Modular curves, Arakelov theory, algorithmic applications

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 1 september 2010
klokke 11:15 uur
door

Pieter Jan Bruin

geboren te Gouda
in 1983

Samenstelling van de promotiecommissie:

Promotor: prof. dr. S. J. Edixhoven

Copromotor: dr. R. S. de Jong

Overige leden: prof. dr. J.-M. Couveignes (Université de Toulouse II–Le Mirail)
prof. dr. K. Khuri-Makdisi (American University of Beirut)
prof. dr. J. Kramer (Humboldt-Universität zu Berlin)
prof. dr. H. W. Lenstra jr.
prof. dr. P. Stevenhagen

Modular curves, Arakelov theory, algorithmic applications

THOMAS STIELTJES INSTITUTE FOR MATHEMATICS

Het onderzoek voor dit proefschrift werd ondersteund door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Contents

Introduction

This thesis is about arithmetic, analytic and algorithmic aspects of modular curves and modular forms. The arithmetic and analytic aspects are linked by the viewpoint that modular curves are examples of *arithmetic surfaces*. For this reason, *Arakelov theory* (intersection theory on arithmetic surfaces) occupies a prominent place in this thesis. Apart from this, a substantial part of it is devoted to studying modular curves over finite fields, and their Jacobian varieties, from an *algorithmic* viewpoint.

The end product of this thesis is an algorithm for computing *modular Galois representations*. These are certain two-dimensional representations of the absolute Galois group of the rational numbers that are attached to Hecke eigenforms over finite fields. The running time of our algorithm is (under minor restrictions) polynomial in the length of the input. This main result is a generalisation of that of the book [17], which was written by Jean-Marc Couveignes and Bas Edixhoven with contributions from Johan Bosman, Robin de Jong and Franz Merkl.

Although describing such an algorithm has been my principal motivation, several intermediate results are developed in sufficient generality to make them of interest to the study of modular curves and modular forms in a wider sense.

In the remainder of this introduction, we explain the motivating question and outline the strategy for computing modular Galois representations. After that, we state the results of this thesis in more detail, and we compare them to those of Couveignes, Edixhoven et al. We then discuss some applications of our algorithm. The introduction is concluded with a summary of the chapters of this thesis.

Modular Galois representations

By work of Eichler, Shimura, Igusa, Deligne and Serre, one can associate to any Hecke eigenform over a finite field $\mathbf{F}$ a two-dimensional $\mathbf{F}$-linear representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. This means the following. Let n and k be positive integers, and let f be a modular form of weight k for the group

$$\Gamma_1(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \;\middle|\; \begin{array}{r} a \equiv d \equiv 1 \pmod{n}, \\ c \equiv 0 \pmod{n} \end{array} \right\}$$

over a finite field $\mathbf{F}$ of characteristic l. We suppose that f is an eigenform for the Hecke algebra of weight k for $\Gamma_1(n)$. Let K_{nl} be the largest extension of $\mathbf{Q}$ inside $\overline{\mathbf{Q}}$ that is ramified only at primes dividing nl. For every prime number $p \nmid nl$, let Frob_p

denote a Frobenius element at p in $\mathrm{Gal}(K_{nl}/\mathbf{Q})$; this is well-defined up to conjugation. Then there exists a two-dimensional semi-simple representation

$$\rho_f\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}} W_f \quad (\cong \mathrm{GL}_2(\mathbf{F})),$$

where W_f is a two-dimensional $\mathbf{F}$-vector space, such that ρ_f is unramified at all prime numbers $p \nmid nl$ (in other words, ρ_f factors via $\mathrm{Gal}(K_{nl}/\mathbf{Q})$) and such that for every prime number $p \nmid nl$, the characteristic polynomial of $\rho_f(\mathrm{Frob}_p)$ equals $t^2 - a_p t + \epsilon(p)p^{k-1}$, where a_p and $\epsilon(p)$ are the eigenvalues of the Hecke operators T_p and $\langle p\rangle$ on f. The representation ρ_f is unique up to isomorphism; it is called the *modular Galois representation* associated to f.

The main result of this thesis

The goal of the last chapter of this thesis is to give an efficient algorithm for computing representations of the form ρ_f, where f is an eigenform over a finite field $\mathbf{F}$. By "computing ρ_f" we mean producing the following data:

(1) the finite Galois extension K_f of $\mathbf{Q}$ such that ρ_f factors as

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K_f/\mathbf{Q}) \rightarrowtail \mathrm{Aut}_{\mathbf{F}} W_f,$$

given by the multiplication table of some $\mathbf{Q}$-basis $(b_1, \ldots, b_r)$ of K_f;

(2) for every $\sigma \in \mathrm{Gal}(K_f/\mathbf{Q})$, the matrix of σ with respect to the basis $(b_1, \ldots, b_r)$ and the matrix of $\rho_f(\sigma)$ with respect to some fixed $\mathbf{F}$-basis of W_f.

We give a probabilistic algorithm that computes ρ_f. We consider the situation where the weight k is less than the characteristic of $\mathbf{F}$ and where n is of the form ab, where a is a fixed positive integer and b is a squarefree positive integer coprime to a. In this situation we prove that the running time of the algorithm is bounded by a polynomial in the level and weight of the form in question and the cardinality of $\mathbf{F}$. This is essentially optimal, given the fact that the length of the input and output of such an algorithm is already polynomial in the same quantities.

The strategy

The main application of the results in this thesis is a generalisation of that of the book [17] of Couveignes, Edixhoven et al. The basic strategy is the same as that of [17], but there are various differences. We will now explain this strategy, as well as the differences.

The first step, due to Edixhoven, is to reduce the problem to computing representations of the form

$$\rho_{\mathrm{J}_1(n)[\mathfrak{m}]}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}}\bigl(\mathrm{J}_1(n)[\mathfrak{m}](\overline{\mathbf{Q}})\bigr),$$

where n is a positive integer, $\mathrm{J}_1(n)$ is the Jacobian of the modular curve $\mathrm{X}_1(n)$, $\mathfrak{m}$ is a maximal ideal of the Hecke algebra $\mathbf{T}_1(n) \subseteq \mathrm{End}\,\mathrm{J}_1(n)$, $\mathrm{J}_1(n)[\mathfrak{m}]$ is the largest closed subscheme of $\mathrm{J}_1(n)$ annihilated by $\mathfrak{m}$, $\mathbf{F}$ is the residue field $\mathbf{T}_1(n)/\mathfrak{m}$ and $\rho_{\mathrm{J}_1(n)[\mathfrak{m}]}$ is the natural homomorphism. Computing $\rho_{\mathrm{J}_1(n)[\mathfrak{m}]}$ essentially comes down to computing the $\mathbf{F}$-vector space scheme $\mathrm{J}_1(n)[\mathfrak{m}]$ over $\mathbf{Q}$.

