcal{L}Y}|_{X'} \arrow[d, hook] \\ F|_{X'} \arrow[r, "(0\times id)"] & E|_{X'} \times_{X'} F|_{X'} \arrow[r] & E|_{X'} \end{tikzcd}$$

To be clear, the morphism $F|_{X'} \to E|_{X'} \times_{X'} F|_{X'}$ is the vertex map times the identity. It's clear the bottom row comes from a distinguished triangle in the derived category and the top row comes from Remark 2.11.

Corollary 3.15. *Suppose given an fs pullback square*

$$\begin{tikzcd} X' \arrow[r, "p"] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell", very near start] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

of DM-type morphisms between log algebraic stacks which admit stratifications by quotient stacks. Outfit q with a Log POT E and f with a Log POT F; give p, f' the pullback obstruction theories. Then

$$f'^{!} \circ q^{!} = p^{!} \circ f^{!}$$

in the sense that the operations send any log smooth equidimensional input stack to the same class in A_*X'.

Proof. Denote by $r : X' \to Y$ the map $f \circ p = q \circ f'$. Apply Theorem 3.12 to both commutative triangles using the compatibility datum constructed in Remark 3.14 to see that

$$p^{!} \circ f^{!} = r^{!} = f'^{!} \circ q^{!}. \qquad \square$$

4. The log Costello formula

This section proves a log analogue of the Costello formula [2006, Theorem 5.0.1]. The original Costello formula is wrong as stated due to a missing properness hypothesis; this is corrected in [Herr and Wise 2022].

Theorem 4.1. *Consider an fs pullback square of DM-type maps between algebraic stacks*:

$$\begin{tikzcd} X' \arrow[r, "p"] \arrow[d, "f'"'] \arrow[dr, phantom, "\ulcorner \ell"] & X \arrow[d, "f"] \\ Y' \arrow[r, "q"] & Y \end{tikzcd}$$

Assume

- $Y' \to Y$ *is of some pure degree* $d \in \mathbb{Q}$ *as in* [Herr and Wise 2022, Definition 2.3],
- Y', Y *are both log smooth and equidimensional*,
- *all arrows are DM-type and all stacks are locally noetherian and locally finite type over* $\mathbb{C}$,
- X', X *admit stratifications by global quotient stacks* [Kresch 1999] *and*
- q *is proper.*

Endow f *with a log perfect obstruction theory* E *and give* f' *the pullback obstruction theory. Then*

$$p_*[X', E|_{X'}]^{\ell vir} = d \cdot [X, E]^{\ell vir}$$

in the Chow ring of X.

Remark 4.2. Let $Y' \to Y$ be a map between log smooth, equidimensional stacks which is of pure degree d. Let $W \to Y$ be a smooth, log smooth, integral, and saturated morphism and $\widetilde{W} \to W$ a log blowup. Form the fs pullback diagram:

$$\begin{tikzcd} \widetilde{W}' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell"] & \widetilde{W} \arrow[d] \\ W' \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell"] & W \arrow[d] \\ Y' \arrow[r] & Y \end{tikzcd}$$

The property "of pure degree d" pulls back along smooth morphisms, so it applies to $W' \to W$. Then [Nizioł 2006, Proposition 4.3] shows that $\widetilde{W} \to W$ is birational, so $\widetilde{W}' \to \widetilde{W}$ is also of pure degree d.

Proof of Theorem 4.1. Consider the morphism

$$s : C^{\ell}_{X'/Y'} \to C^{\ell}_{X/Y}.$$

We will prove that s is of pure degree d. Both "of pure degree" and the specific degree d can be checked after pulling back s along a strict, smooth cover of $C^{\ell}_{X/Y}$. Lemmas 2.14, 2.15 show that replacing Y or X by a smooth cover results in such a smooth cover of cones.

We may thereby assume X and Y are log schemes and the map f globally factors as in Construction 1.1

$$X \to X_\theta \to \mathbb{A}_Y^{r+s} \to Y.$$

Note $\mathbb{A}_Y^{r+s} \to Y$ is smooth, log smooth, integral, and saturated, and $X_\theta \to \mathbb{A}_Y^{r+s}$ is a log blowup. We are in the situation of Remark 4.2, so pulling back

$$\begin{tikzcd}
X' \arrow[r] \arrow[d, hook] \arrow[dr, phantom, "\ulcorner \not\ell"] & X \arrow[d, hook] \\
X'_\theta \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell"] & X_\theta \arrow[d] \\
Y' \arrow[r] & Y
\end{tikzcd}$$

results in a map $X'_\theta \to X_\theta$ which is pure of degree d along $X \to X_\theta$. The proof of Costello's formula [2006, Theorem 5.0.1] then asserts that

$$C^{\not\ell}_{X'/X'_\theta} \to C^{\not\ell}_{X/X_\theta}$$

is of pure degree d. The short exact sequences of Proposition 2.5

$$\begin{tikzcd}
T^{\ell}_{X'_\theta/Y'} \arrow[r] \arrow[d] & C^{\not\ell}_{X'/X'_\theta} \arrow[r] \arrow[d, "t"] & C^{\ell}_{X'/Y'} \arrow[d, "s"] \\
T^{\ell}_{X_\theta/Y} \arrow[r] & C^{\not\ell}_{X/X_\theta} \arrow[r] & C^{\ell}_{X/Y}
\end{tikzcd}$$

let us conclude that s is as well. □

Remark 4.3. The original statement of Costello's formula did not require q to be proper. In fact, one can allow q to be simply "pure" in the sense of [Herr and Wise 2022, Definition 2.3]. Without any such assumption, one has counterexamples:

- Let $Y = \mathbb{A}^1$, $Y' = \mathbb{G}_m$, X the origin, and f, q natural inclusions.
- Let $Y = \mathbb{A}^1$ and Y' be the bug-eyed line, $\mathbb{A}^1$ with doubled origin. Let X again be the origin and f, q the natural maps.

Remark 4.4. We prove a K theoretic version of Costello's formula in degree $d = 1$ and the corresponding Hironaka pushforward theorem in [Chou et al. 2020, Theorem 2.7]. That proof is necessarily global, because K theory is sensitive to higher-codimension phenomena. The same proof can be used in Chow.

5. The product formula

Let V, W be log smooth, quasiprojective schemes throughout this section. We denote the stacks of *prestable curves* and *stable curves* which have n-markings and genus g by $\mathfrak{M}_{g,n}$, $\overline{M}_{g,n}$, respectively

[Stacks 2005–, 0DMG]. They are endowed with divisorial log structures coming from the locus of singular curves [Gross and Siebert 2013, 1.5, Appendix A; Kato 2000].

Definition 5.1 (log stable maps). The stack of log stable maps $\mathscr{M}^{\ell}_{g,n}(V)$ has fiber over an fs log scheme T the category of diagrams of fs log schemes

$$\begin{tikzcd} C \arrow[r] \arrow[d] & V \\ T & \end{tikzcd}$$

with $C \to T$ a log smooth curve [Kato 2000, Definition 1.2] of genus g and n marked points, such that the underlying diagram of schemes is a stable map of curves.

Remarkably, the log algebraic stack $\mathscr{M}^{\ell}_{g,n}(\operatorname{Spec}\mathbb{C})$ of log curves without a map is isomorphic to the ordinary stack of stable curves $\overline{M}_{g,n}$ with log structure induced by the boundary of degenerate curves [Kato 2000, Theorem 4.5]. The log structures of $\mathscr{M}^{\ell}_{g,n}(V)$ for a general fs target may be more complicated, as they have to do with the "tropical deformation space" of the curve [Gross and Siebert 2013, Example 1.17(1)].

Construction 5.2 [Gross and Siebert 2013, Section 5]. We recall the construction [loc. cit., Section 5] of the natural Log POT for $\mathscr{M}^{\ell}_{g,n}(V) \to \mathfrak{M}_{g,n}$ to clarify differences in notation.

Write $\mathcal{U} \to \mathfrak{M}_{g,n}$ for the universal curve. Define $\mathcal{U}_V$ as the fs pullback, naturally equipped with a tautological map to V:

$$\begin{tikzcd} V & \mathcal{U}_V \arrow[l] \arrow[r, "\pi_V"] \arrow[d] \arrow[dr, phantom, "\ulcorner \ell"] & \mathscr{M}^{\ell}_{g,n}(V) \arrow[d] \\ & \mathcal{U} \arrow[r] & \mathfrak{M}_{g,n} \end{tikzcd}$$

This diagram induces maps between log cotangent complexes

$$\mathbb{L}^{\ell}_{V}|_{\mathcal{U}_V} \longrightarrow \mathbb{L}^{\ell}_{\mathcal{U}_V/\mathcal{U}} \xleftarrow{t} \mathbb{L}^{\ell}_{\mathscr{M}^{\ell}_{g,n}(V)/\mathfrak{M}_{g,n}}|_{\mathcal{U}_V}.$$

The map $\mathcal{U} \to \mathfrak{M}_{g,n}$ is integral, saturated, and log smooth according to its functor of points, so its underlying map of stacks is flat and the fs pullback square is also an ordinary pullback.

Then t is an isomorphism [Olsson 2005, 1.1(iv)], and the log cotangent complex of V is [loc. cit., 1.1(iii)]

$$\mathbb{L}^{\ell}_{V} = \Omega^{\ell}_{V}[0].$$

We've written [0] to consider a coherent sheaf as a chain complex concentrated in degree 0. Via the isomorphism t and this identification, we have obtained a map

$$\Omega^{\ell}_{V}[0]|_{\mathcal{U}_V} \to \mathbb{L}^{\ell}_{\mathcal{U}/\mathfrak{M}_{g,n}}|_{\mathcal{U}_V}. \tag{7}$$

We need the ordinary relative dualizing sheaf $\omega_{\pi_V^\circ}$ and the identification

$$L\pi_V^!(\,\cdot\,) = \omega_{\pi_V^\circ} \overset{L}{\otimes} L\pi_V^*(\,\cdot\,).$$

Tensor (7) by $\omega_{\pi_V^\circ}$ and use the adjunction

$$\Omega_V^\ell[0]\,|_{\mathcal{U}_V} \overset{L}{\otimes} \omega_{\pi_V^\circ} \longrightarrow L\pi_V^! \mathbb{L}^\ell_{\mathscr{M}^\ell_{g,n}(V)/\mathfrak{M}_{g,n}} \quad \text{and} \quad E(V) := R\pi_{V*}(\Omega_V^\ell[0]\,|_{\mathcal{U}_V} \overset{L}{\otimes} \omega_{\pi_V^\circ}) \longrightarrow \mathbb{L}^\ell_{\mathscr{M}^\ell_{g,n}(V)/\mathfrak{M}_{g,n}}.$$

We won't repeat the verification [Gross and Siebert 2013, Proposition 5.1] that $E(V)$ is a Log POT.

Remark 5.3. The map (7) comes from the map on normal cones

$$C^\ell_{\mathscr{M}^\ell_{g,n}(V)/\mathfrak{M}_{g,n}}|_{\mathcal{U}_V} \xrightarrow{\sim} C^\ell_{\mathcal{U}_V/\mathcal{U}} \longrightarrow BT_V^\ell|_{\mathcal{U}_V}.$$

We needed duality, so we opted for the other perspective.

Remark 5.4 (variants). The reader may choose to work in the relative setting of a log smooth and quasiprojective map $V \to S$. Obstruction Theories are obtained in the same way.

The contact order of a log stable map is locally constant and amounts to another piece of discrete data like the genus or number of marked points. We only fix genus and number of markings to be consistent with [Lee and Qu 2018]. The reader may readily vary the numerical type conditions in our formulas.

We need one more stack, $\mathfrak{D}$: Points of $\mathfrak{D}$ over T are diagrams $(C' \leftarrow C \to C'')$ of genus g, n-pointed prestable curves over T whose maps are partial stabilizations (they lie over the identities in $\overline{M}_{g,n}$) that don't both contract any component. In other words, $C \to C' \times C''$ itself is a stable map. This stack is only necessary to form an fs pullback square:

Situation 5.5 [Lee and Qu 2018, Section 2]. Recall the fs pullback square:

$$\begin{tikzcd} \mathscr{M}^\ell_{g,n}(V \times W) \arrow[r] \arrow[d, "c"] \arrow[dr, phantom, "\ulcorner \ell"] & \mathscr{M}^\ell_{g,n}(V) \times \mathscr{M}^\ell_{g,n}(W) \arrow[d, "a"] \\ \mathfrak{D} \arrow[r, "\widetilde{\Delta}"] & \mathfrak{M}_{g,n} \times \mathfrak{M}_{g,n} \end{tikzcd} \tag{8}$$

Let $C \to V \times W$ be a log stable map over a base T. The maps $(C \to V)$, $(C \to W)$ needn't be stable; denote their stabilizations by $(C' \to V)$, $(C'' \to W)$, respectively.

The top horizontal arrow in (8) sends $(C \to V \times W)$ to the induced log stable maps $(C' \to V, C'' \to W)$. The vertical arrow c sends $(C \to V \times W)$ to the partial stabilizations $(C' \leftarrow C \to C'')$. The map $\widetilde{\Delta}$ sends a diagram $(C' \leftarrow C \to C'')$ to the pair of prestable curves C', C''. Finally, a sends a pair of log stable maps $(C' \to V, C'' \to W)$ to the prestable curves (C', C'').

This square has a factorization:

$$\begin{tikzcd}
\mathscr{M}^{\ell}_{g,n}(V\times W) \arrow[r, "h"] \arrow[d, "c"] \arrow[dr, phantom, "\ulcorner\ell"] & Q \arrow[r] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \mathscr{M}^{\ell}_{g,n}(V)\times\mathscr{M}^{\ell}_{g,n}(W) \arrow[d, "a"] \\
\mathfrak{D} \arrow[r, "l"] & Q' \arrow[r, "\phi"] \arrow[d] \arrow[dr, phantom, "\ulcorner\ell"] & \mathfrak{M}_{g,n}\times\mathfrak{M}_{g,n} \arrow[d, "s\times s"] \\
& \overline{M}_{g,n} \arrow[r, "\Delta"] & \overline{M}_{g,n}\times\overline{M}_{g,n}
\end{tikzcd} \tag{9}$$

Where $s:\mathfrak{M}_{g,n}\to\overline{M}_{g,n}$ stabilizes a prestable curve.

To be clear, $Q=\mathscr{M}^{\ell}_{g,n}(V)\times^{\ell}_{\overline{M}_{g,n}}\mathscr{M}^{\ell}_{g,n}(W)$ and $Q'=\mathfrak{M}_{g,n}\times^{\ell}_{\overline{M}_{g,n}}\mathfrak{M}_{g,n}$ are the analogues of Lee and Qu's [2018] P, $\mathfrak{P}$, etc.

Theorem 5.6 (the "log Gromov–Witten product formula"). *With V, W log smooth, quasiprojective schemes,*

$$h_*[\mathscr{M}^{\ell}_{g,n}(V\times W), E(V\times W)]^{\ell vir}=\Delta^!([\mathscr{M}^{\ell}_{g,n}(V), E(V)]^{\ell vir}\times[\mathscr{M}^{\ell}_{g,n}(W), E(W)]^{\ell vir}).$$

Our proof will be the same as K. Behrend's [1999]: we compute the log normal cone of the map $Q\to Q'$ in two different ways.

Remark 5.7 (on diagram (9)). We equip a with the product $E(V)\boxplus E(W)$ of the natural Log POT's of Construction 5.2, adopting the notation

$$E\boxplus E':=E|_{V\times W}\oplus E'|_{V\times W}.$$

The cotangent complex $\mathbb{L}^{\ell}_{\Delta}$ is of perfect amplitude in [-1, 0] because its source and target are log smooth. Therefore $C^{\ell}_{\Delta}=N^{\ell}_{\Delta}$ serves as a natural Log POT for itself. We equip ϕ with the pullback obstruction theory, resulting in

$$\Delta^!=\phi^!$$

by Remark 3.6. We endow the square bounded by ϕ and a with the natural compatibility datum afforded all such squares as in Remark 3.14.

All of the arrows in diagrams (8) and (9) are of DM-type.

Lemma 5.8. *The stabilization map $s:\mathfrak{M}_{g,n}\to\overline{M}_{g,n}$ is log smooth.*

Proof. The cover $\bigsqcup_m\overline{M}_{g,n+m}\to\mathfrak{M}_{g,n}$ given by forgetting marked points and not stabilizing is strict smooth [Lee and Qu 2018, 1.2.1]. This map is in particular Kummer and surjective, and [Illusie et al. 2013, Theorem 0.2] applies with $\mathbb{P}=$ "log smooth" once we argue that the composite $\bigsqcup_m\overline{M}_{g,n+m}\to\overline{M}_{g,n}$ is log smooth.

The forgetful map $\overline{M}_{g,n+1}\to\overline{M}_{g,n}$ is the universal curve, so it is tautologically log smooth and flat. We see the map $\overline{M}_{g,n+m}\to\overline{M}_{g,n}$ is log smooth by iterating this forgetfulness, and this completes the argument. □

Remark 5.9. The map $\mathfrak{D} \to \mathfrak{M}_{g,n}$ which records the initial curve is log étale since the original map was étale [Behrend 1999, Lemma 4] and ours is the fsification thereof. The stack Q' is log smooth because the map $Q' \to \overline{M}_{g,n}$ is pulled back from $s \times s$.

Given a log étale map $X' \to X$ of log smooth log algebraic stacks with X equidimensional, we claim X' must be as well. The maps $X' \subseteq \mathcal{L}X'$, $X \subseteq \mathcal{L}X$ are dense because of the log smoothness assumption and the map $\mathcal{L}X' \to \mathcal{L}X$ is étale. Thus $\mathcal{L}X$ and $\mathcal{L}X'$ are equidimensional, as well as $X' \subseteq \mathcal{L}X'$. This argument shows that fsification preserves equidimensionality of log smooth stacks, so our fs versions of $\mathfrak{D}$, Q' are equidimensional because the original versions [Behrend 1999] were.