The problem of computing $\mathrm{J}_1(n)[\mathfrak{m}]$ is approached by choosing a closed immersion

$$\iota\colon \mathrm{J}_1(n)[\mathfrak{m}] \rightarrowtail \mathbf{A}^1_{\mathbf{Q}}$$

of $\mathbf{Q}$-schemes. The image of ι is of the form $V = \operatorname{Spec} \mathbf{Q}[x]/(P)$ for some $P \in \mathbf{Q}[x]$. This V gets the structure of a finite $\mathbf{F}$-vector space scheme, which is given by polynomials with rational coefficients. The essential idea that makes it possible to efficiently compute V is due to Couveignes. It is to *approximate* V, either over the complex numbers or modulo sufficiently many prime numbers, to sufficient precision to reconstruct it exactly. To find out what precision is sufficient, we need to bound the heights of the coefficients defining V.

In [17] both a deterministic and a probabilistic algorithm are given. The deterministic algorithm uses computations over the complex numbers; the probabilistic variant uses computations over finite fields. It seems hard to remove the probabilistic aspect from the algorithms for computing in Jacobians of curves over finite fields.

In this thesis, we only give an algorithm that works over finite fields. Let us briefly explain the reason for this. The computations in $\mathrm{J}_1(n)$ are done using divisors on $\mathrm{X}_1(n)$ as follows. Let g be the genus of $\mathrm{X}_1(n)$. We fix a divisor D_0 of degree g on $\mathrm{X}_1(n)$. This gives a birational morphism

$$\begin{aligned} \operatorname{Sym}^g \mathrm{X}_1(n) &\to \mathrm{J}_1(n) \\ D &\mapsto [D - D_0]. \end{aligned}$$

In [17], the divisor D_0 is chosen such that this map is an isomorphism over $\mathrm{J}_1(n)[\mathfrak{m}]$. The method of choosing such a divisor that is used in [17] does not work in our more general situation. This problem is solved as follows. We take $D_0 = gO$, where O is a rational cusp of $\mathrm{X}_1(n)$. With this choice, there may be points of $\mathrm{J}_1(n)[\mathfrak{m}]$ for which the representation in the form $[D - D_0]$ is not unique. For every $x \in \mathrm{J}_1(n)[\mathfrak{m}](\overline{\mathbf{Q}})$ we therefore consider the least integer d_x such that $x = [D_x - d_x O]$ for some effective divisor D_x of degree d_x. These D_x are unique; the downside is that we need to compute the d_x. We show how to do this in the variant that uses finite fields, but it is not yet clear how to do the analogous computations over the complex numbers.

The algorithms that we use for computing in Jacobians of modular curves over finite field are different from those used in [17]. Instead of algorithms for computing with singular plane curves, we use the algorithms for computing in Jacobians of projective curves developed by Khuri-Makdisi in [56] and [57], and we transfer the methods of Couveignes [16] to this setting.

A bound on the heights of the coefficients of the data to be computed, and therefore a bound on the running time of the algorithm, is derived using Arakelov intersection theory on models of modular curves over rings of integers of number fields. We follow roughly the same strategy that was applied in [17], but there are some notable differences. First, we have avoided introducing Faltings's δ-invariant, which means we do not need bounds on θ-functions of Jacobians of modular curves. Second, our methods allow us to derive bounds on the amount of work that has to be done to find the numbers d_x defined above. Finally, we introduce new analytic methods to find sharper bounds for various Arakelov-theoretic quantities associated to modular curves.

Applications

We now outline some applications of our main result. This thesis contains no proofs of the theorems below; we refer to [17, Chapter 15] for arguments that can be used to prove them. I hope to give more attention to these applications in a future article.

Computing coefficients of modular forms

The history of [17], and therefore also of this thesis, started with a question that René Schoof asked to Bas Edixhoven in 1995. Ramanujan's τ-function is defined by

$$q\prod_{m=1}^{\infty}(1-q^m)^{24}=\sum_{n=1}^{\infty}\tau(n)q^n.$$

This power series is the q-expansion of the unique cusp form Δ of weight 12 for $\mathrm{SL}_2(\mathbf{Z})$. Schoof's question was: given a prime number p, can one compute $\tau(p)$ in time polynomial in $\log p$? This question is answered affirmatively in [17].

For modular forms for $\Gamma_1(n)$ with $n>1$, the results of this thesis imply the following generalisation.

Theorem. *Let a be a positive integer. There is a probabilistic algorithm that, given a positive integer k, a squarefree positive integer b coprime to a, the q-expansion of a Hecke eigenform f of weight k for $\Gamma_1(ab)$ up to sufficient precision to determine f uniquely, and a positive integer m in factored form, computes the m-th coefficient of f, and that runs in expected time polynomial in b, k and $\log m$ under the generalised Riemann hypothesis for number fields.*

The Riemann hypothesis is needed to ensure the existence of sufficiently many primes of small norm in the number field generated by the coefficients of f. It does not suffice to apply the prime number theorem for each of these fields; we need an error term for the prime number theorem that is sufficiently small relative to the discriminant. More precisely, we use the result that if K is of a number field of discriminant Δ_K for which the generalised Riemann hypothesis holds and $\pi_K(x)$ denotes the number of prime ideals of the ring of integers of K of norm at most x, then

$$\left|\pi_K(x)-\int_2^x\frac{dy}{\log y}\right|\le c\sqrt{x}\log(|\Delta_K|x^{[K:\mathbf{Q}]})\quad\text{for all } x\ge 2,$$

where c is a positive real number not depending on K or x; see Weinberger [111].

Sums of squares

One particularly interesting family of modular forms consists of *θ-series* associated to integral lattices. Let L be an integral lattice of rank k and level n. The θ-series of L is defined by

$$\theta_L=\sum_{x\in L}q^{\langle x,x\rangle}\in\mathbf{Z}[[q]].$$

This power series is the q-expansion of a modular form of weight $k/2$ for $\Gamma_1(4n)$. Our results imply that if k is even, then given θ_L up to sufficient order and a positive integer m in factored form, the m-th coefficient of θ_L can be computed quickly, at least for fixed n.