Lemma 5.10. *The obstruction theories $E(V)$, $E(W)$, $E(V \times W)$ are compatible in the sense that*

$$\widetilde{\Delta}^*(E(V) \boxplus E(W)) \simeq E(V \times W).$$

Proof. We completely echo the proof of [Behrend 1999, Proposition 6].

Consider the diagram of universal log curves and tautological maps with the notation:

$$\begin{tikzcd}
V & & V \times W \arrow[ll] \\
\mathcal{U}_V \arrow[u, "f_V"] \arrow[d, "\pi_V"'] \arrow[dr, phantom, "\ell^{\urcorner}"] & \widetilde{\mathcal{U}}_V \arrow[l, "s_V"'] \arrow[dr, "\widetilde{\pi}_V"'] & \mathcal{U}_{V \times W} \arrow[l, "q_V"'] \arrow[u, "f_{V \times W}"'] \arrow[d, "\pi_{V \times W}"] \\
\mathscr{M}^{\ell}_{g,n}(V) & & \mathscr{M}^{\ell}_{g,n}(V \times W) \arrow[ll, "r_V"]
\end{tikzcd}$$

We claim $F \to Rq_{V*}q_V^* F$ is an isomorphism for any vector bundle F on $\mathcal{U}_V$. The map q_V represents partial stabilization. We make the argument for contracting one $\mathbb{P}^1$ at a time.

We first compute that $R^p q_{V*}q_V^* F = 0$ for $p \neq 0$. This claim is local in $\mathcal{U}_V$, so assume F is trivial. The fiber of $R^p q_{V*} q_V^* F$ at a point x is $H^p(q_V^{-1}(x), q_V^* F)$. Hence the fibers $q_V^{-1}(x)$ are either a point or $\mathbb{P}^1$. On each fiber, the cohomology of the trivial vector bundle is concentrated in degree 0 [Stacks 2005–, 01XS]. Not only are F and $q_{V*}q_V^* F$ abstractly isomorphic in that case, but the natural map is an isomorphism [Fantechi et al. 2005, Exercise 9.3.11].

The universal curve π_V is tautologically flat, integral, and saturated. The fs pullback square it belongs to is therefore also an ordinary flat pullback, subject to cohomology and base change [Stacks 2005–, Tag 08IB]. This gives

$$Lr_V^* R\pi_{V*} Lf_V^* \Omega_V = R\widetilde{\pi}_{V*} Ls_V^* Lf_V^* \Omega_V = R\widetilde{\pi}_{V*} Rq_{V*} q_V^* Ls_V^* Lf_V^* \Omega_V = R\pi_{V \times W*} Lf_{V \times W}^* (\Omega_V|_{V \times W}).$$

All the same goes for W. Add the two together to get

$$Lr_V^* R\pi_{V*} Lf_V^* \Omega_V \boxplus Lr_W^* R\pi_{W*} Lf_W^* \Omega_W \;=\; R\pi_{V \times W*} Lf_{V \times W}^* (\Omega_V \boxplus \Omega_W).$$

This is dual to the compatibility we set out to prove, so we are through. $\square$

Proof of Theorem 5.6. Compute the log virtual fundamental class $[Q, E(V) \boxplus E(W)]^{vir}$ in two different ways:

$$\begin{aligned}[Q, E(V) \boxplus E(W)]^{vir} &:= [C^{\ell}_{Q/Q'} \subseteq E(V) \boxplus E(W)] \\ &= a^!(Q') \\ &= a^!\phi^!(\mathfrak{M}_{g,n} \times \mathfrak{M}_{g,n}) \\ &= \phi^! a^!(\mathfrak{M}_{g,n} \times \mathfrak{M}_{g,n}) \\ &= \Delta^![\mathscr{M}^{\ell}_{g,n}(V) \times \mathscr{M}^{\ell}_{g,n}(W), E(V) \boxplus E(W)]^{vir}.\end{aligned}$$

On the other hand,

$$[Q, E(V) \boxplus E(W)]^{vir} = h_*[\mathscr{M}^{\ell}_{g,n}(V \times W), E(V \times W)]^{vir}$$

by the log Costello formula, Theorem 4.1. □

Acknowledgments

The present article is part of the author's Ph.D. thesis at the University of Colorado, Boulder under the supervision of Jonathan Wise. Not a result was envisioned, obtained, or fixed without his tremendous support, guidance, and patience.

The author also benefited from correspondence with Dhruv Ranganathan and Lawrence Barrott. Sam Molcho weakened the hypotheses of Theorem 3.10. Writing the K-theoretic counterpart [Chou et al. 2020] to this paper with Y.P. Lee and You-Cheng Chou simplified some arguments.

The author thanks the referee and Y.P. Lee for carefully reading and editing the manuscript.

The author is grateful to the NSF for partial financial support from RTG grant #1840190.

References

[Abramovich and Chen 2014] D. Abramovich and Q. Chen, "Stable logarithmic maps to Deligne–Faltings pairs, II", *Asian J. Math.* **18**:3 (2014), 465–488. MR Zbl

[Abramovich and Wise 2018] D. Abramovich and J. Wise, "Birational invariance in logarithmic Gromov–Witten theory", *Compos. Math.* **154**:3 (2018), 595–620. MR Zbl

[Abramovich et al. 2017] D. Abramovich, Q. Chen, S. Marcus, and J. Wise, "Boundedness of the space of stable logarithmic maps", *J. Eur. Math. Soc.* (*JEMS*) **19**:9 (2017), 2783–2809. MR Zbl

[Abramovich et al. 2020] D. Abramovich, Q. Chen, M. Gross, and B. Siebert, "Punctured logarithmic maps", preprint, 2020. arXiv 2009.07720

[Barrott 2018] L. J. Barrott, "Logarithmic Chow theory", preprint, 2018. arXiv 1810.03746

[Behrend 1999] K. Behrend, "The product formula for Gromov–Witten invariants", *J. Algebraic Geom.* **8**:3 (1999), 529–541. MR Zbl

[Behrend and Fantechi 1997] K. Behrend and B. Fantechi, "The intrinsic normal cone", *Invent. Math.* **128**:1 (1997), 45–88. MR Zbl

[Cavalieri et al. 2020] R. Cavalieri, M. Chan, M. Ulirsch, and J. Wise, "A moduli stack of tropical curves", *Forum Math. Sigma* **8** (2020), Paper No. e23, 93. MR Zbl

[Chen 2014] Q. Chen, "Stable logarithmic maps to Deligne–Faltings pairs, I", *Ann. of Math.* (2) **180**:2 (2014), 455–521. MR Zbl

[Chou et al. 2020] Y.-C. Chou, L. Herr, and Y.-P. Lee, "The Log Product Formula in quantum K-theory", preprint, 2020. arXiv 2012.09379

[Costello 2006] K. Costello, "Higher genus Gromov–Witten invariants as genus zero invariants of symmetric products", *Ann. of Math.* **164**:2 (2006), 561–601. MR Zbl

[Fantechi et al. 2005] B. Fantechi, L. Göttsche, L. Illusie, S. L. Kleiman, N. Nitsure, and A. Vistoli, *Fundamental algebraic geometry*, Mathematical Surveys and Monographs **123**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[Gross and Siebert 2013] M. Gross and B. Siebert, "Logarithmic Gromov–Witten invariants", *J. Amer. Math. Soc.* **26**:2 (2013), 451–510. MR Zbl

[Herr and Wise 2022] L. Herr and J. Wise, "Costello's pushforward formula: errata and generalization", *manuscripta math.* (2022).

[Herr et al. 2023] L. Herr, S. Molcho, R. Pandharipande, and J. Wise, "Rendimento dei conti", 2023. Forthcoming.

[Holmes et al. 2019] D. Holmes, A. Pixton, and J. Schmitt, "Multiplicativity of the double ramification cycle", *Doc. Math.* **24** (2019), 545–562. MR Zbl

[Illusie 1997] L. Illusie, "Géometrie logarithmique", 1997. Notes of a course at IHP.

[Illusie et al. 2013] L. Illusie, C. Nakayama, and T. Tsuji, "On log flat descent", *Proc. Japan Acad. Ser. A Math. Sci.* **89**:1 (2013), 1–5. MR Zbl

[Ito et al. 2020] T. Ito, K. Kato, C. Nakayama, and S. Usui, "On log motives", *Tunis. J. Math.* **2**:4 (2020), 733–789. MR Zbl

[Kato 2000] F. Kato, "Log smooth deformation and moduli of log smooth curves", *Internat. J. Math.* **11**:2 (2000), 215–232. MR Zbl

[Khan and Rydh 2018] A. A. Khan and D. Rydh, "Virtual Cartier divisors and blow-ups", preprint, 2018. arXiv 1802.05702

[Kontsevich and Manin 1996] M. Kontsevich and Y. Manin, "Quantum cohomology of a product", *Invent. Math.* **124**:1-3 (1996), 313–339. MR Zbl

[Kresch 1999] A. Kresch, "Cycle groups for Artin stacks", *Invent. Math.* **138**:3 (1999), 495–536. MR Zbl

[Laumon and Moret-Bailly 2000] G. Laumon and L. Moret-Bailly, *Champs algébriques*, Ergeb. Math. Grenzgeb. **39**, Springer, 2000. MR Zbl

[Lee and Qu 2018] Y.-P. Lee and F. Qu, "A product formula for log Gromov–Witten invariants", *J. Math. Soc. Japan* **70**:1 (2018), 229–242. MR Zbl

[Li 2001] J. Li, "Stable morphisms to singular schemes and relative stable morphisms", *J. Differential Geom.* **57**:3 (2001), 509–578. MR Zbl

[Manolache 2012] C. Manolache, "Virtual pull-backs", *J. Algebraic Geom.* **21**:2 (2012), 201–245. MR Zbl

[Mochizuki 2015] S. Mochizuki, "Monomorphisms in categories of log schemes", *Kodai Math. J.* **38**:2 (2015), 365–429. MR Zbl

[Molcho and Ranganathan 2021] S. Molcho and D. Ranganathan, "A case study of intersections on blowups of the moduli of curves", preprint, 2021. arXiv 2106.15194

[Nabijou and Ranganathan 2022] N. Nabijou and D. Ranganathan, "Gromov–Witten theory with maximal contacts", *Forum Math. Sigma* **10** (2022), Paper No. e5, 34. MR Zbl

[Niziol 2006] W. a. Nizioł, "Toric singularities: log-blow-ups and global resolutions", *J. Algebraic Geom.* **15**:1 (2006), 1–29. MR

[Ogus 2018] A. Ogus, *Lectures on logarithmic algebraic geometry*, Cambridge Studies in Advanced Mathematics **178**, Cambridge University Press, 2018. MR Zbl

[Olsson 2003] M. C. Olsson, "Logarithmic geometry and algebraic stacks", *Ann. Sci. École Norm. Sup.* (4) **36**:5 (2003), 747–791. MR Zbl

[Olsson 2005] M. C. Olsson, "The logarithmic cotangent complex", *Math. Ann.* **333**:4 (2005), 859–931. MR Zbl

[Olsson 2007] M. Olsson, "Sheaves on Artin stacks", *J. Reine Angew. Math.* **603** (2007), 55–112. MR Zbl

[Ranganathan 2019] D. Ranganathan, "A note on the cycle of curves in a product of pairs", preprint, 2019. arXiv 1910.00239

[Ranganathan 2022] D. Ranganathan, "Logarithmic Gromov–Witten theory with expansions", *Algebr. Geom.* **9**:6 (2022), 714–761. 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, 1972. MR Zbl

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Wise 2011] J. Wise, "Obstruction theories and virtual fundamental classes", preprint, 2011. arXiv 1111.4200

Communicated by Martin Olsson
Received 2020-10-18 Revised 2022-04-05 Accepted 2022-08-18

herr@math.utah.edu Universiteit Leiden, Netherlands

ALGEBRA AND NUMBER THEORY 17:7 (2023)
https://doi.org/10.2140/ant.2023.17.1325

Intersecting geodesics on the modular surface

Junehyuk Jung and Naser Talebizadeh Sardari

We introduce the *modular intersection kernel*, and we use it to study how geodesics intersect on the full modular surface $\mathbb{X} = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$. Let C_d be the union of closed geodesics with discriminant d and let $\beta \subset \mathbb{X}$ be a compact geodesic segment. As an application of Duke's theorem to the modular intersection kernel, we prove that $\{(p, \theta_p) : p \in \beta \cap C_d\}$ becomes equidistributed with respect to $\sin\theta\, ds\, d\theta$ on $\beta \times [0, \pi]$ with a power saving rate as $d \to +\infty$. Here θ_p is the angle of intersection between β and C_d at p. This settles the main conjectures introduced by Rickards(2021).

We prove a similar result for the distribution of angles of intersections between C_{d_1} and C_{d_2} with a power-saving rate in d_1 and d_2 as $d_1 + d_2 \to \infty$. Previous works on the corresponding problem for compact surfaces do not apply to $\mathbb{X}$, because of the singular behavior of the modular intersection kernel near the cusp. We analyze the singular behavior of the modular intersection kernel by approximating it by general (not necessarily spherical) point-pair invariants on $\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})$ and then by studying their full spectral expansion.

1. Introduction

Let Y be a negatively curved surface of finite area. The prime geodesic theorem [Sarnak 1980] states that the number of primitive closed geodesics having length less than L, which we denote by $\pi(L)$, satisfies

$$\pi(L) \sim \frac{e^L}{L},$$

as $L \to \infty$. A natural problem is to understand how primitive closed geodesics of length less than L are positioned or distributed in Y as $L \to \infty$. In particular, one may ask

(1) how the number of transversal intersections $I(\alpha_1, \alpha_2)$ between two primitive closed geodesics α_1 and α_2 is distributed, or

(2) how the set of angles of intersections between α_1 and α_2 is distributed,

We thank D. Jakobson, V. Blomer, D. Milicevic, C. Pagano, M. Lee, M. Lipnowski, I. Khayutin, P. Humphries, Y. Kim, and J. Yim for valuable comments. Jung thanks A. Reid for the discussion that led to this project. Jung was supported by NSF grant DMS-1900993, and by Sloan Research Fellowship. Talebizadeh Sardari was supported partially by the National Science Foundation under Grant No. DMS-2015305, and is grateful to Max Planck Institute for Mathematics in Bonn for its hospitality and financial support.
MSC2020: 11F03, 11F70, 11F72.
Keywords: Closed geodesics, Modular forms, Intersection angles.

as one varies α_1, or both α_1 and α_2? Bonahon [1986] defined the *intersection form* $i : \mathcal{C} \times \mathcal{C} \to \mathbb{R}^+$ on the space of currents $\mathcal{C}$ such that when μ_i $(i = 1, 2)$ is the unique invariant measure corresponding to α_i, then $i(\mu_1, \mu_2) = I(\alpha_1, \alpha_2)$. When Y is compact, Pollicott and Sharp [2006] used an extension of the intersection form to understand the distribution of angles of self-intersections of closed geodesic α having length less than L, as $L \to \infty$. When Y is a compact hyperbolic surface, using the intersection form, Herrera Jaramillo [2015] proved that the distribution of $I(\alpha_1, \alpha_2)/(l(\alpha_1)l(\alpha_2))$ for closed geodesics α_1, α_2 of length $< L$, is concentrated near $1/(2\pi^2(g-1)) = 2/(\pi\ \mathrm{vol}(Y))$ with exponentially decaying tail, as $L \to \infty$. Here $l(\cdot)$ is the length function, and g is the genus of Y.

In this article, we study a refined problem:

(3) How are the locations and angles of intersections between α_1 and α_2 jointly distributed relative to α_2, as one varies α_1, or both α_1 and α_2?

Let $\mathbb{X} = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$ be the full modular surface. The connection between the geometry of geodesics on $\mathbb{X}$ and number theory has a rich history. Artin [1924] discovered a relation between geometry of geodesics in $\mathbb{X}$ and continued fraction expansion. As a result, he proved that there is a hyperbolic geodesic in $\mathbb{X}$ that comes arbitrarily close to any given hyperbolic segment in $\mathbb{X}$. So this geodesic is not only dense, but dense in all directions simultaneously. Another deep connection is discovered in the spectacular work of Katok [1985]. She showed that certain holomorphic Poincaré series (introduced by Petersson) associated with closed geodesics on a Fuchsian group of the first kind, span the corresponding space of cusp forms. Moreover, she proved a formula relating the intersection angles between pairs of closed geodesics to the periods of these holomorphic Poincaré series.

On $\mathbb{X}$, primitive oriented closed geodesics are in one-to-one correspondence with conjugacy classes of primitive hyperbolic elements of $\mathrm{PSL}_2(\mathbb{Z})$. Moreover there is a bijection between the primitive hyperbolic conjugacy classes and the $\mathrm{SL}_2(\mathbb{Z})$ equivalence classes of primitive integral binary quadratic forms of nonsquare positive discriminant [Luo et al. 2009; Sarnak 1982]. So by the discriminant of a primitive closed geodesic, we mean the discriminant of the corresponding binary quadratic form. In particular, if the hyperbolic class γ is associated to the binary quadratic form Q then γ^{-1} is associated to $-Q$.

Let (x_d, y_d) be the fundamental solution of Pell's equation $x^2 - dy^2 = 4$, and let $\varepsilon_d := \frac{1}{2}(x_d + \sqrt{d}\, y_d) > 1$. Each oriented primitive closed geodesics of discriminant d has a unique lift to a closed geodesic of length $2\log\varepsilon_d$ in the unit tangent bundle $S\mathbb{X}$. Let $h(d)$ be the number of inequivalent primitive integral binary quadratic forms of discriminant d. We denote the disjoint union of these $h(d)$ closed geodesics by $\mathscr{C}_d \subset S\mathbb{X}$, which has total length $2h(d)\log\varepsilon_d$.