A case that is worth mentioning specifically is the classical question in how many ways a positive integer can be written as a sum of a given number of squares. For this we introduce Jacobi's θ-series

$$\theta = 1 + 2\sum_{n=1}^{\infty} q^{n^2}.$$

If m and k are positive integers, we write

$$r_k(m) = \#\{(x_1, \ldots, x_k) \in \mathbf{Z}^k \mid x_1^2 + \cdots + x_k^2 = m\}.$$

An elementary combinatorial argument shows that

$$\theta^k = \sum_{n=0}^{\infty} r_k(m) q^m.$$

It is known that θ is the q-expansion of a modular form of weight $1/2$ for $\Gamma_1(4)$. We therefore obtain the following new result on the complexity of evaluating $r_k(m)$.

Theorem. *There is a probabilistic algorithm that, given an even positive integer k and a positive integer m in factored form, computes $r_k(m)$ in time polynomial in k and $\log m$ under the generalised Riemann hypothesis for number fields.*

It was proved recently by Ila Varma [110] that for every even $k \geq 12$, the decomposition of θ^k as as a linear combination of Hecke eigenforms contains cusp forms without complex multiplication. No method was previously known for computing the coefficients of such forms efficiently.

Computing Hecke operators

A consequence of being able to compute coefficients of modular forms is that one can also compute Hecke algebras, in the following sense. Let $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n)))$ be the Hecke algebra acting on cusp forms of weight k for $\Gamma_1(n)$. We represent $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n)))$ by its multiplication table with respect to a suitable $\mathbf{Z}$-basis $(b_1, \ldots, b_r)$, together with the matrices with respect to $(b_1, \ldots, b_r)$ of the Hecke operators T_p for all prime numbers $p \leq \frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma_1(n)]$ and of the diamond operators $\langle d \rangle$ for all $d \in (\mathbf{Z}/n\mathbf{Z})^\times$.

Theorem. *There exists a probabilistic algorithm that, given a positive integer k, a squarefree positive integer n and a positive integer m in factored form, computes the matrix of the Hecke operator T_m in $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n)))$ with respect to $(b_1, \ldots, b_r)$, and that runs in time polynomial in n and $\log m$ under the generalised Riemann hypothesis for number fields.*

The case $k = 2$ of this theorem implies a new result on counting points on modular curves over finite fields. This is because from the elements T_p and $\langle p \rangle$ in $\mathbf{T}(\mathrm{S}_2(\Gamma_1(n)))$ one can compute the characteristic polynomial of the Frobenius endomorphism Frob_p on the l-adic Tate module of $\mathrm{J}_1(n)_{\mathbf{F}_p}$, where l is a prime number different from p.

Theorem. *There exists a probabilistic algorithm that, given a squarefree positive integer n and a prime number $p \nmid n$, computes the zeta function of the modular curve $\mathrm{X}_1(n)$ over $\mathbf{F}_p$, and that runs in time polynomial in n and $\log p$ under the generalised Riemann hypothesis for number fields.*

In particular, this theorem implies that given n and a prime power q coprime to n, the number of rational points on $\mathrm{X}_1(n)$ over the field of q elements can be computed in time polynomial in n and $\log q$ under the generalised Riemann hypothesis.

Explicit realisations of certain Galois groups

The Abelian representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ are well understood: by the Kronecker–Weber theorem, the largest Abelian extension of $\mathbf{Q}$ is obtained by adjoining all roots of unity, and the largest Abelian quotient of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is isomorphic to $\widehat{\mathbf{Z}}^\times$.

Serre's conjecture, which is now a theorem thanks to Khare and Wintenberger, with an important step due to Kisin (see [54], [55] and [61]), asserts that every two-dimensional, odd, irreducible representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over a finite field is associated to a modular form. Our results therefore imply that an important class of non-Abelian extensions of $\mathbf{Q}$ can be computed efficiently.

Computational work based on the work of Couveignes, Edixhoven et al. has been carried out by Johan Bosman using the complex analytic method; see [17, Chapter 7]. In [9], Bosman also gave an explicit polynomial of degree 17 over $\mathbf{Q}$ with Galois group $\mathrm{SL}_2(\mathbf{F}_{16})$; the corresponding Galois representation is attached to a modular form of weight 2 for $\Gamma_0(137)$. It follows from the results of this thesis that analogous calculations of Galois groups can be done efficiently in greater generality.

Overview of the chapters

In **Chapter I**, we introduce modular curves, modular forms and modular Galois representations. This chapter consists mostly of known material.

In **Chapter II**, we prove several analytic results on modular curves that are needed in the later chapters. The most important of these are explicit bounds on Petersson norms and suprema for cusp forms, and on Green functions of quotients of the upper half-plane by cofinite Fuchsian groups.

In **Chapter III**, we describe Arakelov's intersection theory on aritmetic surfaces. We give the results from Arakelov theory that we need for the bounds of the running time of the algorithm that is described in Chapter V. We also find fairly explicit bounds on many Arakelov-theoretic invariants of modular curves.

In **Chapter IV**, we collect the computational tools that are needed for the algorithm. This chapter largely consists of algorithms for computing with projective curves and their Jacobians. We describe a collection of algorithms developed by Khuri-Makdisi, and we develop new algorithms that allow us to work with finite morphisms between curves and with curves over finite fields.

In **Chapter V**, we describe the promised algorithm for computing Galois representations associated to modular forms over finite fields. The algorithm is based on the tools developed in Chapter IV. We use the Arakelov-theoretic methods introduced in Chapter III to bound the heights of the data that need to be computed, and thus to bound the expected running time of our algorithm.

Chapter I

Modular curves, modular forms and Galois representations

In this chapter we collect the necessary preliminaries on modular curves, modular forms and modular Galois representations. We focus entirely on the algebraic side; the analytic side will be explained in Chapter II. Essentially all the material in this chapter is known; only Theorem 3.5 seems to be new.

The set-up of this chapter is geared towards quickly introducing the material and notation we need, rather than towards giving an anywhere near complete introduction. The reader is therefore encouraged to consult one of the many existing texts in which this material, and much more, is explained. These include Deligne and Rapoport [23], Katz and Mazur [53], Conrad [14], Diamond and Im [25], and Diamond and Shurman [26].

1. Modular curves

1.1. Moduli spaces of generalised elliptic curves

To begin with, we describe some of the work of Deligne and Rapoport [23], Drinfeld (unpublished), Katz and Mazur [53], Edixhoven (unpublished) and Conrad [14] on the moduli spaces of (generalised) elliptic curves.