Note that the closed geodesic on $\mathbb{X}$ has length $\log\varepsilon_d$ or $2\log\varepsilon_d$ according as Q is or is not equivalent to $-Q$ [Duke 1988, page 75]. We now let C_d be the union of primitive (unoriented) closed geodesics of discriminant d on $\mathbb{X}$, and note that $l(C_d) = h(d)\log\varepsilon_d$ is the total length of C_d.

Theorem 1.1. *Fix $T > 100$, and let β be a compact oriented geodesic segment of length < 1 in the region determined by $y < T$ on $\mathbb{X}$. For $0 < \theta_1 < \theta_2 < \pi$, let $I_{\theta_1,\theta_2}(\beta, C_d)$ be the number of intersections between*

β and C_d with the angle between θ_1 and θ_2. (Here the angle between β and C_d at $p \in \beta \cap C_d$ is measured counterclockwise from the tangent to β at p to the tangent to C_d at p.)

Then we have

$$\frac{I_{\theta_1,\theta_2}(\beta, C_d)}{l(\beta)l(C_d)} = \frac{3}{\pi^2}\int_{\theta_1}^{\theta_2} \sin\theta d\theta + O_\epsilon\,(d^{-25/3584+\epsilon}),$$

uniformly in β, θ_1, and θ_2, under the assumption that

$$\theta_2 - \theta_1 \gg d^{-25/7168},$$

and that

$$l(\beta) \gg d^{-25/7168}.$$

(Here and elsewhere, $A \ll_\tau B$ means $|A| \le C(\tau)B$ for some constant $C(\tau)$ that depends only on τ.)

Remark 1.1. This statement is false if C_d is replaced by individual geodesics. For instance, the set of intersections between β and a closed geodesic α does not necessarily become equidistributed as $l(\alpha) \to \infty$. To see this, take a finite sheeted covering S of $\mathbb{X}$ whose genus is ≥ 2. Then according to Rivin's work [2001] there are arbitrarily long simple closed geodesics on S. Note that these simple closed geodesics must be contained in a compact part of S [Jung and Reid 2021]. This implies that there is a compact set $C \subset \mathbb{X}$ which contains arbitrarily long primitive closed geodesics. Take a geodesic segment β in $\mathbb{X} - C$. Then there are infinitely many closed geodesics which do not intersect β.

Remark 1.2. The exponent $-\frac{25}{3584}$ can be improved slightly by refining our argument (for instance, by inputting the Weyl-like subconvex bound [Petrow and Young 2019] instead of the Burgess-like subconvex bound [Heath-Brown 1980]), but in order to keep the exposition simple, we do not discuss the optimal rate in the current article.

As an immediate consequence, we deduce that the intersection points and corresponding angles become equidistributed, resolving the main conjectures introduced by Rickards [2021].

Corollary 1.2. *Fix a closed geodesic α. Then for any fixed segment $\beta \subset \alpha$, and any fixed $0 < \theta_1 < \theta_2 < \pi$, we have*

$$\lim_{d\to\infty} \frac{I_{\theta_1,\theta_2}(\beta, C_d)}{I(\alpha, C_d)} = \frac{l(\beta)}{l(\alpha)} \int_{\theta_1}^{\theta_2} \frac{\sin\theta}{2}\, d\theta.$$

Remark 1.3. Rickards's work is motivated by the work of Darmon and Vonk [2022] on the arithmetic (p-arithmetic) intersection between pairs of oriented closed geodesics on the modular surfaces (Shimura curves). The arithmetic intersection between oriented closed geodesics α_1 and α_2 of discriminants D_1 and D_2 only depends on D_1 and D_2 and the angles of intersections between α_1 and α_2. Darmon and Vonk [2022, Conjecture 2] conjectured that the p-arithmetic intersection is algebraic and belongs to the composition of the Hilbert class field of real quadratic fields of discriminants D_1 and D_2.

To prove our main results, we introduce the *modular intersection kernel*. For $\delta > 0$ and $\theta_1, \theta_2 \in (0, \pi)$, let $k_\delta^{\theta_1,\theta_2} : S\mathbb{H} \times S\mathbb{H} \to \mathbb{R}$ be the integral kernel defined by

$$k_\delta^{\theta_1,\theta_2}((x_1, \xi_1), (x_2, \xi_2)) = 1,$$

if the geodesic segments of length δ from x_i with the initial vector ξ_i intersect at an angle $\in (\theta_1, \theta_2)$, and 0 otherwise. Under the identification $S\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})$, for a given discrete subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$, we define the *modular intersection kernel* $K_\delta^{\theta_1,\theta_2} : \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \times \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{R}$ by taking the average of $k_\delta^{\theta_1,\theta_2}$ over Γ:

$$K_\delta^{\theta_1,\theta_2}(g_1, g_2) = \sum_{\gamma\in\Gamma} k_\delta^{\theta_1,\theta_2}(g_1, \gamma g_2).$$

The basic idea of the proof of Theorem 1.1 then is as follows. Heuristically,

$$I_{\theta_1,\theta_2}(\beta, C_d)$$

should be well approximated by

$$\frac{1}{2\delta^2} \int_{\mathscr{C}_d} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2, \tag{1-1}$$

where $\tilde{\beta} \subset S\mathbb{X}$ is a lift of β with either of orientations of β

$$\tilde{\beta}(t) = (\beta(t), \beta'(t)),$$

under assuming that $\beta(t)$ is parametrized by the arc length. As noted in [Luo et al. 2009], Duke's theorem [1988] can be extended to the equidistribution of $\mathscr{C}_d$ in $\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})$ as $d \to \infty$. Observing that

$$\frac{1}{2\delta^2} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, g)\, ds_1 \tag{1-2}$$

is a compactly supported function in g for compact β, (1-1) is

$$\sim \frac{l(\mathscr{C}_d)}{2\delta^2} \int_g \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, g)\, ds_1\, d\mu_g,$$

which is asymptotically $(3/\pi^2) l(C_d) l(\beta) \int_{\theta_1}^{\theta_2} \sin\alpha\, d\alpha$ as $\delta \to 0$, by an explicit computation.

Note that (1-2) is a discontinuous function. Therefore, in order to obtain the rate of convergence, we need a smooth approximation of (1-2), and a quantified version of Duke's theorem with explicit dependency on the test functions. To this end, we follow the argument sketched in [Luo et al. 2009] to prove:

Theorem 1.3. *Assume that $f \in C_0^\infty(\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R}))$ has support in the region determined by $y < T$. Then we have*

$$\frac{1}{l(\mathscr{C}_d)} \int_{\mathscr{C}_d} f(s)\, ds = \frac{3}{\pi^2} \int_{\mathrm{PSL}_2(\mathbb{Z})\backslash \mathrm{PSL}_2(\mathbb{R})} f(g) d\mu_g + O_\epsilon(\log T\, d^{-25/512+\epsilon} \|f\|_{W^{6,\infty}}).$$

Here $\|\cdot\|_{W^{k,p}}$ is the Sobolev norm:

$$\|f\|_{W^{k,p}} = \max_{|\alpha|\le k} \|\partial_\theta^{\alpha_1}(y\partial_x)^{\alpha_2}(y\partial_y)^{\alpha_3} f\|_{L^p}.$$

Remark 1.4. The proof of Theorem 1.1 is based on the equidistribution of the lifts of C_d in the unit tangent bundle. For this reason, one may generalize Theorem 1.1 to any surfaces and any sequence of closed geodesics whose lifts become equidistributed on the unit tangent bundle.

1A. ***Intersecting two closed geodesics.*** We now consider the number of intersections between two closed geodesics when both vary.

Theorem 1.4. *The following estimate holds uniformly in* $d_1, d_2 > 0$, *and* $0 < \theta_1 < \theta_2 < \pi$ *such that* $\theta_2 - \theta_1 \gg (d_1 d_2)^{-25/3072}$

$$\frac{I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2})}{l(C_{d_1}) l(C_{d_2})} = \frac{3}{\pi^2}\int_{\theta_1}^{\theta_2} \sin\theta \, d\theta + O_\epsilon((d_1 d_2)^{-25/6144+\epsilon}).$$

Note that if Γ is cocompact, then the modular intersection kernel coincides with the *intersection kernel* from [Lalley 2014] when $\theta = \pi$ and $\delta > 0$ is sufficiently small. However, when $\Gamma\backslash\mathbb{H}$ is noncompact, then they are never the same; for instance, we have $K_\delta^{\theta_1,\theta_2}(g, g) = \Omega(y)$ as $y \to \infty$ (Proposition 2.2). In particular, $K_\delta^{\theta_1,\theta_2}$ is not a Hilbert–Schmidt kernel, so the usual spectral theory does not apply. This is the main technical difficulty of dealing with the modular intersection kernel for noncompact quotients of $\mathbb{H}$. As it will be shown in the subsequent chapters, when both α_1 and α_2 are closed geodesics, $I_{\theta_1,\theta_2}(\alpha_1, \alpha_2)/(l(\alpha_1)l(\alpha_2))$ is the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}/(l(\alpha_1)l(\alpha_2))$ over $\alpha_1 \times \alpha_2$. When α_1 and α_2 vary over closed geodesics of length $< L$, as $L \to \infty$, we expect that the integral converges to the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}$ over $\Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$, since $\alpha_1 \times \alpha_2$ becomes equidistributed in $\Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$, as $L \to \infty$. However, unboundedness of the modular intersection kernel K causes issues of interchanging the limit and the integral. In particular, the argument of [Pollicott and Sharp 2006] using intersection form does not apply in this case. Hence, in order to prove Theorem 1.4, we study the full spectral expansion of $K_\delta^{\theta_1,\theta_2}(g_1, g_2)$. This is similar to the existing work on the weight m Selberg's trace formula [Hejhal 1976], except that we have to deal with all weights simultaneously, and that the modular intersection kernel is not diagonalizable in general. We go over this carefully in Section 5. Once the spectral expansion is obtained, the integral of $\delta^{-2}K_\delta^{\theta_1,\theta_2}$ over $\alpha_1 \times \alpha_2$ becomes a linear combination of the period integrals of the form

$$\int_{\alpha_1} \phi_1 \, ds \times \int_{\alpha_2} \phi_2 \, ds.$$

We may now use the same estimates that we use in order to prove the effective Duke's theorem to bound these, which leads to Theorem 1.4, generalizing [Pollicott and Sharp 2006] to a noncompact hyperbolic surface.

2. The modular intersection kernel

2A. *Parametrization.* Recall that $\mathrm{PSL}_2(\mathbb{R})$ acts transitively on $\mathbb{H}$ and on $S\mathbb{H}$ with the fractional transformations. For $g \in \mathrm{PSL}_2(\mathbb{R})$, $z \in \mathbb{H}$ and $u \in S\mathbb{H}$ we write these actions by gz and gu. We parametrize the points of $\mathbb{H}$ and $S\mathbb{H}$ with $x+iy$ and $(x+iy, \exp(i\theta))$. Let

$$\Pi((x+iy, \exp(i\theta))) := x+iy,$$

be the projection map from $S\mathbb{H}$ to $\mathbb{H}$.

Fix $z_0 = i$ and $u_0 = (i, \exp(i\pi/2))$. Let $g = naR_\theta \in \mathrm{PSL}_2(\mathbb{R})$ be the Iwasawa decomposition where

$$n = n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a = a(y) = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}, \quad \text{and} \quad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Then we have

$$gz_0 = x+iy \quad \text{and} \quad gu_0 = \big(x+iy, \exp\big(i\big(\tfrac{\pi}{2}+2\theta\big)\big)\big).$$

For the rest of the paper, we identify $S\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$ by sending $g \in \mathrm{PSL}_2(\mathbb{R})$ to gu_0. We often use the following fact in our computation without mentioning.

Proposition 2.1. *The image under $\gamma \in \mathrm{SL}_2(\mathbb{R})$ of the geodesic segment of length δ corresponding to $g = (x, \xi)$ is the geodesic segment of length δ corresponding to γg.*

We use the volume form given by $dV = (dx\, dy\, d\theta)/y^2$. The volume of $S\mathbb{X}$ is then $\pi^2/3$.

2B. *Preliminary estimates.* We first recall here the definition of the modular intersection kernel described in the introduction. For $\delta > 0$ and $\theta_1, \theta_2 \in (0, \pi)$, we define the integral kernel

$$k_\delta^{\theta_1,\theta_2} : S\mathbb{H} \times S\mathbb{H} \to \mathbb{R}$$

by

$$k_\delta^{\theta_1,\theta_2}((x_1, \xi_1), (x_2, \xi_2)) = 1,$$

if the geodesic segment of length δ on $\mathbb{H}$ from x_1 with the initial vector ξ_1 and the segment from x_2 with the initial vector ξ_2 intersect at an angle $\in (\theta_1, \theta_2)$, and 0 otherwise. Here the angle of the intersection of geodesic segments l_1 and l_2 at $p \in l_1 \cap l_2$ is measured counterclockwise from l_1 to l_2. Under the identification $S\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})$ from Section 2A, we note here that

$$k_\delta^{\theta_1,\theta_2}(gg_1, gg_2) = k_\delta^{\theta_1,\theta_2}(g_1, g_2)$$

for any $g, g_1, g_2 \in \mathrm{PSL}_2(\mathbb{R})$.

Now for a given discrete subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$, we define the modular intersection kernel $K_\delta^{\theta_1,\theta_2} : \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \times \Gamma\backslash \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{R}$ by taking the average of $k_\delta^{\theta_1,\theta_2}$ over Γ:

$$K_\delta^{\theta_1,\theta_2}(g_1, g_2) = \sum_{\gamma\in\Gamma} k_\delta^{\theta_1,\theta_2}(g_1, \gamma g_2).$$

Note that when Γ is cocompact, and $\delta > 0$ is less than a half of the injectivity radius of $\Gamma\backslash\mathbb{H}$, we have $K_\delta^{\theta_1,\theta_2} \leq 1$. However, when $\Gamma\backslash\mathbb{H}$ is noncompact, $K_\delta^{\theta_1,\theta_2}(g_1, g_2)$ becomes arbitrarily large near the diagonal $g_1 = g_2$ as $y_1, y_2 \to \infty$. This is illustrated in the following proposition when $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$.

Proposition 2.2. *Fix* $0 < \theta < \pi$. *Then for any* $1 > \delta > 0$, *we have*

$$K_\delta^{0,\theta}(g, g) = \Omega_\theta(\delta y).$$

Proof. Consider

$$g = (\mathrm{Re}^{i(\pi/2+\alpha(\delta))}, e^{i\alpha(\delta)}) \in S\mathbb{H},$$

where $\alpha(\delta)$ is chosen such that the geodesic segment

$$\beta_g := \left\{\mathrm{Re}^{i\theta} : \left|\theta - \tfrac{\pi}{2}\right| < \alpha(\delta)\right\} \subset \mathbb{H}$$

has length δ. Note that the length of the segment does not depend on R and that $\alpha(\delta) \sim \delta$ as $\delta \to 0$. From this, we infer that β_g and $\beta_g + n$ with $0 < n \ll R\delta$ intersect.

The angle of intersection is explicitly given by $2 \arcsin \frac{n}{R}$. So for all sufficiently small $0 < \delta < \theta$, we have

$$k_\delta^{\theta_1,\theta_2}\left(g, \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} g\right) = 1,$$

for $0 < n \ll R\delta$. This implies that

$$K_\delta^{\theta_1,\theta_2}(g, g) \gg \delta R \gg \delta y. \qquad \square$$

In view of Proposition 2.2, the following proposition provides a nice upper bound of the modular intersection kernel.

Proposition 2.3. *Let* $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ *and let* $1 > \delta > 0$. *Let* h *be a compactly supported function on* $S\mathbb{H}$, *where we assume that* $h((\,\cdot\,, \xi))$ *is supported in* $B_\delta(i)$ *for any* $\xi \in S^1$. *Define* $H : \Gamma\backslash S\mathbb{H} \times \Gamma\backslash S\mathbb{H}$ *by*

$$H(g_1, g_2) = \sum_{\gamma\in\Gamma} h(g_1^{-1}\gamma g_2)$$

for $g_1, g_2 \in \Gamma\backslash \mathrm{PSL}_2(\mathbb{R})$. *Then for* $g_i = (z_i, \xi_i)$ *with* $\mathrm{dist}_{\Gamma\backslash\mathbb{H}}(z_1, z_2) > 2\delta$, *we have*

$$H(g_1, g_2) = 0.$$

When $y_1 > 0$ *and* $y_2 > 0$ *are sufficiently large, we have*

$$H(g_1, g_2) \ll \delta\sqrt{y_1 y_2}\|h\|_{L^\infty}.$$

Proof. If $H > 0$, then there exists $\gamma \in \Gamma$ such that

$$h(g_1^{-1}\gamma g_2) > 0.$$

This implies that the balls of radius δ centered at z_1 and γz_2 intersect, hence

$$\mathrm{dist}_{\mathbb{H}}(z_1, \gamma z_2) < 2\delta,$$

which contradicts the assumption.

Now to prove the second estimate, we first note that when y_2 is sufficiently large, we have $y(\gamma g_2) < 1$ unless $\gamma = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Therefore $h(g_1^{-1}\gamma g_2) > 0$ only if $\gamma = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Note that $h(g_1^{-1}\gamma g_2) = 1$ holds only if $\mathrm{dist}_{\mathbb{H}}(z_1, n + z_2) < 2\delta$. This is equivalent to

$$\operatorname{arccosh}\left(1 + \frac{(n + x_2 - x_1)^2 + (y_1 - y_2)^2}{y_1 y_2}\right) < 2\delta,$$

and so

$$(n + x_2 - x_1)^2 < y_1 y_2(\cosh(2\delta) - 1) - (y_1 - y_2)^2 \leq y_1 y_2(\cosh(2\delta) - 1),$$

from which we infer that there are at most $\ll \delta\sqrt{y_1 y_2}$ choices of γ which makes $h(g_1, \gamma g_2) > 0$. $\square$

Now we analyze the modular intersection kernel when one variable is assumed to be contained in a compact set. We first note that if δ is less than half of the injectivity radius of g_0 in $\Gamma\backslash S\mathbb{H}$, then for each $g \in S\mathbb{H}$, there is at most one $\gamma \in \Gamma$ such that

$$k_\delta^{\theta_1,\theta_2}(g_0, \gamma g) \neq 0.$$

Therefore $K_\delta^{\theta_1,\theta_2}(g_0, \cdot\,)$ coincides with $k_\delta^{\theta_1,\theta_2}(g_0, \cdot\,)$ in the 2δ-neighborhood of g_0, which is a translation of $k_\delta^{\theta_1,\theta_2}((i, i), \cdot\,)$ around (i, i).