Let S be a scheme. For each positive integer n, the *standard n-gon* (or *Néron n-gon*) over S is the S-scheme obtained by taking n copies of $\mathbf{P}^1_S$ and identifying the section ∞ on the i-th copy with the section 0 on the $(i+1)$-th copy. For $n = 1$, one needs to be a bit careful. The result in this case is the closed subscheme of $\mathbf{P}^2_S$ defined by the equation $y^2z + xyz = x^3$; see Conrad [14, § 2.1].)

A *semi-stable curve of genus 1* over S is a proper, finitely presented and flat morphism $f\colon C \to S$ such that every geometric fibre of f is either a smooth curve of genus 1 or a Néron n-gon for some n. If $f\colon C \to S$ is a semi-stable curve of genus 1, we write C^{sm} for the open subscheme of C consisting of the points at which f is smooth. If $f\colon C \to S$ is a semi-stable curve of genus 1, then the relative dualising sheaf $\Omega_{C/S}$ is a line bundle on C, and the direct image $f_*\Omega_{C/S}$ is a line bundle on S.

A *generalised elliptic curve* over S is a triple $(E, +, 0)$ consisting of a semi-stable curve E of genus 1 over S, a morphism $+\colon E^{\mathrm{sm}} \times_S E \to E$ of S-schemes and a section $0 \in E^{\mathrm{sm}}(S)$ such that the following conditions hold (see Conrad [14, Definition 2.1.4]):

(1) $+$ restricts to a commutative group scheme structure on E^{sm} with identity section 0;

(2) $+$ is an action of E^{sm} on E such that on singular geometric fibres the translation action by each rational point in the smooth locus induces a rotation on the graph of irreducible components.

Let E be a generalised elliptic curve over S. A point $P \in E^{\mathrm{sm}}(S)$ is called a *point of exact order* n if the relative Cartier divisor

$$\langle P\rangle^{(n)} = \sum_{i=1}^{n} [iP]$$

on E^{sm} is a closed subgroup scheme of E^{sm}; see Katz and Mazur [53, § 1.4]. A $\Gamma_1(n)$-*structure* on E is a group homomorphism $\phi\colon \mathbf{Z}/n\mathbf{Z} \to E^{\mathrm{sm}}(S)$ such that $\phi(1)$ is a point of exact order n. A *cyclic subgroup of order* n on E is a subgroup scheme that locally for the *fppf*-topology on S is of the form $\langle P\rangle^{(n)}$ for some point P of exact order n.

Let G be a cyclic subgroup of order n on E. For every divisor d of n, there is a canonical subgroup scheme G_d of G that, again locally for the *fppf*-topology on S, is given by choosing a generator P of G and defining

$$G_d = \langle (n/d)P\rangle^{(d)};$$

see Katz and Mazur [53, Theorem 6.7.2]. This G_d is called the *standard cyclic subgroup of order* d *of* G.

Let E be a generalised elliptic curve over a scheme S, let n be a positive integer, and let p be a prime number. For $p \nmid n$, we define a $\Gamma_1(n;p)$-*structure* on E to be a pair (P, G) consisting of a $\Gamma_1(n)$-structure P on E^{sm} and a cyclic subgroup G of order p on E^{sm} such that the Cartier divisor $\sum_{j\in\mathbf{Z}/p\mathbf{Z}}(jP + G)$ on E is ample. For $p \mid n$, we define a $\Gamma_1(n;p)$-structure in the same way, but we add the condition

$$\sum_{j\in\mathbf{Z}/p\mathbf{Z}} (j(n/p)P + G_p) = E^{\mathrm{sm}}[p],$$

where $G_p \subseteq G$ is the standard cyclic subgroup of order p as defined above.

Let Γ denote $\Gamma_1(n)$ or $\Gamma_1(n;p)$. There exists a moduli stack $\mathcal{M}_\Gamma$ classifying Γ-structures. (For background on stacks, we refer to the book [63] of Laumon and Moret-Bailly.) It is known that $\mathcal{M}_\Gamma$ is a proper flat Deligne–Mumford stack over $\mathbf{Z}$; see Conrad [14, Theorem 1.2.1]. Furthermore, $\mathcal{M}_\Gamma$ is regular and has geometrically connected fibres of pure dimension 1 over $\operatorname{Spec}\mathbf{Z}$. The coarse moduli spaces of $\mathcal{M}_\Gamma$ for $\Gamma = \Gamma_1(n)$ and $\Gamma = \Gamma_1(n;p)$ are denoted by $\mathrm{X}_1(n)$ and $\mathrm{X}_1(n;p)$, respectively.

The stack $\mathcal{M}_\Gamma$ has an open substack consisting exactly of the points with trivial automorphism group, and this open substack is representable by a scheme. This implies, for example, that $\mathcal{M}_{\Gamma_1(n)}$ and $\mathcal{M}_{\Gamma_1(n;p)}$ are representable over $\operatorname{Spec}\mathbf{Z}[1/n]$ for $n \geq 5$ and p prime.

There is a canonical open substack $\mathcal{M}_\Gamma^\circ$ of $\mathcal{M}_\Gamma$ classifying smooth elliptic curves with Γ-structure, and a *divisor of cusps* $\mathcal{M}_\Gamma^\infty$ classifying Néron polygons with Γ-structure. If $\Gamma = \Gamma_1(n;p)$, we identify $\mathcal{M}_\Gamma^\circ$ with the stack classifying pairs of the form $(E \xrightarrow{\phi} E', P)$, where ϕ is a cyclic isogeny of degree p whose kernel has trivial intersection with $\langle P\rangle^{(n)}$, in the following way. Given a cyclic subgroup G, we take $E' = E/G$ and take ϕ to be the quotient map; conversely, given ϕ, we take G to be the kernel of ϕ.

If $E \to S$ is a generalised elliptic curve and $\Omega_{E/S}$ is the relative dualising sheaf, then $f_*\Omega_{E/S}$ is a line bundle on S whose formation is compatible with base change on S. This gives us the *line bundle of modular forms of weight 1* on $\mathcal{M}_\Gamma$, denoted by ω_Γ.

1.2. Maps between moduli spaces

There are various canonical morphisms between the moduli stacks defined above; see Conrad [14, Lemma 4.2.3]. These preserve the open substack $\mathcal{M}_\Gamma^\circ$ and $\mathcal{M}_\Gamma^\infty$.

First, let n be a positive integer, and let p be a prime number. The p-th *Hecke correspondence* on $\mathcal{M}_{\Gamma_1(n)}$ is the diagram

$$\begin{tikzcd} & \mathcal{M}_{\Gamma_1(n;p)} \arrow[dl, "q_1"'] \arrow[dr, "q_2"] & \\ \mathcal{M}_{\Gamma_1(n)} & & \mathcal{M}_{\Gamma_1(n)} \end{tikzcd} \tag{1.1}$$

where q_1 and q_2 are defined on the open substack $\mathcal{M}_{\Gamma_1(n;p)}^\circ$ classifying smooth elliptic curves by

$$q_1(E \xrightarrow{\phi} E', P) = (E, P) \quad \text{and} \quad q_2(E \xrightarrow{\phi} E', P) = (E', \phi \circ P).$$

By Conrad's result in [14, Theorem 1.2.2] the morphisms q_1 and q_2 extend uniquely to finite flat morphisms $\mathcal{M}_{\Gamma_1(n;p)} \to \mathcal{M}_{\Gamma_1(n)}$.

Furthermore, for all $d \in (\mathbf{Z}/n\mathbf{Z})^\times$, we define an automorphism

$$r_d\colon \mathcal{M}_{\Gamma_1(n)} \xrightarrow{\sim} \mathcal{M}_{\Gamma_1(n)} \tag{1.2}$$

by the modular interpretation

$$r_d(E, P) = (E, dP)$$

for all generalised elliptic curves E together with a $\Gamma_1(n)$-structure P.

Finally, let m be a divisor of n. Then for each divisor $d \mid (n/m)$ there exists a natural morphism

$$b_d^{n,m}\colon \mathcal{M}_{\Gamma_1(n)} \to \mathcal{M}_{\Gamma_1(m)}$$

defined on (smooth) elliptic curves with $\Gamma_1(n)$-structure by sending a pair (E, P) to $(E/\langle(n/d)P\rangle_d, (n/md)P \bmod \langle(n/d)P\rangle_d)$.

1.3. Jacobians of modular curves

Let n be an integer such that $n \geq 5$. The modular stack $\mathcal{M}_{\Gamma_1(n)}$ over $\operatorname{Spec} \mathbf{Z}[1/n]$ is representable by a proper smooth curve $\mathrm{X}_1(n)$ over $\operatorname{Spec} \mathbf{Z}[1/n]$ with geometrically connected fibres. Because of this, there exists an Abelian scheme $\mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ over $\operatorname{Spec} \mathbf{Z}[1/n]$ representing the functor $\operatorname{Pic}^0_{\mathrm{X}_1(n)/\mathbf{Z}[1/n]}$, i.e. the connected component of the identity element of the Picard functor. For details, we refer to Bosch, Lütkebohmert and Raynaud [8, Chapter 9].

For any prime number p we can now view the Hecke correspondence (1.1) as a correspondence on $\mathrm{X}_1(n)$, and use it to define an endomorphism T_p of $\mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$, called the p-th *Hecke operator*, as

$$T_p = \operatorname{Alb}(q_2) \circ \operatorname{Pic}(q_1).$$

This is *a priori* defined on $\mathrm{J}_1(n)_{\mathbf{Z}[1/np]}$, but it extends uniquely to an endomorphism of $\mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ since the latter is an Abelian scheme. For $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ we define the *diamond operator* $\langle d \rangle$ on $\mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ to be the automorphism

$$\langle d \rangle = \operatorname{Alb}(r_d).$$

We define the *Hecke algebra* for $\Gamma_1(n)$ as the subring

$$\mathbf{T}_1(n) \subseteq \operatorname{End} \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$$

generated by the endomorphisms T_p for p prime and $\langle d \rangle$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$. It is known that the Hecke algebra $\mathbf{T}_1(n)$ is commutative; see for example Miyake [80, § 4.5].

We introduce some more notation for the case that $n = n_1 n_2$ with given coprime integers n_1 and n_2. Then the Chinese remainder theorem implies that

$$(\mathbf{Z}/n_1 n_2 \mathbf{Z})^\times \cong (\mathbf{Z}/n_1\mathbf{Z})^\times \times (\mathbf{Z}/n_2\mathbf{Z})^\times.$$

For $d_1 \in (\mathbf{Z}/n_1\mathbf{Z})^\times$, we define

$$\langle d_1 \rangle_{n_1} = \langle d \rangle,$$

where d is the unique element of $(\mathbf{Z}/n\mathbf{Z})^\times$ with $(d \bmod n_1) = d_1$ and $(d \bmod n_2) = 1$. Then we can decompose $\langle d \rangle$ for any $d \in (\mathbf{Z}/n\mathbf{Z})$ as

$$\langle d \rangle = \langle d \bmod n_1 \rangle_{n_1} \langle d \bmod n_2 \rangle_{n_2}.$$

It is also useful at times to consider the *duals* of the Hecke operators, which are defined by

$$T_p^\vee = \operatorname{Alb}(q_1) \circ \operatorname{Pic}(q_2) \quad \text{for } p \text{ prime}$$

and

$$\langle d \rangle^\vee = \operatorname{Pic}(r_d) = \langle d^{-1} \rangle \quad \text{for } d \in (\mathbf{Z}/n\mathbf{Z})^\times.$$

Remark. In the case where p divides n, the operator T_p is often denoted by U_p in the literature, but we will not do this. Also, some authors, such as Ribet [88, page 444], define the operators T_p and $\langle d \rangle$ in the opposite way, i.e. as the duals of the endomorphisms defined above. The subring of $\operatorname{End} \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ generated by these endomorphisms is isomorphic to the Hecke algebra $\mathbf{T}_1(n)$ defined above via the Rosati involution on $\operatorname{End} \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$. (The Rosati involution is actually an anti-isomorphism, but this does not matter since $\mathbf{T}_1(n)$ is commutative.) There does not seem to be a strong reason to prefer either of the two definitions, but our choice is motivated by the convention that in the representation ρ_f associated to an eigenform f with $T_p f = a_p f$ for p prime and $\langle d\rangle f = \epsilon(d) f$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$, the characteristic polynomial of a Frobenius element at a prime number p is $X^2 - a_p X + \epsilon(p)p^{k-1}$, as opposed to $X^2 - (a_p/\epsilon(p))X + p^{k-1}/\epsilon(p)$; compare § 1.4.

For later use, we state here a result on the non-vanishing of certain finite subgroup schemes of Jacobians of modular curves.

Lemma 1.1. *Let A be a complex Abelian variety (viewed as a complex manifold) and R a commutative subring of* $\operatorname{End} A$. *For every maximal ideal $\mathfrak{m}$ of R, the subgroup*

$$A[\mathfrak{m}] = \{x \in A \mid rx = 0 \text{ for all } r \in \mathfrak{m}\}$$

is non-zero.