Lemma 2.4. *For $0 < \theta_1 < \theta_2 < \pi$, we have*

$$\int_{\mathbb{H}} k_\delta^{\theta_1,\theta_2}((i, i), g)\, dV = (\cos\theta_1 - \cos\theta_2)\delta^2.$$

Assume that $0 < \delta < 1$. Then for any $\varepsilon = o(\delta)$ and $\varepsilon = o(\theta_2 - \theta_1)$ there exist a smooth majorant $M_\delta^{\theta_1,\theta_2}$ and a smooth minorant $m_\delta^{\theta_1,\theta_2}$, i.e.,

$$0 \leq m_\delta^{\theta_1,\theta_2} \leq k_\delta^{\theta_1,\theta_2}((i, i), \cdot\,) \leq M_\delta^{\theta_1,\theta_2},$$

such that

$$\int m_\delta^{\theta_1,\theta_2}\, dV \quad \textit{and} \quad \int M_\delta^{\theta_1,\theta_2}\, dV$$

are both

$$(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon)),$$

and that

$$\|m_\delta^{\theta_1,\theta_2}\|_{W^{k,\infty}} + \|M_\delta^{\theta_1,\theta_2}\|_{W^{k,\infty}} = O_k(\varepsilon^{-k}).$$

Proof. Note that the action of the geodesic flow of time t on $S\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$ is the multiplication from the right by $a(e^t)$. For given $\varphi \in (\theta_1, \theta_2)$, we describe the collection of $g \in \mathrm{PSL}_2(\mathbb{R})$ for which the corresponding geodesic segment of length δ intersects $\{iy : e^\delta > y > 1\}$ transversally at angle φ. Note that this happens only when

$$g a(e^{t_2/2}) = \begin{cases} a(e^{t_1/2}) R_{\varphi/2}, \\ a(e^{t_1/2}) R_{(\varphi+\pi)/2}. \end{cases}$$

for some $0 < t_1, t_2 < \delta$. Hence

$$g = \begin{cases} a(e^{t_1/2}) R_{\varphi/2} a(e^{-t_2/2}), \\ a(e^{t_1/2}) R_{(\varphi+\pi)/2} a(e^{-t_2/2}). \end{cases}$$

Consider $\Psi : AKA \to \mathrm{PSL}_2(\mathbb{R})$ given by

$$(t_1, \varphi, t_2) \mapsto a(e^{t_1/2}) R_{\varphi/2} a(e^{-t_2/2})$$

The determinant of the Jacobian of Ψ is a nonzero multiple of $|\sin\varphi|$ (we refer the readers to the Appendix for the computation), and so this defines a local diffeomorphism away from $\varphi = 0$ and π. Observe that Ψ is injective away from $\varphi = 0$ and π. From this we infer that the support of $k_\delta^{\theta_1,\theta_2}((i,i), g)$ is the image of the open box

$$\{(t_1, \varphi, t_2) : 0 < t_1, t_2 < \delta, \theta_1 < \varphi < \theta_2 \text{ or } \theta_1 + \pi < \varphi < \theta_2 + \pi\}$$

under Ψ, and

$$\begin{aligned} \int_{\mathbb{H}} k_\delta^{\theta_1,\theta_2}((i,i), g)\, dV &= \tfrac{1}{2} \int_0^\delta \int_0^\delta \int_{\theta_1}^{\theta_2} |\sin(\varphi)|\, d\varphi\, dt_1\, dt_2 + \tfrac{1}{2} \int_0^\delta \int_0^\delta \int_{\theta_1+\pi}^{\theta_2+\pi} |\sin(\varphi)|\, d\varphi\, dt_1\, dt_2 \\ &= (\cos\theta_1 - \cos\theta_2)\delta^2, \end{aligned}$$

where we used $dV = \frac{1}{2}|\sin\varphi|\, d\varphi\, dt_1\, dt_2$ ((A-1)).

Note that the support of $k_\delta^{\theta_1,\theta_2}((i,i), \cdot\,)$ is an open set which has a piecewise smooth boundary. Therefore, under the assumption that $\varepsilon = o(\delta)$ and $\varepsilon = o(\theta_2 - \theta_1)$, there exist smooth majorant and minorant whose L^1 norms are $(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon))$, and whose k-th derivatives are $O_k(\varepsilon^{-k})$. $\square$

As an immediate application, we have the following corollary.

Corollary 2.5. *Fix a compact subset $C \subset \Gamma\backslash S\mathbb{H}$, and assume that δ is less than the half of the infimum of injectivity radius of $g \in C$ in $\Gamma\backslash S\mathbb{H}$. Then for any given compact geodesic segment $\beta \subset C$, and for any given $\varepsilon > 0$ which is $o(\delta)$ and $o(\theta_2 - \theta_1)$,*

$$\int_\beta K_\delta^{\theta_1,\theta_2}(s, \cdot\,)\, ds$$

admits a smooth majorant $M_{\beta,\delta}^{\theta_1,\theta_2}$ and a smooth minorant $m_{\beta,\delta}^{\theta_1,\theta_2}$ such that

$$\|m_{\beta,\delta}^{\theta_1,\theta_2}\|_{L^1}, \|M_{\beta,\delta}^{\theta_1,\theta_2}\|_{L^1} = l(\beta)(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon)),$$

and that

$$\|m_{\beta,\delta}^{\theta_1,\theta_2}\|_{W^{k,\infty}} + \|M_{\beta,\delta}^{\theta_1,\theta_2 i}\|_{W^{k,\infty}} = O_k(l(\beta)\varepsilon^{-k}).$$

2C. *Intersection numbers.* In this section, we prove formulas relating the number of intersections between two geodesics to the integral of the modular intersection kernel over the two geodesics.

Lemma 2.6. *Let $\alpha_i = \{\alpha_i(t) : t \in [0, l(\alpha_i))\}$ be closed geodesics in $\Gamma\backslash\mathbb{H}$ parametrized by the arc length, and let $\widetilde{\alpha}_i = \{(\alpha_i(t), \alpha_i'(t)) : t \in [0, l(\alpha_i))\} \subset S\mathbb{H}$ be the lifts of α_i for $i = 1, 2$. Then for any $\delta > 0$,*

$$I_{\theta_1,\theta_2}(\alpha_1, \alpha_2) = \frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2.$$

Remark 2.1. For each α_i, there are two choices of parametrization by the arc length, namely $\alpha_i(t)$ and $\alpha_i(-t)$, but the integral does not depend on the choice of the parametrizations.

Proof. By abuse of notations, we think of each α_i with $t \in [0, l(\alpha_i))$ a geodesic segment in $\mathbb{H}$ and accordingly $\widetilde{\alpha}_i$ a corresponding curve in $S\mathbb{H}$. For a geodesic segment $\alpha \subset \mathbb{H}$ parametrized by $t \in [a, b]$, let $[\alpha] \subset \mathbb{H}$ be the biinfinite geodesic $\{\alpha(t) : t \in \mathbb{R}\}$ that contains α. Then we express the integral as follows:

$$\begin{aligned}
\frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 &= \sum_{\gamma\in\Gamma} \frac{1}{\delta^2} \int_{\gamma\widetilde{\alpha}_2} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma/\Gamma_{[\alpha_2]}} \frac{1}{\delta^2} \int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}} \sum_{\gamma'\in\Gamma_{[\alpha_1]}} \frac{1}{\delta^2} \int_{\gamma'\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{\alpha}_1} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\
&= \sum_{\gamma\in\Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}} \frac{1}{\delta^2} \int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{[\alpha_1]}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2.
\end{aligned}$$

Here $\Gamma_{[\alpha_i]}$ is the stabilizer subgroup of Γ with respect to $[\alpha_i]$.

Now because two geodesics in $\mathbb{H}$ may intersect at most once, for each intersection point $p \in \alpha_1 \cap \alpha_2$ on $\Gamma\backslash\mathbb{H}$, there exists a unique $\gamma \in \Gamma/\Gamma_{[\alpha_2]}$ such that α_1 and $\gamma[\alpha_2]$ intersect at a lift of p. Also, because $[\alpha_1]$ is a disjoint union of $\gamma'\alpha_1$ with $\gamma' \in \Gamma_{[\alpha_1]}$, each $\{\gamma'\gamma : \gamma' \in \Gamma_{[\alpha_1]}\}$ contains at most one $\gamma'\gamma$ such that $\gamma'\gamma[\alpha_2]$ intersects α_1.

Therefore the intersections of α_1 and α_2 are in one-to-one correspondence with $\gamma \in \Gamma_{[\alpha_1]}\backslash\Gamma/\Gamma_{[\alpha_2]}$ such that $\gamma[\alpha_2]$ intersects $[\alpha_1]$. We complete the proof by observing that

$$\int_{\gamma\widetilde{[\alpha_2]}} \int_{\widetilde{[\alpha_1]}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = 1,$$

if $[\alpha_1]$ and $\gamma[\alpha_2]$ intersect at an angle $\in (\theta_1, \theta_2)$, and $= 0$ otherwise. □

Now let $\beta = \{\beta(t) : t \in [0, l(\beta))\}$ be a compact geodesic segment in $\Gamma\backslash\mathbb{H}$, and let α_2 be a closed geodesic as before. Then

$$\frac{1}{\delta^2} \int_{\widetilde{\alpha}_2} \int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2$$

does not always give $I(\beta, \alpha_2)$. Instead, it is a weighted sum over the intersections of $\beta_0 := \{\beta(t) : t \in [0, l(\beta) + \delta)\}$ and α_2. We prove the following.

Lemma 2.7. *With the same notations as above, assume that $0 < \delta < l(\beta)$ and that β_0 has no self intersection. For $0 < \theta_1 < \theta_2 < \pi$, let $S(\beta_0, \alpha_2)_{\theta_1,\theta_2}$ be the set of intersections between β_0 and α_2 where the intersection angle is $\in (\theta_1, \theta_2)$. Then we have*

$$\frac{1}{\delta^2}\int_{\tilde{\alpha}_2}\int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = \sum_{p\in S(\beta_0,\alpha_2)_{\theta_1,\theta_2}} \min\left\{\frac{\beta^{-1}(p)}{\delta}, 1, \frac{l(\beta)+\delta-\beta^{-1}(p)}{\delta}\right\}.$$

Proof. As in the proof of Lemma 2.6, we first have

$$\begin{aligned}\frac{1}{\delta^2}\int_{\tilde{\alpha}_2}\int_{\tilde{\beta}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 &= \sum_{\gamma\in\Gamma}\frac{1}{\delta^2}\int_{\gamma\tilde{\alpha}_2}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2\\ &= \sum_{\gamma\in\Gamma/\Gamma_{[\alpha_2]}}\frac{1}{\delta^2}\int_{\gamma\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2.\end{aligned}$$

Note that because we assumed that β_0 has no self-intersection, $p \in S(\beta_0, \alpha_2)_{\theta_1,\theta_2}$ is in one-to-one correspondence with $\gamma \in \Gamma/\Gamma_{[\alpha_2]}$ such that β_0 and $\gamma\widetilde{[\alpha_2]}$ intersect at p at an angle $\in (\theta_1, \theta_2)$. We denote by γ_p the γ corresponding to p. Observe that

$$\int_{\gamma\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = 0,$$

if $\gamma\widetilde{[\alpha_2]} \cap \beta_0 = \varnothing$. So it is sufficient to prove that

$$\frac{1}{\delta^2}\int_{\gamma_p\widetilde{[\alpha_2]}}\int_{\tilde{\beta}} k_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = \min\left\{\frac{\beta^{-1}(p)}{\delta}, 1, \frac{l(\beta)+\delta-\beta^{-1}(p)}{\delta}\right\}.$$

This follows by observing that

$$k_\delta^{\theta_1,\theta_2}((\beta(t_1), \beta'(t_1)), (\gamma_p\alpha_2(t_2), (\gamma_p\alpha_2)'(t_2))) = 1$$

for

$$(t_1, t_2) \in (\beta^{-1}(p) - \delta, \beta^{-1}(p)) \times (\alpha_2^{-1}(p) - \delta, \alpha_2^{-1}(p)),$$

and 0 otherwise, whereas the integral over $\tilde{\beta}$ is over the range $t_1 \in (0, l(\beta))$. $\square$

3. Spectral theory

3A. ***Spectral expansion.*** We first go over the spectral decomposition of $L^2(S\mathbb{X})$. Readers may find more details on the subject in [Kubota 1973; Lang 1985]. On $G = \mathrm{PSL}_2(\mathbb{R})$, there is a differential operator of order 2 that commutes with the G action,

$$\Omega = y^2\partial_x^2 + y^2\partial_y^2 + y\partial_x\partial_\theta,$$

which is called the Casimir operator. An equivariant eigenfunction of Ω is a function $f \in C^\infty(S\mathbb{X})$ that satisfies

$$\Omega f = \lambda f$$

for some $\lambda \in \mathbb{R}$, and

$$f(gR_\theta) = e^{-im\theta} f(g) \tag{3-1}$$

for some $m \in 2\mathbb{Z}$. We say that a function has weight m if it satisfies (3-1).

Each irreducible (cuspidal) subrepresentation of the right regular representation

$$\rho_g : f(h) \mapsto f(hg)$$

on $L^2(S\mathbb{X})$ is generated by an equivariant eigenfunction of Ω.

We let $\boldsymbol{E}^+$ and $\boldsymbol{E}^-$ to be the raising and lowering operator acting on equivariant functions on $L^2(S\mathbb{X})$, which are given by [Jakobson 1994]

$$\boldsymbol{E}^+ = e^{-2i\theta}(2iy\partial_x + 2y\partial_y + i\partial_\theta) \quad \text{and} \quad \boldsymbol{E}^- = e^{2i\theta}(2iy\partial_x - 2y\partial_y + i\partial_\theta). \tag{3-2}$$

Note that $\boldsymbol{E}^+$ (resp. $\boldsymbol{E}^-$) maps a weight m eigenfunction of Ω to a weight $m+2$ (resp. $m-2$) eigenfunction of Ω.

For an even integer m let

$$\psi_{s,m}(g) = y^s e^{-im\theta}.$$

Note that $\psi_{s,m}$ is invariant under the action of the unipotent upper triangular matrices. The weight m Eisenstein series is then given by

$$E_m(g,s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \psi_{s,m}(\gamma g),$$

where $\Gamma_\infty = \left\{ \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z} \right\}$ is the stabilizer subgroup of Γ with respect to the cusp $i\infty$. Although the right-hand side of the equation is absolutely convergent only for $\mathrm{Re}(s) > 1$, the weight m Eisenstein series has a meromorphic continuation to the entire complex plane.

Let Θ be the closure of

$$\left\{ \int_{-\infty}^{\infty} h(t) E_m\left(g, \tfrac{1}{2} + it\right) dt : h(t) \in C_0^\infty(\mathbb{R}), m \in 2\mathbb{Z} \right\}$$

in $L^2(S\mathbb{X})$, and let

$$L^2_{\mathrm{cusp}}(S\mathbb{X}) = \left\{ f \in L^2(S\mathbb{X}) : \int_0^1 f(n(x)g)\, dx = 0 \text{ for almost every } g \in S\mathbb{X} \right\}$$

be the space of cusp forms. Then we have the decomposition

$$L^2(S\mathbb{X}) = \langle \{1\} \rangle \oplus \Theta \oplus L^2_{\mathrm{cusp}}(S\mathbb{X}),$$

where $\langle \{1\} \rangle$ is the subspace spanned by a constant function.

We express the cuspidal subspace as a direct sum of subspaces generated by Maass forms and modular forms as in [Luo et al. 2009, (1.10)],

$$L^2_{\text{cusp}}(S\mathbb{X}) = \sum_{j=1}^{\infty} W_{\pi_j^0} \bigoplus \sum_{m\geq 12} \sum_{j=1}^{d_m} (W_{\pi_j^m} \oplus W_{\pi_j^{-m}}),$$

where each $W_{\pi_j^m}$ corresponds to a G and Hecke irreducible subspace of a right regular representation on L^2_{cusp}. Here d_m is the dimension of the space of holomorphic cusp forms of weight m for $\mathrm{PSL}_2(\mathbb{Z})$. Each π_j^0 corresponds to a Maass–Hecke cusp form which we denote by ϕ_j^0. For $m > 0$, π_j^m corresponds to a holomorphic Hecke cusp form ϕ_j^m. We identify a weight m function on $\Gamma\backslash\mathbb{H}$

$$f(\gamma z) = (cz+d)^m f(z) \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \gamma \in \Gamma$$

with a weight m Γ-invariant function F on $\mathrm{PSL}_2(\mathbb{R})$ via

$$F(g) = y^{m/2} f(z) e^{-im\theta}. \tag{3-3}$$

When $m \geq 0$, viewing ϕ_j^m as a function on $S\mathbb{X}$, each $W_{\pi_j^m}$ is spanned by

$$\dots, \quad (\boldsymbol{E}^-)^3\phi_j^m, \quad (\boldsymbol{E}^-)^2\phi_j^m, \quad \boldsymbol{E}^-\phi_j^m, \quad \phi_j^m, \quad \boldsymbol{E}^+\phi_j^m, \quad (\boldsymbol{E}^+)^2\phi_j^m, \quad (\boldsymbol{E}^+)^3\phi_j^m, \quad \dots$$

Note that when $m > 0$, $\boldsymbol{E}^-\phi_j^m = 0$.

For $m < 0$, we set

$$W_{\pi_j^{-m}} = \overline{W_{\pi_j^m}} = \{\bar{f} : f \in W_{\pi_j^m}\}.$$

Now let

$$U_{\pi_j^0} = W_{\pi_j^0} \quad \text{and} \quad U_{\pi_j^m} = W_{\pi_j^m} \oplus W_{\pi_j^{-m}},$$

when $m > 0$. We specify an orthonormal basis of each $U_{\pi_j^m}$ as follows.

The Maass cusp form case $m = 0$: Let $-\left(\frac{1}{4} + t_j^2\right)$ be the Laplacian eigenvalue of ϕ_j^0,[†] for some real t_j. We set $\phi_{j,0}^0 = \phi_j^0$, and define $\phi_{j,l}^0$ for $l \in 2\mathbb{Z}$ inductively by

$$\boldsymbol{E}^-\phi_{j,l}^0 = (l+1-2it_j)\phi_{j,l-2}^0 \quad \text{and} \quad \boldsymbol{E}^+\phi_{j,l}^0 = (l+1+2it_j)\phi_{j,l+2}^0. \tag{3-4}$$

The holomorphic Hecke cusp form case $m > 0$: We set $\phi_{j,m}^m = \phi_j^m$ and $\phi_{j,-m}^m = \overline{\phi_j^m}$, and define $\phi_{j,l}^m$ for $l \in 2\mathbb{Z}$ inductively by

$$\boldsymbol{E}^-\phi_{j,l}^m = (l-m)\phi_{j,l-2}^m \quad \text{and} \quad \boldsymbol{E}^+\phi_{j,l}^m = (l+m)\phi_{j,l+2}^m. \tag{3-5}$$

Finally, note that we have the following relation among the weight m Eisenstein series.

$$\boldsymbol{E}^- E_m\left(g, \tfrac{1}{2}+it\right) = (m+1-2it)E_{m-2}\left(g, \tfrac{1}{2}+it\right), \quad \text{and}$$
$$\boldsymbol{E}^+ E_m\left(g, \tfrac{1}{2}+it\right) = (m+1+2it)E_{m+2}\left(g, \tfrac{1}{2}+it\right).$$

[†]Formally, it is the eigenvalue of the Laplace–Beltrami operator on $\mathbb{X}$ that corresponds to ϕ_j^0.

With these notations, we have:

Proposition 3.1. *Let $f \in L^2(S\mathbb{X})$. Then we have*

$$f(g)=\frac{3}{\pi^2}\int_{S\mathbb{X}} f(g_1)\,dg_1+\sum_{\substack{m\geq 0\\ 2\,|\,m}}\sum_{j=1}^{d_m}\sum_{\substack{l\in 2\mathbb{Z}\\ |l|\geq m}}\langle f,\phi_{j,l}^m\rangle_{S\mathbb{X}}\phi_{j,l}^m(g)+\sum_{m\in 2\mathbb{Z}}\frac{1}{4\pi}\int_{-\infty}^{\infty}\bigl\langle f,E_m\bigl(\cdot,\tfrac12+it\bigr)\bigr\rangle_{S\mathbb{X}}E_m\bigl(g,\tfrac12+it\bigr)\,dt,$$

where we set $d_0=+\infty$.

4. Effective equidistribution

4A. *Invariant linear form.* Define μ_d to be the integral over discriminant d oriented closed geodesics on $S\mathbb{X}$,

$$\mu_d(F):=\int_{\mathscr{C}_d} F(s)\,ds=\sum_{\mathrm{disc}(q)=d}\int_{C(q)} F(s)\,ds.$$

where $C(q)\subset S\mathbb{X}$ is the oriented closed geodesic associated to the binary quadratic form q [Luo et al. 2009, 2.3]. Then for any $F\in U_{\pi_j^m}$, we have

$$\mu_d(F)=\mu_d(\phi_j^m)\eta_j^m(F)$$

for some linear form η_j^m on $U_{\pi_j^m}$ invariant under the diagonal action [loc. cit., Section 3.7.1], which we describe below following [loc. cit., Section 3.2]. (Note that the parameter s in [loc. cit.] is replaced by $2it$ in this article for consistency.)

The Maass cusp form case $m=0$: Let $\phi_{j,l}^0$ be the Maass form defined by (3-4). When $4\,|\,l$ and $l\geq 4$, we have

$$\eta_j^0(\phi_{j,l}^0)=\eta_j^0(\phi_{j,-l}^0)=\frac{(1-2it_j)(5-2it_j)\cdots(l-3-2it_j)}{(3+2it_j)(7+2it_j)\cdots(l-1+2it_j)},\tag{4-1}$$

and $\eta_j^0(\phi_{j,l}^0)$ is identically 0 if $l\equiv 2 \pmod 4$. Note that $\{\phi_{j,l}^0\}_{l\in 2\mathbb{Z}}$ is an orthogonal basis of $U_{\pi_j^0}$, and normalized so that,

$$\|\phi_{j,l}^0\|_{L^2}=\|\phi_j^0\|_{L^2}.$$

The holomorphic Hecke cusp form case $m>0$: Let $\phi_{j,l}^m$ be the holomorphic Hecke cusp form defined by (3-5). When $m\equiv 2 \pmod 4$, η_j^m is identically 0.

When $m\equiv 0 \pmod 4$, for $l\geq 4$ with $4\,|\,l$,

$$\eta_j^m(\phi_{j,m+l}^m)=\eta_j^m(\phi_{j,-m-l}^m)=\frac{1\cdot 3\cdot 5\cdots(l/2-1)}{(m+1)(m+3)\cdots(m+l/2-1)},\tag{4-2}$$

and $\eta_j^m(\phi_{j,m+l}^m)$ vanishes for $l\equiv 2 \pmod 4$.

Note that $\{\phi_{j,l}^m\}_{l\in 2\mathbb{Z},|l|\geq m}$ is an orthogonal basis of $U_{\pi_j^m}$, and normalized so that

$$\|\phi_{j,l}^m\|_{L^2}=\|\phi_j^m\|_{L^2}.$$

for $l\in 2\mathbb{Z}$, $|l|\geq m$.

Eisenstein series case: By the above identities and following [Luo et al. 2009, Section 3], we have

$$\mu_d\big(E_m\big(g,\tfrac{1}{2}+it\big)\big)=\eta(m,t)\mu_d\big(E_0\big(g,\tfrac{1}{2}+it\big)\big),$$

where for $m\geq 4$ such that $4\mid m$,

$$\eta(m,t)=\eta(-m,t)=\frac{(1-2it)(5-2it)\cdots(2m-3-2it)}{(3+2it)(7+2it)\cdots(2m-1+2it)},\tag{4-3}$$

and $\eta(m,t)$ is identically 0 if $m\equiv 2 \pmod 4$.

4B. *Period integrals.*

4B1. *Holomorphic cusp forms.* In this section, we give an upper bound on the period integrals of holomorphic forms. We first use the results of Shintani to relate the period integrals of holomorphic cusp forms to the Fourier coefficients of half integral holomorphic forms. We then apply the result of Kohnen and Zagier [1981] which gives an explicit version of the Waldspurger's formula for the Fourier coefficients of half integral holomorphic forms. An upper bound on these period integrals is deduced by using the subconvexity bounds on the central value of the L-functions and the Ramanujan bound on the Fourier coefficients of holomorphic modular forms.

Note that $c(d)$ is identically zero when $m\equiv 2 \pmod 4$, and so we assume that $4\mid m$. Let $\hat{\phi}_j^m$ be a normalization of the Hecke holomorphic cusp form ϕ_j^m of weight m such that $a_1=1$. Let

$$c(d):=\sum_{\mathrm{disc}(q)=d}\int_{C(q)}\hat{\phi}_j^m(z)q(z,1)^{m/2-1}\,dz,$$

where $\hat{\phi}_j^m(z)$ is the associated holomorphic modular form defined on the upper half plane and the integration is on the upper half plane (3-3). By [Luo et al. 2009, (2.4) page 14], we have

$$|c(d)|=|d|^{m/4-1/2}|\mu_d(\hat{\phi}_j^m)|.\tag{4-4}$$

Let

$$\theta(z,\hat{\phi}_j^m):=\sum_{d\geq 1}c(d)e(dz).$$

By [Shintani 1975, Theorem 2], $\theta(z,\phi_j^m)$ is a Hecke holomorphic cusp form of weight $(m+1)/2$ and level $\Gamma_0(4)$. By [Luo et al. 2009, (6.2), page 37], we have the following explicit version of Rallis inner product formula

$$\langle\theta(\hat{\phi}_j^m),\theta(\hat{\phi}_j^m)\rangle=\frac{(m/2-1)!}{2^m\pi^{m/2}}L\big(\tfrac{1}{2},\phi_j^m\big)\langle\hat{\phi}_j^m,\hat{\phi}_j^m\rangle.$$

Suppose that $d=Db^2$ with D a fundamental discriminant. By [Kohnen and Zagier 1981, Theorem 1], for D a fundamental discriminant with $D>0$ and $4\mid m$, we have

$$\frac{c(D)^2}{\langle\theta(\hat{\phi}_j^m),\theta(\hat{\phi}_j^m)\rangle}=\frac{(m/2-1)!}{\pi^{m/2}}D^{(m-1)/2}\frac{L(1/2,\phi_j^m\otimes\chi_D)}{\langle\hat{\phi}_j^m,\hat{\phi}_j^m\rangle},$$

which implies that

$$|c(D)| = D^{(m-1)/4} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2}.$$

By using the Ramanujan bound on the Fourier coefficients of integral weight cusp forms and the above, we have

$$|c(d)| \ll_\epsilon b^{m-1/2+\epsilon} |c(D)| \ll_\epsilon d^{m-1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2},$$

and so

$$|\mu_d(\hat{\phi}_j^m)| \ll_\epsilon |d|^{1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} \big(L\big(\tfrac{1}{2}, \phi_j^m\big) L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big)\big)^{1/2},$$

by (4-4).

We now use the convexity bound

$$L\big(\tfrac{1}{2}, \phi_j^m\big) \ll_\epsilon m^{1/2+\epsilon},$$

and the subconvexity bound [Blomer et al. 2007, Theorem 1]

$$L\big(\tfrac{1}{2}, \phi_j^m \otimes \chi_D\big) \ll_\epsilon m^{(75+12\theta)/16} D^{1/2-(1/8)(1-2\theta)+\epsilon},$$

where $\theta = \frac{7}{64}$ is the best exponent toward Ramanujan conjecture for Maass forms, to see that

$$|\mu_d(\hat{\phi}_j^m)| \ll_\epsilon d^{1/4+\epsilon} \frac{(m/2-1)!}{2^{m/2}\pi^{m/2}} m^{2.64} D^{1/4-25/512}.$$

It is well-known that

$$\langle \hat{\phi}_j^m, \hat{\phi}_j^m \rangle = \frac{\Gamma(m)}{(4\pi)^m} L(1, \operatorname{sym}^2 \phi_j^m)$$

up to a constant. Hence, by Stirling's approximation

$$|\mu_d(\phi_j^m)| \ll_\epsilon d^{1/4+\epsilon} m^{2.9} D^{1/4-25/512} \ll d^{1/2-25/512+\epsilon} m^{2.9}. \tag{4-5}$$

4B2. *Maass forms.* In this section, we give an upper bound on the period integrals of Maass forms. We first recall some results of Katok and Sarnak [1993] that generalize the work of Shintani [1975] to Maass forms and related the period integrals to the Fourier coefficients of half integral weight Maass forms. Then we use an explicit version of the Waldspurger formula [Baruch and Mao 2010] and give a nontrivial bound on these period integrals by using the subconvexity bound on the central value of the L-functions and the best bound toward Ramanujan conjecture for Maass forms.

Let ϕ_j^0 be a Hecke–Maass form with $\langle \phi_j^0, \phi_j^0 \rangle = 1$ and with the Laplacian eigenvalue $-\big(\frac{1}{4} + t_j^2\big)$. For $d > 0$, let

$$\rho(d) := \frac{1}{\sqrt{8}\pi^{1/4} d^{3/4}} \sum_{\operatorname{disc}(q)=d} \int_{C(q)} \phi_j^0 \, ds$$

be the associated period integral, and let

$$\theta((u+iv),\phi_j^0) := \sum_{d\neq 0} \rho(d) W_{\operatorname{sgn}(d)/4, it_j/2}(4\pi|d|v)e(du),$$

where $W_{\operatorname{sgn}(d)/4,it_j/2}$ is the usual Whittaker function. Here $\rho(d)$ for $d<0$ is the sum of ϕ_j^0 over the CM points with the discriminant d appropriately normalized; see [Katok and Sarnak 1993, page 197] or [Sardari 2021, Section 3.3] for a detailed discussion.

Note from [Katok and Sarnak 1993] that $\theta((u+iv),\phi_j^0)$ is a weight $\frac{1}{2}$ Hecke–Maass form with the Laplacian eigenvalue $-\bigl(\frac{1}{4}+\frac{t_j^2}{4}\bigr)$. By [Katok and Sarnak 1993, (5.6), page 224] or [Luo et al. 2009, (6.4), page 38], we have the following version of the Rallis inner product formula

$$\langle \theta(\phi_j^0), \theta(\phi_j^0)\rangle = \tfrac{3}{2}\Lambda\bigl(\tfrac{1}{2},\phi_j^0\bigr),$$

where

$$\Lambda(s,\phi_j^0) = \pi^{-s}\Gamma\left(\frac{s+it_j}{2}\right)\Gamma\left(\frac{s-it_j}{2}\right)L(s,\phi_j^0)$$

is the completed L-function.

By an explicit form of Waldspurger formula [Baruch and Mao 2010, Theorem 1.4], and the best exponent toward the Ramanujan conjecture [Lester and Radziwiłł 2020, Corollary 6.1], we have

$$\frac{\rho(d)}{\langle \theta(\phi_j^0),\theta(\phi_j^0)\rangle^{1/2}} \ll_\epsilon \frac{1}{\sqrt{|d|}}\left(\frac{L(1/2,\phi_j^0\otimes\chi_D)}{L(1,\operatorname{sym}^2\phi_j^0)}\right)^{1/2} b^{7/64+\epsilon}|t_j|^{-\operatorname{sgn}(d)/4}e^{\pi|t_j|/4},$$

where $d=Db^2$ with D a fundamental discriminant. Note from Stirling's formula that

$$\Gamma\left(\frac{1/2+it_j}{2}\right)\Gamma\left(\frac{1/2-it_j}{2}\right) \ll |t_j|^{-1/2}e^{-\pi|t_j|/2},$$

from which we infer that

$$\begin{aligned}
\mu_d(\phi_j^0) &\ll d^{3/4}|\rho(d)|\\
&\ll_\epsilon d^{1/4}\bigl(\Lambda\bigl(\tfrac{1}{2},\phi_j^0\bigr)\bigr)^{1/2}\left(\frac{L(1/2,\phi_j^0\otimes\chi_D)}{L(1,\operatorname{sym}^2\phi_j^0)}\right)^{1/2} b^{7/64+\epsilon}|t_j|^{-\operatorname{sgn}(d)/4}e^{\pi|t_j|/4}\\
&\ll_\epsilon d^{1/4}\bigl(L\bigl(\tfrac{1}{2},\phi_j^0\bigr)L\bigl(\tfrac{1}{2},\phi_j^0\otimes\chi_D\bigr)\bigr)^{1/2} b^{7/64+\epsilon}|t_j|^{-((\operatorname{sgn}(d)+1)/4)+\epsilon}.
\end{aligned}$$

We now use the convexity bound,

$$L\bigl(\tfrac{1}{2},\phi_j^0\bigr) \ll_\epsilon |t_j|^{1/2+\epsilon},$$

and the subconvexity bound [Blomer et al. 2007, Theorem 1],

$$L\bigl(\tfrac{1}{2},\phi_j^0\otimes\chi_D\bigr) \ll_\epsilon |t_j|^{(31+4\theta+\epsilon)/16}D^{1/2-(1-2\theta)/8+\epsilon},$$

to conclude that

$$\mu_d(\phi_j^0) \ll_\epsilon d^{1/4+\epsilon}|t_j|^{3/4}b^{7/64+\epsilon}D^{1/4-25/512} \ll d^{1/2-25/512+\epsilon}|t_j|^{3/4}. \tag{4-6}$$

4B3. *Eisenstein series.* For a nonsquare integer $d \equiv 0, 1 \pmod 4$, let $d = Db^2$ where D is a fundamental discriminant. Then we have the following explicit formula for the period integral of the Eisenstein series [Zagier 1981, page 282]:[†]

$$\mu_d(E_0(\cdot\,, s)) = \frac{\Gamma(s/2)^2 d^{s/2} L(s, d)}{\Gamma(s)\zeta(2s)}, \tag{4-7}$$

where

$$L(s, d) = L(s, \chi_D)\Biggl(\sum_{a\,|\,b} \mu(a)\bigl(\tfrac{D}{a}\bigr) a^{-s} \sigma_{1-2s}\bigl(\tfrac{b}{a}\bigr)\Biggr). \tag{4-8}$$

Here $L(s, \chi_D)$ is the Dirichlet L-function attached to the quadratic Dirichlet character $\chi_D(\cdot) = \bigl(\frac{D}{\cdot}\bigr)$, $\mu(\cdot)$ is the Möbius function, and $\sigma_v(\cdot) = \sum_{a\,|\,\cdot} a^v$ is the divisor function.