Proof. The homology group $\mathrm{H}_1(A, \mathbf{Z})$ is a faithful, finitely generated R-module. For any maximal ideal $\mathfrak{m} \subset R$, the localisation $\mathrm{H}_1(A, \mathbf{Z})_\mathfrak{m}$ is therefore a faithful, finitely generated module over the local ring $R_\mathfrak{m}$. Because R is finitely generated as a Abelian group, $\mathfrak{m}$ contains a prime number l and $A[\mathfrak{m}]$ is contained in the group $A[l]$ of l-torsion points of A. From the canonical isomorphism

$$A[l] \cong \mathrm{H}_1(A, \mathbf{Z})/l\mathrm{H}_1(A, \mathbf{Z})$$

we get a canonical isomorphism

$$\begin{aligned} A[\mathfrak{m}] &\cong (\mathrm{H}_1(A, \mathbf{Z})/l\mathrm{H}_1(A, \mathbf{Z}))[\mathfrak{m}] \\ &\cong (\mathrm{H}_1(A, \mathbf{Z})_\mathfrak{m}/l\mathrm{H}_1(A, \mathbf{Z})_\mathfrak{m})[\mathfrak{m}]. \end{aligned}$$

Since l is in the maximal ideal $\mathfrak{m}R_\mathfrak{m}$ of $R_\mathfrak{m}$, Nakayama's lemma implies that the $R_\mathfrak{m}$-module $\mathrm{H}_1(A, \mathbf{Z})_\mathfrak{m}/l\mathrm{H}_1(A, \mathbf{Z})_\mathfrak{m}$ is non-zero. As this module has finite cardinality, it admits a composition chain whose constituents are isomorphic to $R/\mathfrak{m}$ (the only simple $R_\mathfrak{m}$-module). The above isomorphism now shows that $A[\mathfrak{m}] \neq 0$. □

Lemma 1.2. *Let n be an integer with $n \geq 5$, and let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}_1(n)$. Let $J = \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$, and let $J[\mathfrak{m}]$ be the maximal closed subscheme of J annihilated by $\mathfrak{m}$. Then $J[\mathfrak{m}]$ is a non-zero closed subgroup scheme of J and is étale over* $\operatorname{Spec} \mathbf{Z}[1/nl]$, *where l is the residue characteristic of $\mathfrak{m}$.* □

Proof. The claim that $J[\mathfrak{m}]$ is non-zero follows from Lemma 1.1. Since $J[l]$ is étale, the closed subscheme of $J[l]$ that is sent to zero by any Hecke operator is a union of irreducible components of $J[l]$. The scheme $J[\mathfrak{m}]$ is the intersection of these subschemes and is therefore étale as well. □

1.4. The Eichler–Shimura relation

Let n be a positive integer, and let p be a prime number not dividing n. We write Frob_p for the the Frobenius endomorphism of the Abelian variety

$$\mathrm{J}_1(n)_{\mathbf{F}_p} = \mathrm{J}_1(n)_{\mathbf{Z}[1/n]} \times \operatorname{Spec} \mathbf{F}_p$$

and Ver_p for the *Verschiebung*, i.e. the unique endomorphism of $\mathrm{J}_1(n)_{\mathbf{F}_p}$ such that

$$\mathrm{Frob}_p \mathrm{Ver}_p = \mathrm{Ver}_p \mathrm{Frob}_p = p \in \operatorname{End} \mathrm{J}_1(n)_{\mathbf{F}_p}.$$

Then the *Eichler–Shimura relation*

$$T_p = \mathrm{Frob}_p + \langle p\rangle \mathrm{Ver}_p \tag{1.3}$$

holds in $\operatorname{End} \mathrm{J}_1(n)_{\mathbf{F}_p}$; see Diamond and Im [25, § 8.5 and § 10.2] or Gross [41, Proposition 3.12]. Moreover, if l is a prime number different from the characteristic of p, then the Tate module

$$\mathrm{V}_l(\mathrm{J}_1(n)_{\mathbf{F}_p}) = \mathbf{Q}_l \otimes_{\mathbf{Z}_l} \varprojlim_r \mathrm{J}_1(n)_{\mathbf{F}_p}[l^r](\overline{\mathbf{F}}_p)$$

is a free module of rank 2 over $\mathbf{Q}_l \otimes \mathbf{T}_1(n)$, and the characteristic polynomial of Frob_p on this space is equal to

$$\chi_{\mathbf{Q}_l \otimes \mathbf{T}_1(n)}(\mathrm{Frob}_p) = x^2 - T_p x + p\langle p\rangle \in \mathbf{T}_1(n)[x];$$

see Diamond and Im [25], § 12.5 or Gross [41, Proposition 11.8].

2. Modular forms

Let Γ denote $\Gamma_1(n)$ or $\Gamma_1(n;p)$ for a positive integer n and a prime number p. We define the moduli stack $\mathcal{M}_\Gamma$ over $\operatorname{Spec} \mathbf{Z}$ and the line bundle ω_Γ on $\mathcal{M}_\Gamma$ as in § 1.1. For any non-negative integer k and any Abelian group A, we define the Abelian group of *modular forms* of weight k for Γ with coefficients in A as

$$\mathrm{M}_k(\Gamma, A) = \mathrm{H}^0(\mathcal{M}_\Gamma, A \otimes_{\mathbf{Z}} \omega_\Gamma^{\otimes k}). \tag{2.1}$$

This gives a functor on the category of Abelian groups. Furthermore, if A and B are Abelian groups, there are multiplication maps

$$\mathrm{M}_k(\Gamma, A) \otimes \mathrm{M}_l(\Gamma, B) \to \mathrm{M}_{k+l}(\Gamma, A \otimes B) \qquad (k, l \geq 0)$$

and if R is any ring, then $\bigoplus_{k\geq 0} \mathrm{M}_k(\Gamma, R)$ is in a natural way a graded R-algebra.

If $n \geq 5$, $k \geq 2$, and A is a $\mathbf{Z}[1/n]$-module, then the canonical map

$$A \otimes_{\mathbf{Z}} \mathrm{M}_k(\Gamma, \mathbf{Z}) \to \mathrm{M}_k(\Gamma, A).$$

is an isomorphism. This is not the case in general. For example, if p is a prime number, the canonical reduction map

$$\mathrm{M}_1(\Gamma, \mathbf{Z}) \to \mathrm{M}_1(\Gamma, \mathbf{F}_p)$$

is not always surjective. For this reason, modular forms of weight 1 often require a more careful treatment.