Now assume that $s = \frac{1}{2} + it$ for some $t \in \mathbb{R}$. By Stirling's formula, we have

$$\frac{\Gamma(s/2)^2}{\Gamma(s)} \ll |t|^{-1/2}.$$

By the zero free region of $\zeta(2s)$ around $2s = 1 + 2it$, we have

$$|\zeta(2s)| \gg_\epsilon t^{-\epsilon}.$$

We also have the convexity bound

$$\zeta(s) \ll |t|^{1/4},$$

and we know from [Heath-Brown 1980] that

$$L\bigl(\tfrac{1}{2} + it, \chi_D\bigr) \ll_\epsilon ((|t| + 1)D)^{3/16+\epsilon}.$$

Finally, observe that we have

$$\sum_{a\,|\,b} \mu(a)\bigl(\tfrac{D}{a}\bigr) a^{-s} \sigma_{1-2s}\bigl(\tfrac{b}{a}\bigr) \ll_\epsilon d^\epsilon.$$

Combining all these estimates, we deduce the following estimate from (4-7) for $s = \frac{1}{2} + it$:

$$\mu_d(E_0(\cdot\,, s)) \ll_\epsilon d^{1/2-1/16+\epsilon}. \tag{4-9}$$

4C. ***Proof of Theorem 1.3.*** For any compactly supported smooth function $F \in C_0^\infty(S\mathbb{X})$, recall from Proposition 3.1 that we have

$$F(g) = \frac{3}{\pi^2}\int_{S\mathbb{X}} F(g_1)\,dg_1 + \sum_{\substack{m\ge 0\\ 2\,|\,m}} \sum_{j=1}^{d_m} \sum_{\substack{l\in 2\mathbb{Z}\\ |l|\ge m}} \langle F, \phi^m_{j,l}\rangle_{S\mathbb{X}} \phi^m_{j,l}(g) + \sum_{m\in 2\mathbb{Z}} \frac{1}{4\pi}\int_{-\infty}^{\infty} \bigl\langle F, E_m\bigl(\cdot\,, \tfrac{1}{2}+it\bigr)\bigr\rangle_{S\mathbb{X}} E_m\bigl(g, \tfrac{1}{2}+it\bigr)\,dt,$$

[†]When $b = 1$, this is a classical result due to Hecke [Siegel 1965, page 88].

and so from the discussion of Section 4A, we have

$$\mu_d(F)=\mu_d\left(\frac{3}{\pi^2}\right)\int_{S\mathbb{X}}F(g)\,dg+\sum_{\substack{m\geq 0\\ 4\mid m}}\sum_{j=1}^{d_m}\mu_d(\phi_j^m)\sum_{\substack{l\in 4\mathbb{Z}\\ |l|\geq m}}\langle F,\phi_{j,l}^m\rangle_{S\mathbb{X}}\eta_j^m(\phi_{j,l}^m)\\ +\sum_{m\in 4\mathbb{Z}}\frac{1}{4\pi}\int_{-\infty}^{\infty}\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\eta\big(m,\tfrac12+it\big)\mu_d\big(E_0\big(\cdot,\tfrac12+it\big)\big)\,dt.$$

Firstly, we have from (4-1), (4-2), and (4-3) that $\eta_j^m(\phi_{j,l}^m)$ and $\eta\big(m,\frac12+it\big)$ are both $O(1)$. Note by successive integration by parts and Cauchy–Schwarz inequality, we have for all $N\geq 1$,

$$\langle F,\phi_{j,l}^m\rangle\ll_N(|l|^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})},$$

when $m>0$, and

$$\langle F,\phi_{j,l}^0\rangle\ll_N(|l|^2+|t_j|^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})}.$$

Likewise, assuming that the support of F is contained in $y<T$, we have

$$\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\ll_N(|m|^2+t^2+1)^{-N}\|F\|_{W^{2N,2}(S\mathbb{X})}\log T,$$

where we used [Kubota 1973, (6.1.6)] and [Jakobson 1994, (1.6), (1.7)].

Now for $m>0$, we take $N=3$ and apply (4-5) to see that

$$\sum_{\substack{m>0\\ 4\mid m}}\sum_{j=1}^{d_m}\mu_d(\phi_j^m)\sum_{\substack{l\in 4\mathbb{Z}\\ |l|\geq m}}\langle F,\phi_{j,l}^m\rangle_{S\mathbb{X}}\eta_j^m(\phi_{j,l}^m)\ll_\epsilon d^{1/2-25/512+\epsilon}\|F\|_{W^{6,2}(S\mathbb{X})},$$

and for $m=0$, we take $N=2$ and apply (4-6) to deduce

$$\sum_{j=1}^{\infty}\mu_d(\phi_j^0)\sum_{l\in 4\mathbb{Z}}\langle F,\phi_{j,l}^0\rangle_{S\mathbb{X}}\eta_j^0(\phi_{j,l}^0)\ll_\epsilon d^{1/2-25/512+\epsilon}\|F\|_{W^{4,2}(S\mathbb{X})}.$$

For the Eisenstein series contribution, we take $N=2$ and apply (4-9) to see

$$\sum_{m\in 4\mathbb{Z}}\frac{1}{4\pi}\int_{-\infty}^{\infty}\big\langle F,E_m\big(\cdot,\tfrac12+it\big)\big\rangle_{S\mathbb{X}}\eta\big(m,\tfrac12+it\big)\mu_d\big(E_0\big(\cdot,\tfrac12+it\big)\big)dt\ll_\epsilon\log T\,d^{7/16+\epsilon}\|F\|_{W^{4,2}(S\mathbb{X})}.$$

Therefore Theorem 1.3 will follow once we establish the following lower bound for the total length of $\mathscr{C}_d$:

$$l(\mathscr{C}_d)=2h(d)\log\epsilon_d\gg_\epsilon d^{1/2-\epsilon}.\tag{4-10}$$

To see this, let $d=Db^2$ where D is a fundamental discriminant. Then by Dirichlet class number formula [Davenport 1967, page 50] for binary quadratic forms discriminant d (or by letting $s\to 1$ in (4-7)), we have

$$h(d)\log(\epsilon_d)=d^{1/2}L(1,d)$$

with the same $L(\cdot\,, d)$ given in (4-8), i.e.,

$$L(1, d) = L(1, \chi_D)\Biggl(\sum_{a\,|\,b} \mu(a)\bigl(\tfrac{D}{a}\bigr)a^{-1}\sigma_{-1}\bigl(\tfrac{b}{a}\bigr)\Biggr).$$

Note that

$$\sum_{a\,|\,b} \mu(a)\bigl(\tfrac{D}{a}\bigr)a^{-1}\sigma_{-1}\bigl(\tfrac{b}{a}\bigr) = \sum_{ca\,|\,b} \mu(a)\bigl(\tfrac{D}{a}\bigr)\tfrac{c}{b} = \tfrac{1}{b}\sum_{e\,|\,b} e \prod_{p\,|\,e}\bigl(1-\bigl(\tfrac{D}{p}\bigr)p^{-1}\bigr),$$

where $e = ac$, and that

$$\tfrac{1}{b}\sum_{e\,|\,b} e \prod_{p\,|\,e}\bigl(1-\bigl(\tfrac{D}{p}\bigr)p^{-1}\bigr) \gg b^{-\epsilon}.$$

Now (4-10) follows by using Siegel's lower bound [1935]

$$L(1, \chi_D) \gg_\epsilon D^{-\epsilon},$$

and this completes the proof of Theorem 1.3.

4D. ***Proof of Theorem 1.1.*** We are now ready to prove Theorem 1.1. Assume that $\beta : [0, l(\beta)] \to \mathbb{X}$ is a sufficiently short compact geodesic segment in the region determined by $y < T$ such that $\beta([-l(\beta), 2l(\beta)])$ has no self intersection. (We fix T for simplicity, but it is possible to vary T with d.) For $\delta = d^{-a}$ with $a > 0$ to be chosen later, such that $l(\beta) \gg \delta$, let

$$\beta_1 := \{\beta(t) : t \in [0, l(\beta)-\delta]\} \quad \text{and} \quad \beta_2 := \{\beta(t) : t \in [-\delta, l(\beta)]\}.$$

Then from Lemma 2.7, we have

$$\frac{1}{\delta^2}\int_{\widetilde{\alpha}_2}\int_{\widetilde{\beta}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \le I^{\theta_1,\theta_2}(\beta, \alpha_2) \le \frac{1}{\delta^2}\int_{\widetilde{\alpha}_2}\int_{\widetilde{\beta}_1} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2$$

for any closed geodesic α_2. Now define $f_1, f_2 \in C_0^\infty(S\mathbb{X})$ using Lemma 2.4 by

$$f_1(g) = \frac{1}{\delta^2}\int_{\widetilde{\beta}_1} m_\delta^{\theta_1,\theta_2}(s_1^{-1}g)\, ds_1 \quad \text{and} \quad f_2(g) = \frac{1}{\delta^2}\int_{\widetilde{\beta}_2} M_\delta^{\theta_1,\theta_2}(s_1^{-1}g)\, ds_1,$$

with $\varepsilon = d^{-2a}$, where we assume that $\theta_2 - \theta_1 \gg d^{-a}$. Note that $m(g_1^{-1}g_2)$ and $M(g_1^{-1}g_2)$ are minorant and majorant of $K_\delta^{\theta_1,\theta_2}(g_1, g_2)$ for $g_1 \in \beta_i$, $g_2 \in S\mathbb{X}$ for all sufficiently large d. Hence, for all sufficiently large d (independent of α_2), we have

$$\int_{\widetilde{\alpha}_2} f_1(s)\, ds \le I^{\theta_1,\theta_2}(\beta, \alpha_2) \le \int_{\widetilde{\alpha}_2} f_2(s)\, ds,$$

and so

$$\int_{\mathscr{C}_d} f_1(s)\, ds \le 2I^{\theta_1,\theta_2}(\beta, C_d) \le \int_{\mathscr{C}_d} f_2(s)\, ds, \tag{4-11}$$

where the factor 2 amounts to the fact that $\mathscr{C}_d$ is a double cover of C_d.

We now apply Theorem 1.3 to see that

$$\frac{1}{l(\mathscr{C}_d)}\int_{\mathscr{C}_d} f_i(s)ds=\frac{3}{\pi^2}\int_{S\mathbb{X}} f_i(g)d\mu_g+O_\epsilon(d^{-25/512+\epsilon}\|f_i\|_{W^{6,\infty}}).$$

Because of the choice of f_1 and f_2, we have

$$\|f_i\|_{W^{6,\infty}}\ll \varepsilon^{-6}l(\beta)\ll d^{12a}l(\beta),$$

and

$$\int_{S\mathbb{X}} f_i(g)\,d\mu_g=(\cos\theta_1-\cos\theta_2)(l(\beta)+O(\delta))(1+O(\varepsilon))=(\cos\theta_1-\cos\theta_2)l(\beta)(1+O(d^{-2a}))$$

by Lemma 2.4. Now we complete the proof of Theorem 1.1 for sufficiently short geodesic segments by choosing $a=\frac{25}{7168}$ and applying these estimates to (4-11). This then implies Theorem 1.1 for any geodesic segment of length <1 by dividing the segment into finitely many sufficiently short geodesic segments, and then applying Theorem 1.1 to each of them.

5. Selberg's pretrace Formula for $\mathrm{PSL}_2(\mathbb{R})$

Let $k\in C_0^\infty(\mathrm{PSL}_2(\mathbb{R}))$, and let K be the integral kernel on $S\mathbb{X}$ defined by

$$K(g_1,g_2)=\sum_{\gamma\in\Gamma} k(g_1,\gamma g_2),$$

where $k(g_1,g_2)=k(g_1^{-1}g_2)$. The corresponding integral operator T_K acts on $f\in L^2(S\mathbb{X})$ by

$$T_K(f):=\int_{S\mathbb{X}} K(g_1,g_2)f(g_2)\,dg_2=\int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1^{-1}g_2)f(g_2)\,dg_2.$$

It follows that $T_K(f)\in L^2(S\mathbb{X})$. In this section, we study the spectral expansion of K in terms of the equivariant eigenfunctions of the Casimir operator, which are explicitly described in Section 3A. In other words, we derive Selberg's pretrace formula for $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$.

5A. ***Cuspidal spectrum.*** In this section, we describe explicitly the spectrum of T_K acting on the cuspidal subspace $L^2_{\mathrm{cusp}}(S\mathbb{X})$. Let $R_g(f)(x)=f(xg)$ be the right regular action of $\mathrm{PSL}_2(\mathbb{R})$ on

$$L^2_{\mathrm{cusp}}(\Gamma\backslash\mathrm{PSL}_2(\mathbb{R}))=L^2_{\mathrm{cusp}}:S\mathbb{X}).$$

Lemma 5.1. *Let π be an irreducible unitary representation of* $\mathrm{PSL}_2(\mathbb{R})$. *Then for any $f\in W_\pi\subset L^2_{\mathrm{cusp}}(S\mathbb{X})$, we have*

$$T_K(f)\in W_\pi.$$

Proof. Observe that

$$T_K(f)(g_1)=\int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1,g_2)f(g_2)\,dg_2=\int_{\mathrm{PSL}_2(\mathbb{R})} k(g_1^{-1}g_2)f(g_2)\,dg_2=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)f(g_1u)\,du,$$

where $u = g_1^{-1} g_2$. Hence, we have

$$T_K(f) = \int_{\mathrm{PSL}_2(\mathbb{R})} k(u) R_u(f)\, du,$$

and because $R_u(f) \in W_\pi$ for every u, we conclude that $T_K(f) \in W_\pi$. □

From 5A, for an abstract irreducible unitary representation π of $\mathrm{PSL}_2(\mathbb{R})$ and $f \in W_\pi$, we define the action of k on f by

$$k * f = \int_{\mathrm{PSL}_2(\mathbb{R})} k(u)\pi(u)(f)\, du,$$

which agrees with $T_K(f)$ when W_π is a subspace of $L^2_{\mathrm{cusp}}(S\mathbb{X})$.

Let $\psi : W_\pi \to W_{\pi'}$ be an isomorphism of representations π and π'. Note that for $f \in W_\pi$ and $f' \in W_{\pi'}$ with $\psi(f) = f'$, we have $\psi(k * f) = k * f'$. We denote by $\phi_m \in W_\pi$ the unique (up to a unit scalar) vector of norm 1 and weight m. We fix the unit scalar except for the spherical or the lowest weight vector, by using the normalized lowering and raising operator that we introduced in (3-4) and (3-5).

Now let

$$h(k, m, n, \pi) := \langle k * \phi_m, \phi_n \rangle, \tag{5-1}$$

and let $M_\pi(m, n)(g) = \langle \pi(g)\phi_m, \phi_n \rangle$ be the matrix coefficient of π. We note that $h(k, m, n, \pi)$ and $M_\pi(m, n)(g)$ do not depend on the choice of the unit scalar of the spherical or the lowest weight vector.

We recall some properties of $M_\pi(m, n)(g)$ in the following lemma.

Lemma 5.2. *We have for every $g \in \mathrm{PSL}_2(\mathbb{R})$,*

$$|M_\pi(m, n)(g)| \le 1,$$

and

$$M_\pi(m, n)(R_{\theta'} g R_\theta) = e^{-im\theta} e^{-in\theta'} M_\pi(m, n)(g).$$

Proof. We have

$$1 = |\pi(g)\phi_m|^2 = \sum_n \langle \pi(g)\phi_m, \phi_n \rangle^2,$$

from which it is immediate that $|M_\pi(m, n)(g)| \le 1$. For the second identity, we have

$$M_\pi(m, n)(R_{\theta'} g R_\theta) = \langle \pi(g)\pi(R_\theta)\phi_m, \pi(R_{-\theta'})\phi_n \rangle = e^{-im\theta} e^{-in\theta'} M_\pi(m, n)(g).$$ □

Define $k_{m,n} \in C_0^\infty(\mathrm{PSL}_2(\mathbb{R}))$ by

$$k_{m,n}(g) := \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} k(R_{\theta'} g R_\theta) e^{-in\theta' - im\theta}\, d\theta'\, d\theta. \tag{5-2}$$

Note that

$$k_{m,n}(R_{\theta_1} g R_{\theta_2}) = e^{in\theta_1} k_{m,n}(g) e^{im\theta_2}. \tag{5-3}$$

The following lemma holds for every unitary irreducible representation of $\mathrm{PSL}_2(\mathbb{R})$.

Lemma 5.3. *We have*

$$h(k,m,n,\pi)=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(u)M_\pi(m,n)(u)\,du,$$

and for all nonnegative integers N_1, N_2, we have the following estimate

$$h(k,m,n,\pi)\ll_{N=N_1+N_2}(1+|m|)^{-N_1}(1+|n|)^{-N_2}\|k\|_{W^{N,1}}.$$

Proof. Recall from the definition that

$$h(k,m,n,\pi)=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)\langle\pi(u)\phi_m,\phi_n\rangle\,du=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)M_\pi(m,n)(u)\,du,$$

and so

$$\begin{aligned}
h(k,m,n,\pi)&=\int_{\mathrm{PSL}_2(\mathbb{R})} k(u)M_\pi(m,n)(u)\,du.\\
&=\frac{1}{4\pi^2}\int_{\mathrm{PSL}_2(\mathbb{R})}\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)M_\pi(m,n)(R_{\theta'}uR_\theta)\,d\theta\,d\theta'\,du\\
&=\frac{1}{4\pi^2}\int_{\mathrm{PSL}_2(\mathbb{R})} M_\pi(m,n)(u)\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)e^{-im\theta}e^{-in\theta'}\,d\theta\,d\theta'\,du\\
&=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(u)M_\pi(m,n)(u)\,du.
\end{aligned}$$

Therefore, by integration by parts, we have

$$\begin{aligned}
h(k,m,n,\pi)&\le\int_{\mathrm{PSL}_2(\mathbb{R})}|k_{m,n}(u)|\,du\\
&=\int_{\mathrm{PSL}_2(\mathbb{R})}\left|\frac{1}{4\pi^2}\int_\theta\int_{\theta'} k(R_{\theta'}uR_\theta)e^{-im\theta}e^{-in\theta'}\,d\theta\,d\theta'\right|du\\
&\ll_N (1+|m|)^{-N_1}(1+|n|)^{-N_2}\|k\|_{W^{N,1}},
\end{aligned}$$

where we used $|M_\pi(m,n)(u)|\le 1$ from Lemma 5.2. This completes the proof of our lemma. $\square$

5A1. *Principal series representation of* $\mathrm{SL}_2(\mathbb{R})$. For our application in the subsequent chapters, we need a refined estimate for $h(k,m,n,\pi)$ when π is a unitary principal series representation. We first give an explicit representation of $h(k,m,n,\pi)$.

Lemma 5.4. *Let W_π be a unitary principal series representation of* $\mathrm{SL}_2(\mathbb{R})$ *with the parameter* $\frac{1}{2}+it$ [Knapp 2001, Chapter VII]. *Let*

$$h(k,m,n,t):=\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g)y^{1/2+it}e^{-im\theta}\,dg,\tag{5-4}$$

where $g=na(y)R_\theta$. Then we have

$$h(k,m,n,\pi)=h(k,m,n,t).$$

Proof. We note that principal series representations are induced from the unitary characters of the upper triangular matrices to $\mathrm{PSL}_2(\mathbb{R})$ [Knapp 2001, Chapter VII]. In this model, a dense subspace of a representation is given by

$$\{f : \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{C} \text{ continuous} : f(xan) = e^{(it+1/2)\log(a)} f(x)\}$$

with the norm

$$|f|^2 = \frac{1}{2\pi}\int_\theta |f(R_\theta)|^2\, d\theta,$$

and the $\mathrm{PSL}_2(\mathbb{R})$ action is given by

$$\pi(g)f(x) = f(g^{-1}x).$$

The weight m unit vectors are explicitly given by

$$\phi_m(R_\theta a(y)n) = e^{im\theta} y^{-(1/2+it)}.$$

Note that the orthonormal basis $\{\phi_m\}$ is normalized as our convention in (3-4), i.e.,

$$\boldsymbol{E}^-\phi_m = (m+1-2it)\phi_{m-2} \quad \text{and} \quad \boldsymbol{E}^+\phi_m = (m+1+2it)\phi_{m+2}.$$

With these, we first see that

$$\begin{aligned} k*\phi_m(R_{\theta'}) &= \int_{\mathrm{PSL}_2(\mathbb{R})} k(u)y(u^{-1}R_{\theta'})^{-(1/2+it)} e^{im\theta(u^{-1}R_{\theta'})}\, du \\ &= \int_{\mathrm{PSL}_2(\mathbb{R})} k(R_{\theta'}v^{-1})y(v)^{-(1/2+it)}e^{im\theta(v)}\, dv, \end{aligned}$$

where $v = u^{-1}R_{\theta'}$ and $v = R_{\theta(v)}a(y(v))n(v)$. We therefore have

$$\begin{aligned} h(k,m,n,\pi) &= \langle k*f_m, f_n\rangle \\ &= \frac{1}{2\pi}\int_{\theta'} k*f_m(R_{\theta'})\bar{f}_n(R_{\theta'})\, d\theta' \\ &= \frac{1}{2\pi}\int_{\theta'} e^{-in\theta'}\int_{\mathrm{PSL}_2(\mathbb{R})} k(R_{\theta'}v^{-1})y(v)^{-(1/2+it)}e^{im\theta(v)}\, dv\, d\theta' \\ &= \frac{1}{2\pi}\int_{\mathrm{PSL}_2(\mathbb{R})} y^{1/2+it}\int_{\theta'} e^{-in\theta'}e^{-im\theta}k(R_{\theta'}w)\, d\theta'\, dw \\ &= \int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(w)y^{1/2+it}e^{-im\theta}\, dw, \end{aligned}$$

where $w = v^{-1}$ and $w = na(y)R_\theta$. Note that $y = y(v)^{-1}$ and $\theta = -\theta(v)$. □

We now prove that $h(k,m,n,t)$ decays fast in all parameters uniformly.

Lemma 5.5. *Suppose that k is supported inside the compact subset $C \subset \mathrm{SL}_2(\mathbb{R})$. Then we have*

$$\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g)y^{1/2+it}e^{-im\theta}\, dg \ll_{N,C} (1+|m|)^{-N_1}(1+|n|)^{-N_2}(1+|t|)^{-N_3}\|k\|_{W^{N,\infty}}$$

for any $N_1, N_2, N_3 \geq 0$, where $N = N_1+N_2+N_3$.

Proof. From the definition, we have

$$\int_{\mathrm{PSL}_2(\mathbb{R})} k_{m,n}(g) y^{1/2+it} e^{-im\theta}\, dg = \frac{1}{4\pi}\int_{\mathbb{H}}\int_0^{2\pi}\int_0^{2\pi} k(R_{\theta_1'} n(x) a(y) R_{\theta_2'}) y^{1/2+it} e^{-in\theta_1' - im\theta_2'}\, d\theta_1'\, d\theta_2' \,\frac{dx\,dy}{y^2},$$

and so the statement follows from integration by parts. □

5B. *Continuous spectrum.* For $k_{m,n}$ given by (5-2), let

$$K_{m,n}(g_1, g_2) := \sum_{\gamma\in\Gamma} k_{m,n}(g_1^{-1}\gamma g_2). \tag{5-5}$$

Then we infer from (5-3) that

$$K_{m,n}(g_1 R_{\theta_1}, g_2 R_{\theta_2}) = e^{-in\theta_1} K_{m,n}(g_1, g_2) e^{im\theta_2},$$

and so it defines an integral operator that maps weight m forms to weight n forms. Denote by $S^m \subset L^2(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))$ the space of weight m forms and by S^m_{cusp} the space of weight m forms in $L^2_{\mathrm{cusp}}(S\mathbb{X})$. We first recall the following result regarding the decomposition of $K_{m,m}$.

Theorem 5.6 [Hejhal 1976]. *The integral kernel*

$$K_{m,m}(g_1, g_2) - \frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,m,t) E_m\big(g_1, \tfrac12 + it\big)\overline{E_m\big(g_2, \tfrac12 + it\big)}\, dt$$

defines a compact operator $S^m_{\mathrm{cusp}} \to S^m_{\mathrm{cusp}}$ that acts trivially on Θ. (Here $h(k,m,m,t)$ is given by (5-4).)

We define $\boldsymbol{E}^a$ to be $(\boldsymbol{E}^+)^a$ if $a > 0$, and $(\boldsymbol{E}^-)^{|a|}$ if $a < 0$. We have

$$\overline{\boldsymbol{E}^a} = (-\boldsymbol{E})^{-a},$$

which follows directly from (3-2). Let $c_{m,n}$ be given by

$$\boldsymbol{E}^{n-m} E_m(g,s) = c_{m,n}(s) E_n(g,s).$$

Observe that

$$\boldsymbol{E}^{n-m} y^s e^{-im\theta} = c_{m,n}(s) y^s e^{-in\theta},$$

and that

$$\overline{c_{m,n}\big(\tfrac12 + it\big)} = c_{n,m}\big(\tfrac12 + it\big) \tag{5-6}$$

for $t\in\mathbb{R}$.

Theorem 5.7. *For $m, n \in 2\mathbb{Z}$,*

$$K_{m,n}(g_1, g_2) - \frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,n,t) E_n\big(g_1, \tfrac12 + it\big)\overline{E_m\big(g_2, \tfrac12 + it\big)}\, dt$$

defines a compact operator $S^m_{\mathrm{cusp}} \to S^n_{\mathrm{cusp}}$ that acts trivially on Θ.

Proof. Note that

$$\int \boldsymbol{E}_{g_2}^{m-n}(K(g_1,g_2)f(g_2))\,dg_2 = 0$$

for every g_1, $m \neq n$, and $f \in C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))$. Hence

$$T_K \boldsymbol{E}^{m-n} : C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R})) \to C_0^\infty(\Gamma\backslash \mathrm{PSL}_2(\mathbb{R}))$$

is an integral operator with the integral kernel

$$K'(g_1,g_2) = \sum_{\gamma\in\Gamma} k'(g_1^{-1}\gamma g_2),$$

where

$$k'(g) = (-\boldsymbol{E})^{m-n}k(g) = \overline{\boldsymbol{E}^{n-m}}k(g).$$

Then by Theorem 5.6, we see that

$$K''(g_1,g_2) = K'_{n,n}(g_1,g_2) - \frac{1}{4\pi}\int_{-\infty}^{\infty} h(k',n,n,t)E_n\big(g_1,\tfrac{1}{2}+it\big)\overline{E_n\big(g_2,\tfrac{1}{2}+it\big)}\,dt$$

defines a compact operator $T_{K''} : S^n_{\mathrm{cusp}} \to S^n_{\mathrm{cusp}}$ that acts trivially on Θ. Note that

$$\int_{-\infty}^{\infty} h(k',n,n,t)E_n\big(g_1,\tfrac{1}{2}+it\big)\overline{E_n\big(g_2,\tfrac{1}{2}+it\big)}\,dt \\ = \int_{-\infty}^{\infty} \frac{h(k',n,n,t)}{\overline{c_{m,n}(1/2+it)}}E_n\big(g_1,\tfrac{1}{2}+it\big)\overline{\boldsymbol{E}^{n-m}E_m\big(g_2,\tfrac{1}{2}+it\big)}\,dt.$$

Let

$$K'''(g_1,g_2) := K_{m,n}(g_1,g_2) - \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{h(k',n,n,t)}{\overline{c_{m,n}(1/2+it)}}E_n\big(g_1,\tfrac{1}{2}+it\big)\overline{E_m\big(g_2,\tfrac{1}{2}+it\big)}\,dt.$$

Note that

$$T_{K''} = T_{K'''} \circ \boldsymbol{E}^{m-n}.$$

Firstly, since $\boldsymbol{E}^{m-n}$ does not annihilate the Eisenstein series, $T_{K'''}$ acts trivially on Θ.

If $m > n \geq 0$ or $m < n \leq 0$, then as a map $S^n_{\mathrm{cusp}} \to S^m_{\mathrm{cusp}}$, $ker(\boldsymbol{E}^{m-n})$ is empty, and we may decompose S^m_{cusp} as

$$S^m_{\mathrm{cusp}} = \Im(\boldsymbol{E}^{m-n}) \oplus R,$$

where R is a finite dimensional subspace of S^m_{cusp} spanned by modular forms of weight $> n$ and their images under raising operators in S^m_{cusp}. Note that

$$(\boldsymbol{E}^{m-n})^{-1} : \Im(\boldsymbol{E}^{m-n}) \to S^n_{\mathrm{cusp}}$$

is a bounded operator, hence

$$T_{K'''}|_{Im(\boldsymbol{E}^{m-n})} = T_{K''} \circ (\boldsymbol{E}^{m-n})^{-1}$$

is a compact operator. This implies that $T_{K'''}$ is a direct sum of a compact operator and finite dimensional linear operator, which is a compact operator.

If $n > m \geq 0$ or $n < m \leq 0$, then $\boldsymbol{E}^{m-n} : S^n_{\text{cusp}} \to S^m_{\text{cusp}}$ is surjective, and so we may define a bounded operator

$$(\boldsymbol{E}^{m-n})^{-1} : S_m \to (\ker(\boldsymbol{E}^{m-n}))^{\perp}$$

from which it follows that

$$T_{K'''} = T_{K''} \circ (\boldsymbol{E}^{m-n})^{-1}$$

is a compact operator.

If $n > 0 > m$ or $m > 0 > n$, then we further decompose $T_{K''}$ to

$$S^n_{\text{cusp}} \xrightarrow{\boldsymbol{E}^{-n}} S^0_{\text{cusp}} \xrightarrow{\boldsymbol{E}^{m}} S^m_{\text{cusp}} \xrightarrow{T_{K'''}} S^n_{\text{cusp}},$$

and then combine the above arguments to see that $T_{K''}$ is a compact operator.

Finally, observe that

$$h(k', n, n, t) = \int_{\mathrm{PSL}_2(\mathbb{R})} (\overline{\boldsymbol{E}^{n-m}} k(g)) y^{\frac{1}{2}+it} e^{in\theta}\, dg = c_{n,m}\left(\tfrac{1}{2}+it\right) \int_{\mathrm{PSL}_2(\mathbb{R})} k(g) y^{\frac{1}{2}+it} e^{im\theta}\, dg,$$

and we complete the proof using (5-6). □

5C. *General case.* We are now ready to describe Selberg's pretrace formula for $\mathrm{PSL}_2(\mathbb{R})$.

Theorem 5.8. *For $k \in C_0^\infty(\mathrm{PSL}_2(\mathbb{R}))$, let K be the integral kernel on $S\mathbb{X}$ defined by*

$$K(g_1, g_2) = \sum_{\gamma \in \Gamma} k(g_1, \gamma g_2).$$

Then we have

$$K(g_1, g_2) = \frac{9}{\pi^4} \iint K(g_1, g_2)\, dg_1\, dg_2 + \sum_{\substack{e \geq 0 \\ 2 \mid e}} \sum_{j=1}^{d_e} \sum_{\substack{m,n \in 2\mathbb{Z} \\ |m|,|n| \geq e}} h(k, m, n, \pi^e_j) \phi^e_{j,n}(g_1) \overline{\phi^e_{j,m}(g_2)}$$
$$+ \frac{1}{4\pi} \sum_{m,n \in 2\mathbb{Z}} \int_{-\infty}^{\infty} h(k, m, n, t) E_n\left(g_1, \tfrac{1}{2}+it\right) \overline{E_m\left(g_2, \tfrac{1}{2}+it\right)}\, dt,$$

where π^e_j is the irreducible unitary representation of $\mathrm{PSL}_2(\mathbb{R})$ associated to ϕ^e_j.

Proof. We first note from (5-2) and (5-5) that

$$\begin{aligned}
K_{m,n}(g_1,g_2) &= \sum_{\gamma\in\Gamma}\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} k(R_{\theta_1'}g_1^{-1}\gamma g_2 R_{\theta_2'})e^{-in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\
&= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\sum_{\gamma\in\Gamma} k(R_{-\theta_1'}g_1^{-1}\gamma g_2 R_{\theta_2'})e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\
&= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} K(g_1R_{\theta_1'}, g_2R_{\theta_2'})e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' \\
&= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} K((x_1,y_1,\theta_1'),(x_2,y_2,\theta_2'))e^{in\theta_1'-im\theta_2'}\,d\theta_1'\,d\theta_2' e^{-in\theta_1+im\theta_2}.
\end{aligned}$$

Therefore, we have the Fourier expansion of K,

$$K(g_1,g_2)=\sum_{n,m\in 2\mathbb{Z}} K_{m,n}(g_1,g_2),$$

where the summation is uniform for g_1 and g_2 in compacta.

We infer from Theorem 5.7 that

$$K_{m,n}(g_1,g_2)-\frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,n,t)E_n\bigl(g_1,\tfrac12+it\bigr)\overline{E_m\bigl(g_2,\tfrac12+it\bigr)}\,dt$$

defines a compact operator acting on L_{cusp} that acts trivially on Θ. Because it only acts nontrivially on weight m forms, we see that

$$\begin{aligned}
K_{m,n}(g_1,g_2)&-\frac{1}{4\pi}\int_{-\infty}^{\infty} h(k,m,n,t)E_n\bigl(g_1,\tfrac12+it\bigr)\overline{E_m\bigl(g_2,\tfrac12+it\bigr)}\,dt \\
&= \frac{9}{\pi^4}\iint K_{m,n}(g_1,g_2)\,dg_1\,dg_2+\sum_{\substack{e\ge 0\\ 2\mid e}}^{\min\{|m|,|n|\}}\sum_{j=1}^{d_e} h(k,m,n,\pi_j^e)\phi_{j,n}^e(g_1)\overline{\phi_{j,m}^e(g_2)},
\end{aligned}$$

where we used (5-1), and the fact that

$$\int_{-\infty}^{\infty} h(k,m,n,t)E_n\bigl(g_1,\tfrac12+it\bigr)\overline{E_m\bigl(g_2,\tfrac12+it\bigr)}\,dt$$

acts trivially on L^2_{cusp}. Note that the integral on the right-hand side of the equation vanishes unless $m=n=0$, in which case it is identical to

$$\frac{9}{\pi^4}\iint K(g_1,g_2)\,dg_1\,dg_2. \qquad \square$$

5D. ***Proof of Theorem 1.4.*** We now present a proof of Theorem 1.4. By Theorem 5.8, we have

$$\frac{1}{l(\mathscr{C}_{d_1})l(\mathscr{C}_{d_2})}\int_{\mathscr{C}_{d_2}}\int_{\mathscr{C}_{d_1}} K(s_1,s_2)\,ds_1\,ds_2 = M+D+\frac{1}{4\pi}E,$$

where

$$M = \frac{9}{\pi^4} \iint K(g_1, g_2)\, dg_1\, dg_2,$$

$$D = \sum_{\substack{e\geq 0\\ 2\,|\,e}} \sum_{j=1}^{d_e} \sum_{\substack{m,n\in 2\mathbb{Z}\\ |m|,|n|\geq e}} h(k,m,n,\pi_j^e) \frac{\mu_{d_1}(\phi_{j,n}^e)}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(\phi_{j,m}^e)}}{l(\mathscr{C}_{d_2})}$$

$$= \sum_{\substack{e\geq 0\\ 4\,|\,e}} \sum_{j=1}^{d_e} \frac{\mu_{d_1}(\phi_j^e)}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(\phi_j^e)}}{l(\mathscr{C}_{d_2})} \sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\geq e}} h(k,m,n,\pi_j^e)\eta_j^e(\phi_{j,n}^e)\overline{\eta_j^e(\phi_{j,m}^e)},$$

and

$$E = \sum_{m,n\in 2\mathbb{Z}} \int_{-\infty}^{\infty} h(k,m,n,t) \frac{\mu_{d_1}(E_n(\,\cdot\,,1/2+it))}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(E_m(\,\cdot\,,1/2+it))}}{l(\mathscr{C}_{d_2})}\, dt$$

$$= \sum_{m,n\in 4\mathbb{Z}} \int_{-\infty}^{\infty} h(k,m,n,t) \frac{\mu_{d_1}(E_0(\,\cdot\,,1/2+it))}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(E_0(\,\cdot\,,1/2+it))}}{l(\mathscr{C}_{d_2})} \eta\big(n,\tfrac{1}{2}+it\big)\overline{\eta\big(m,\tfrac{1}{2}+it\big)}dt.$$