2.1. Cusp forms

We recall from § 1.1 that the moduli stack $\mathcal{M}_\Gamma$ is the union of the open substack $\mathcal{M}^\circ_\Gamma$, classifying (smooth) elliptic curves, and the divisor of cusps, classifying Néron polygons. For any Abelian group A, we define the subgroup of *cusp forms* inside the group $\mathrm{M}_k(\Gamma, A)$ of modular forms as

$$\mathrm{S}_k(\Gamma, A) = \mathrm{H}^0\bigl(\mathcal{M}_\Gamma, A \otimes_{\mathbf{Z}} \omega_\Gamma^{\otimes k}(-\mathrm{cusps})\bigr).$$

As is the case for the full space of modular forms, for $n \geq 5$, $k \geq 2$ and A a $\mathbf{Z}[1/n]$-module, the map

$$A \otimes_{\mathbf{Z}} \mathrm{S}_k(\Gamma, \mathbf{Z}) \to \mathrm{S}_k(\Gamma, A).$$

is an isomorphism; this is not true in general.

The maps $b_e^{n,d}$ defined in § 1.2 respect the divisor of cusps. This implies that the induced maps

$$(b_e^{n,d})^*\colon \mathrm{M}_k(\Gamma_1(d), A) \to \mathrm{M}_k(\Gamma_1(n), A) \quad (d \mid n \text{ and } e \mid n/d)$$

preserve the subgroup of cusp forms.

2.2. Hecke algebras on spaces of modular forms

Let n and k be positive integers, and let A be any Abelian group. The Abelian group $\mathrm{M}_k(\Gamma_1(n), A)$ of modular forms of weight k for $\Gamma_1(n)$ with coefficients in A, as defined in (2.1), admits a natural action of the *Hecke operators* T_p for p prime and $\langle d\rangle$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$. We will briefly sketch how these operators are defined.

First, for every prime number p, the Hecke correspondence (1.1) induces an endomorphism of $\mathrm{M}_k(\Gamma_1(n), A)$, denoted by T_p. Its definition is somewhat complicated, especially if A has non-trivial p-torsion. We therefore assume that multiplication by p is injective on A, and we refer to Conrad [14, § 4.5] for the construction in the general case. On the open substack $\mathcal{M}^\circ_{\Gamma_1(n)}$ of $\mathcal{M}_{\Gamma_1(n)}$, we have the universal p-isogeny ϕ as in § 1.1. There is an induced pull-back map

$$\phi^*\colon q_2^*\omega_{\Gamma_1(n)} \to \omega_{\Gamma_1(n;p)} = q_1^*\omega_{\Gamma_1(n)}$$

on $\mathcal{M}^\circ_{\Gamma_1(n;p)}$. This map can be extended to all of $\mathcal{M}_{\Gamma_1(n;p)}$; see Conrad [14, Theorem 1.2.2]. Furthermore, the fact that q_1 is finite flat implies that there is a natural trace map

$$\mathrm{tr}_{q_1}\colon \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n;p)}, \omega^{\otimes k}_{\Gamma_1(n;p)}) = \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n;p)}, q_1^*\omega^{\otimes k}_{\Gamma_1(n)}) \longrightarrow \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n)}, \omega^{\otimes k}_{\Gamma_1(n)}).$$

The Hecke operator T_p on the Abelian group

$$\mathrm{M}_k(\Gamma_1(n), A) = \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n)}, \omega^{\otimes k}_{\Gamma_1(n)})$$

can now be defined by

$$pT_p = \mathrm{tr}_{q_1} \circ \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n;p)}, \phi^*) \circ q_2^*.$$

Indeed, we have assumed that multiplication by p is injective on p, and the right-hand side is divisible by p; see Conrad [14, Theorem 4.5.1].

Next we introduce the *diamond operator* $\langle d\rangle$ on $\mathrm{M}_k(\Gamma_1(n), A)$ for every $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ as the automorphism of $\mathrm{M}_k(\Gamma_1(n), A)$ induced by pull-back via the automorphism r_d of $\mathcal{M}_{\Gamma_1(n)}$. Here we have used the fact that $r_d^*\omega_{\Gamma_1(n)}$ is naturally isomorphic to $\omega_{\Gamma_1(n)}$.

The *Hecke algebra* on the space of modular forms with coefficients in A is the subring

$$\mathbf{T}(\mathrm{M}_k(\Gamma_1(n), A)) \subseteq \operatorname{End}\mathrm{M}_k(\Gamma_1(n), A)$$

generated by the Hecke operators acting on $\mathrm{M}_k(\Gamma_1(n), A)$.

If K is a field, a *(Hecke) eigenform* of weight k for $\Gamma_1(n)$ over K is a non-zero element

$$f \in \mathrm{M}_k(\Gamma_1(n), K)$$

such that the one-dimensional K-linear subspace of $\mathrm{M}_k(\Gamma_1(n), K)$ spanned by f is stable under the action of $\mathbf{T}(\mathrm{M}_k(\Gamma_1(n), K)$.

Since the maps defining the Hecke correspondences respect the divisor of cusps, the action of the Hecke algebra $\mathbf{T}(\mathrm{M}_k(\Gamma_1(n), A))$ preserves the subgroup $\mathrm{S}_k(\Gamma_1(n), A)$ of cusp forms. In other words, we have a canonical ring homomorphism

$$\mathbf{T}(\mathrm{M}_k(\Gamma_1(n), A)) \to \operatorname{End}\mathrm{S}_k(\Gamma_1(n), A).$$

The image of this homomorphism is a subring of $\operatorname{End}\mathrm{S}_k(\Gamma_1(n), A)$ called the Hecke algebra on the space of cusp forms. We denote it by $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n), A))$.