For D with $e>0$, we use (4-2), (4-5), Lemma 5.3 with $N_1 = N_2 = 5$, and (4-10) to see that

$$\sum_{\substack{e>0\\ 4\,|\,e}} \sum_{j=1}^{d_e} \frac{\mu_{d_1}(\phi_j^e)}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(\phi_j^e)}}{l(\mathscr{C}_{d_2})} \sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\geq e}} h(k,m,n,\pi_j^e)\eta_j^e(\phi_{j,n}^e)\overline{\eta_j^e(\phi_{j,m}^e)}$$

$$\ll_\epsilon \sum_{\substack{e>0\\ 4\,|\,e}} e^{6.8}(d_1d_2)^{-25/512+\epsilon} \sum_{\substack{m,n\in 4\mathbb{Z}\\ |m|,|n|\geq e}} |m|^{-5}|n|^{-5}\|k\|_{W^{10,\infty}}$$

$$\ll (d_1d_2)^{-25/512+\epsilon}\|k\|_{W^{10,\infty}}.$$

For D with $e=0$, we use (4-1), (4-6), Lemma 5.5 with $N_1 = N_2 = 2$ and $N_3 = 4$, and (4-10) to see that

$$\sum_{j=1}^{\infty} \frac{\mu_{d_1}(\phi_j^0)}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(\phi_j^0)}}{l(\mathscr{C}_{d_2})} \sum_{m,n\in 4\mathbb{Z}} h(k,m,n,\pi_j^0)\eta_j^0(\phi_{j,n}^0)\overline{\eta_j^0(\phi_{j,m}^0)}$$

$$\ll_\epsilon \sum_{j=1}^{\infty} (d_1d_2)^{-25/512+\epsilon}|t_j|^{3/2} \sum_{m,n\in 4\mathbb{Z}} (1+|m|)^{-2}(1+|n|)^{-2}(1+|t_j|)^{-4}\|k\|_{W^{8,\infty}}$$

$$\ll (d_1d_2)^{-25/512+\epsilon}\|k\|_{W^{8,\infty}}.$$

For E, we use (4-3), (4-9), Lemma 5.5 with $N_1 = N_2 = 2$ and $N_3 = 3$, and (4-10) to see that

$$\sum_{m,n\in 4\mathbb{Z}} \int_{-\infty}^{\infty} h(k,m,n,t) \frac{\mu_{d_1}(E_0(\,\cdot\,,1/2+it))}{l(\mathscr{C}_{d_1})} \frac{\overline{\mu_{d_2}(E_0(\,\cdot\,,1/2+it))}}{l(\mathscr{C}_{d_2})} \eta\big(n,\tfrac{1}{2}+it\big)\overline{\eta\big(m,\tfrac{1}{2}+it\big)}\, dt$$

$$\ll_\epsilon \sum_{m,n\in 4\mathbb{Z}} \int_{-\infty}^{\infty} (d_1d_2)^{-1/16+\epsilon}(1+|m|)^{-2}(1+|n|)^{-2}(|t|+1)^{-2}\|k\|_{W^{7,\infty}}\, dt$$

$$\ll (d_1d_2)^{-1/16+\epsilon}\|k\|_{W^{7,\infty}}.$$

Now observe that

$$\iint K(g_1, g_2)\, dg_1\, dg_2 = \int_{S\mathbb{X}} \int_{S\mathbb{H}} k(g_1^{-1} g_2)\, dg_2\, dg_1 = \frac{\pi^2}{3} \int_{S\mathbb{H}} k(g)\, dg,$$

and so

$$M = \frac{3}{\pi^2} \int_{S\mathbb{H}} k(g)\, dg.$$

So far, we proved the following:

Theorem 5.9. *For any $k \in C_0^\infty(S\mathbb{H})$, we have*

$$\frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K(s_1, s_2)\, ds_1\, ds_2 = \frac{3}{\pi^2} \int_{S\mathbb{H}} k(g)\, dg + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \|k\|_{W^{10,\infty}}).$$

Remark 5.1. Note that this is *not* the same as equidistribution of $\mathscr{C}_{d_1} \times \mathscr{C}_{d_2}$ in $S\mathbb{X} \times S\mathbb{X}$. For instance, if we replace K with any compactly supported smooth function in $S\mathbb{X} \times S\mathbb{X}$, then the equality may not hold when d_1 is fixed and d_2 tends to ∞.

In order to prove Theorem 1.4, we make specific choices of k in Theorem 5.9. We let K_1 and K_2 to be the kernel corresponding to $k = m_\delta^{\theta_1,\theta_2}$ and $k = M_\delta^{\theta_1,\theta_2}$ defined in Lemma 2.4, respectively. Then by Lemma 2.4, we have

$$\begin{aligned}\frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_1(s_1, s_2)\, ds_1\, ds_2 &\le \frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 \\ &\le \frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_2(s_1, s_2)\, ds_1\, ds_2,\end{aligned}$$

while we know from Lemma 2.6 that

$$\int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_\delta^{\theta_1,\theta_2}(s_1, s_2)\, ds_1\, ds_2 = 4\delta^2 I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2}).$$

We now apply Theorem 5.9 and Lemma 2.4 to see that

$$\frac{1}{l(\mathscr{C}_{d_1}) l(\mathscr{C}_{d_2})} \int_{\mathscr{C}_{d_2}} \int_{\mathscr{C}_{d_1}} K_i(s_1, s_2)\, ds_1\, ds_2 = \frac{3}{\pi^2}(\cos\theta_1 - \cos\theta_2)\delta^2(1 + O(\varepsilon)) + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \varepsilon^{-10}).$$

Therefore, we have

$$\frac{I_{\theta_1,\theta_2}(C_{d_1}, C_{d_2})}{l(C_{d_1}) l(C_{d_2})} = \frac{3}{\pi^2}(\cos\theta_1 - \cos\theta_2)(1 + O(\delta^2))(1 + O(\varepsilon)) + O_\epsilon((d_1 d_2)^{-25/512+\epsilon} \varepsilon^{-10} \delta^{-2}),$$

and by choosing $\delta^2 = \varepsilon = (d_1 d_2)^{-25/6144}$, we complete the proof of Theorem 1.4.

Appendix: Jacobian computation

Recall that $\Psi : AKA \to \mathrm{SL}_2(\mathbb{R})$ is given by

$$(t_1, \varphi, t_2) \mapsto \begin{pmatrix} e^{t_1/2} & 0 \\ 0 & e^{-t_1/2} \end{pmatrix} R_{\frac{\varphi}{2}} \begin{pmatrix} e^{-t_2/2} & 0 \\ 0 & e^{t_2/2} \end{pmatrix} = \begin{pmatrix} e^{(t_1-t_2)/2} \cos\frac{\varphi}{2} & -e^{(t_1+t_2)/2} \sin\frac{\varphi}{2} \\ e^{(-t_1-t_2)/2} \sin\frac{\varphi}{2} & e^{(t_2-t_1)/2} \cos\frac{\varphi}{2} \end{pmatrix}.$$

In this section, we compute the pullback of $dV = dx\,dy\,d\theta/y^2$ under Ψ. We start with the identity

$$\begin{pmatrix} e^{(t_1-t_2)/2}\cos\frac{\varphi}{2} & -e^{(t_1+t_2)/2}\sin\frac{\varphi}{2} \\ e^{(-t_1-t_2)/2}\sin\frac{\varphi}{2} & e^{(t_2-t_1)/2}\cos\frac{\varphi}{2} \end{pmatrix} = n(x)a(y)R_\theta = \begin{pmatrix} * & * \\ \frac{\sin\theta}{\sqrt{y}} & \frac{\cos\theta}{\sqrt{y}} \end{pmatrix}.$$

By comparing the image of $i \in \mathbb{H}$, we have

$$x + iy = \frac{e^{(t_1-t_2)/2}\cos\frac{\varphi}{2} i - e^{(t_1+t_2)/2}\sin\frac{\varphi}{2}}{e^{(-t_1-t_2)/2}\sin\frac{\varphi}{2} i + e^{(t_2-t_1)/2}\cos\frac{\varphi}{2}},$$

and for simplicity, we write this as $\frac{A}{B}$. By comparing the second row of each matrix, we have

$$\frac{e^{i\theta}}{\sqrt{y}} = B.$$

From a quick computation, we see that

$$A_{t_1} = \frac{A}{2},\quad B_{t_1} = -\frac{B}{2},\quad A_{t_2} = \frac{\bar{A}}{2},\quad B_{t_2} = \frac{\bar{B}}{2},\quad A_\varphi = -\frac{e^{t_1}}{2}B,\quad B_\varphi = \frac{e^{-t_1}}{2}A,\quad \Im A\bar{B} = 1,\quad y = \frac{1}{|B|^2}.$$

We use these to express the Jacobian matrix in terms of A and B as follows:

$$\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)} = \begin{pmatrix} \operatorname{Re}\frac{A}{B} & \Im\frac{1}{B^2} & \operatorname{Re}\left(-\frac{e^{t_1}}{2} - \frac{e^{-t_1}}{2}\frac{A^2}{B^2}\right) \\ \Im\frac{A}{B} & -\operatorname{Re}\frac{1}{B^2} & \Im\left(-\frac{e^{t_1}}{2} - \frac{e^{-t_1}}{2}\frac{A^2}{B^2}\right) \\ 0 & \frac{1}{2}\Im\frac{\bar{B}}{B} & \frac{e^{-t_1}}{2|B|^2} \end{pmatrix}.$$

From this, we have

$$\begin{aligned} \frac{1}{y^2}\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| &= |B|^4\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| \\ &= \left|-\tfrac{1}{2}e^{-t_1}\operatorname{Re}\left(\frac{\bar{A}}{B}\right) + \tfrac{1}{4}\Im(\bar{B}^2)\Im\left(\bar{A}B\left(e^{t_1} + e^{-t_1}\frac{A^2}{B^2}\right)\right)\right| \\ &= \left|\frac{e^{t_1}}{2}\Im(B^2) + \frac{e^{-t_1}}{4|B|^2}(-2\operatorname{Re}(AB) - |A|^2\Im(B^2))\right|. \end{aligned}$$

Now we use the definition of A and B to compute each term explicitly as follows

$$\begin{aligned} 2\operatorname{Re}(AB) &= -(e^{t_2} + e^{-t_1})\sin\varphi \\ e^{t_1}\Im(B^2) &= \sin\varphi \\ e^{-t_1}|A|^2 &= e^{t_2}\sin^2\frac{\varphi}{2} + e^{-t_2}\cos^2\frac{\varphi}{2} \\ e^{t_1}|B|^2 &= e^{-t_2}\sin^2\frac{\varphi}{2} + e^{t_2}\cos^2\frac{\varphi}{2}, \end{aligned}$$

and so

$$\frac{1}{y^2}\left|\frac{\partial(x,y,\theta)}{\partial(t_1,t_2,\varphi)}\right| = \tfrac{1}{2}|\sin\varphi|.$$

Therefore, we conclude that

$$dV = \tfrac{1}{2}|\sin\varphi|\,dt_1\,dt_2\,d\varphi. \tag{A-1}$$

References

[Artin 1924] E. Artin, "Ein mechanisches System mit quasiergodischen Bahnen", *Abh. Math. Sem. Univ. Hamburg* **3**:1 (1924), 170–175. MR Zbl

[Baruch and Mao 2010] E. M. Baruch and Z. Mao, "A generalized Kohnen–Zagier formula for Maass forms", *J. Lond. Math. Soc.* (2) **82**:1 (2010), 1–16. MR Zbl

[Blomer et al. 2007] V. Blomer, G. Harcos, and P. Michel, "A Burgess-like subconvex bound for twisted L-functions", *Forum Math.* **19**:1 (2007), 61–105. MR Zbl

[Bonahon 1986] F. Bonahon, "Bouts des variétés hyperboliques de dimension 3", *Ann. of Math.* (2) **124**:1 (1986), 71–158. MR Zbl

[Darmon and Vonk 2022] H. Darmon and J. Vonk, "Arithmetic intersections of modular geodesics", *J. Number Theory* **230** (2022), 89–111. MR Zbl

[Davenport 1967] H. Davenport, *Multiplicative number theory*, Lectures in Advanced Mathematics **1**, Markham Publishing Co., Chicago, 1967. MR Zbl

[Duke 1988] W. Duke, "Hyperbolic distribution problems and half-integral weight Maass forms", *Invent. Math.* **92**:1 (1988), 73–90. MR Zbl

[Heath-Brown 1980] D. R. Heath-Brown, "Hybrid bounds for Dirichlet L-functions, II", *Quart. J. Math. Oxford Ser.* (2) **31**:122 (1980), 157–167. MR Zbl

[Hejhal 1976] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2, R)$, *I*, Lecture Notes in Mathematics **548**, Springer, 1976. MR Zbl

[Herrera Jaramillo 2015] Y. A. Herrera Jaramillo, "Intersection numbers of geodesic arcs", *Rev. Colombiana Mat.* **49**:2 (2015), 307–319. MR

[Jakobson 1994] D. Jakobson, "Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$", *Ann. Inst. Fourier (Grenoble)* **44**:5 (1994), 1477–1504. MR Zbl

[Jung and Reid 2021] J. Jung and A. W. Reid, "Embedding closed totally geodesic surfaces in Bianchi orbifolds", preprint, 2021. To appear in *Math. Res. Lett.* arXiv 2003.05427

[Katok 1985] S. Katok, "Closed geodesics, periods and arithmetic of modular forms", *Invent. Math.* **80**:3 (1985), 469–480. MR Zbl

[Katok and Sarnak 1993] S. Katok and P. Sarnak, "Heegner points, cycles and Maass forms", *Israel J. Math.* **84**:1-2 (1993), 193–227. MR Zbl

[Knapp 2001] A. W. Knapp, *Representation theory of semisimple groups*, Princeton University Press, 2001. MR Zbl

[Kohnen and Zagier 1981] W. Kohnen and D. Zagier, "Values of L-series of modular forms at the center of the critical strip", *Invent. Math.* **64**:2 (1981), 175–198. MR Zbl

[Kubota 1973] T. Kubota, *Elementary theory of Eisenstein series*, Kodansha, Ltd., Tokyo, 1973. MR Zbl

[Lalley 2014] S. P. Lalley, "Statistical regularities of self-intersection counts for geodesics on negatively curved surfaces", *Duke Math. J.* **163**:6 (2014), 1191–1261. MR Zbl

[Lang 1985] S. Lang, $\mathrm{SL}_2(\mathbf{R})$, Graduate Texts in Mathematics **105**, Springer, 1985. MR

[Lester and Radziwiłł 2020] S. Lester and M. Radziwiłł, "Quantum unique ergodicity for half-integral weight automorphic forms", *Duke Math. J.* **169**:2 (2020), 279–351. MR Zbl

[Luo et al. 2009] W. Luo, Z. Rudnick, and P. Sarnak, "The variance of arithmetic measures associated to closed geodesics on the modular surface", *J. Mod. Dyn.* **3**:2 (2009), 271–309. MR Zbl

[Petrow and Young 2019] I. Petrow and M. P. Young, "A generalized cubic moment and the Petersson formula for newforms", *Math. Ann.* **373**:1-2 (2019), 287–353. MR Zbl

[Pollicott and Sharp 2006] M. Pollicott and R. Sharp, "Angular self-intersections for closed geodesics on surfaces", *Proc. Amer. Math. Soc.* **134**:2 (2006), 419–426. MR Zbl

[Rickards 2021] J. Rickards, "Computing intersections of closed geodesics on the modular curve", *J. Number Theory* **225** (2021), 374–408. MR Zbl

[Rivin 2001] I. Rivin, "Simple curves on surfaces", *Geom. Dedicata* **87**:1-3 (2001), 345–360. MR Zbl

[Sardari 2021] N. T. Sardari, "The least prime number represented by a binary quadratic form", *J. Eur. Math. Soc.* (*JEMS*) **23**:4 (2021), 1161–1223. MR Zbl

[Sarnak 1980] P. C. Sarnak, *Prime geodesic theorems*, Ph.D. thesis, Stanford University, 1980, https://www.proquest.com/docview/303065936. MR

[Sarnak 1982] P. Sarnak, "Class numbers of indefinite binary quadratic forms", *J. Number Theory* **15**:2 (1982), 229–247. MR Zbl

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

[Siegel 1935] C. L. Siegel, "Über die Classenzahl quadratischer Zahlkörper", *Acta Arith.* **1**:1 (1935), 83–36. Zbl

[Siegel 1965] C. L. Siegel, *Lectures on advanced analytic number theory*, Tata Institute of Fundamental Research Lectures on Mathematics **23**, Tata Institute of Fundamental Research, Bombay, 1965. MR Zbl

[Zagier 1981] D. Zagier, "Eisenstein series and the Riemann zeta function", pp. 275–301 in *Automorphic forms*, *representation theory and arithmetic* (Bombay, 1979), edited by K. G. Ramanathan, Tata Inst. Fund. Res. Studies in Math. **10**, Tata Institute of Fundamental Research, Bombay, 1981. MR Zbl

Communicated by Philippe Michel
Received 2021-12-09 Revised 2022-05-18 Accepted 2022-07-06

junehyuk_jung@brown.edu *Department of Mathematics, Brown University, Providence, RI, United States*

nzt5208@psu.edu *Department of Mathematics, Pennsylvania State University, State College, PA, 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 17 No. 7 2023