Similarly to the case of Hecke operators on Jacobians, we can also consider the *duals* of the Hecke operators defined above on spaces of modular forms. The dual of T_p for p prime is defined by

$$pT_p^\vee = \operatorname{tr}_{q_2} \circ \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n;p)}, \hat{\phi}^*) \circ q_1^*,$$

where $\hat{\phi}^*$ is given on $\mathcal{M}^\circ_{\Gamma_1(n)}$ by pullback via the dual of the universal p-isogeny ϕ. We have

$$T_p^\vee = \langle p\rangle^{-1} T_p \quad \text{for } p \nmid n \text{ prime.}$$

For $p \mid n$ prime, the operators T_p and $T_p^\vee$ do not in general commute. The duals of the diamond operators are defined by

$$\langle d\rangle^\vee = \langle d\rangle^{-1} \quad \text{for } d \in (\mathbf{Z}/n\mathbf{Z})^\times.$$

Finally, we note that the maps $b_e^{n,d}$ introduced in § 1.2 induce natural maps

$$(b_e^{n,d})^*\colon \mathrm{M}_k(\Gamma_1(d), A) \to \mathrm{M}_k(\Gamma_1(n), A) \quad \text{for } d \mid n \text{ and } e \mid n/d.$$

2.3. A connection between Hecke algebras on Jacobians and on spaces of cusp forms

Let n be a positive integer, let l be a prime number not dividing n, and let k be an integer such that

$$2 \le k \le l+1. \tag{2.2}$$

We now come to the rather subtle point that the $\mathbf{F}_l$-vector space $\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l)$ of cusp forms can be viewed as a module over the ring $\mathbf{T}_1(n) \subseteq \operatorname{End} \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ if $k = 2$, and over the ring $\mathbf{T}_1(nl) \subseteq \operatorname{End} \mathrm{J}_1(nl)_{\mathbf{Z}[1/nl]}$ if $3 \le k \le l+1$.

Remark. This fact is a basic ingredient for the algorithms of Chapter V. At first sight, the condition (2.2) puts a restriction on the set of modular forms for which we can compute Galois representations. However, this restriction is only superficial, because up to twists all modular Galois representations arise from eigenforms of weight k over finite fields of characteristic l for which the inequality (2.2) holds; see Serre [99, page 116] or Edixhoven [31, Theorem 3.4]. We will explain this in more detail when we need it.

We start with the case $k = 2$. The Hecke algebra $\mathbf{T}_1(n) \subseteq \operatorname{End} \mathrm{J}_1(n)_{\mathbf{Z}[1/n]}$ acts in a natural way on the space $\mathrm{S}_2(\Gamma_1(n), \mathbf{Z})$. One way to see this is using the injective homomorphism

$$\operatorname{End}(\mathrm{J}_1(n)_{\mathbf{Z}[1/n]}) \to \operatorname{End}(\mathrm{J}_1(n)_{\mathbf{C}}),$$

the isomorphism

$$\mathrm{J}_1(n)_{\mathbf{C}} \cong \mathrm{H}^0(\mathrm{X}_1(n)(\mathbf{C}), \Omega^1_{\mathrm{X}_1(\mathbf{C})})^\vee / \mathrm{H}_1(\mathrm{X}_1(n)(\mathbf{C}), \mathbf{Z})$$

and the Kodaira–Spencer isomorphism

$$\mathrm{H}^0(\mathrm{X}_1(n)_{\mathbf{C}}, \Omega^1_{\mathrm{X}_1(n)}) \xrightarrow{\sim} \mathrm{S}_2(\Gamma_1(n), \mathbf{C}).$$

One can check that these isomorphisms are compatible with the action of the Hecke operators. From the fact that the subgroup $\mathrm{S}_2(\Gamma_1(n), \mathbf{Z})$ of $\mathrm{S}_2(\Gamma_1(n), \mathbf{C})$ is stabilised by the Hecke algebra, one then deduces that there exists a ring isomorphism

$$\mathbf{T}_1(n) \to \mathbf{T}(\mathrm{S}_2(\Gamma_1(n), \mathbf{Z}))$$

sending each of the Hecke operators T_m with $m \ge 1$ and $\langle d \rangle$ with $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ in $\mathbf{T}_1(n)$ to the operator in $\mathbf{T}(\mathrm{S}_2(\Gamma_1(n), \mathbf{Z}))$ denoted by the same symbol. For each prime number l, the existence of the isomorphism

$$\mathbf{F}_l \otimes_{\mathbf{Z}} \mathrm{S}_2(\Gamma_1(n), \mathbf{Z}) \xrightarrow{\sim} \mathrm{S}_2(\Gamma_1(n), \mathbf{F}_l)$$

implies that $\mathbf{T}_1(n)$ also acts on $\mathrm{S}_2(\Gamma_1(n), \mathbf{F}_l)$.

When $3 \le k \le l+1$, the situation is more complicated. In this case the Hecke algebra $\mathbf{T}_1(nl) \subseteq \operatorname{End} \mathrm{J}_1(nl)_{\mathbf{Z}[1/nl]}$ acts in a natural way on the $\mathbf{F}_l$-vector space $\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l)$. In other words, there is a surjective ring homomorphism

$$\mathbf{T}_1(nl) \to \mathbf{T}(\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l))$$

sending each of the operators T_p for p prime and $\langle d\rangle_n$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ (using the notation of § 1.3) to the corresponding operator on $\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l)$. Note that T_l is a somewhat subtle case, since l divides the level on the left-hand side but not on the right-hand side. Furthermore, it sends the operator $\langle d\rangle_l$ to $d^{k-2} \in \mathbf{F}_l^\times$ for each $d \in (\mathbf{Z}/l\mathbf{Z})^\times$. Another way to phrase the effect on the diamond operators is to say that $\langle d\rangle \mapsto \langle d \bmod n\rangle (d \bmod l)^{k-2}$. The construction of the action just described is essentially given by Edixhoven in [31, § 6.7].

2.4. The Tate curve and q-expansions

We give here the basic facts about Tate curves. For details, we refer to Deligne and Rapoport [23, VII, § 1] and Conrad [14, § 2.5].

For every positive integer d, the d-th *Tate curve* is a certain generalised elliptic curve

$$f^{(d)}\colon \underline{\mathrm{Tate}}(q^d) \to \operatorname{Spec} \mathbf{Z}[[q]]$$

that becomes a Néron d-gon after base change to the zero locus of q and that is a (smooth) elliptic curve over $\operatorname{Spec} \mathbf{Z}[[q]][q^{-1}]$, the complement of this zero locus. The relative dualising sheaf $\Omega_{\underline{\mathrm{Tate}}(q^d)/\operatorname{Spec}\mathbf{Z}[[q]]}$ admits a canonical generating section α, giving a trivialisation

$$\mathcal{O}_{\operatorname{Spec}\mathbf{Z}[[q]]} \xrightarrow{\sim} f^{(d)}_* \Omega_{\underline{\mathrm{Tate}}(q^d)/\operatorname{Spec}\mathbf{Z}[[q]]}.$$

Consider a positive integer n, and let d and e be positive integers such that n is the least common multiple of d and e. Then the curve $\underline{\mathrm{Tate}}(q^d)$ over $\operatorname{Spec} \mathbf{Z}[[q, \zeta_e]]$ admits at least one $\Gamma_1(n)$-structure. Each choice of d, e and a $\Gamma_1(n)$-structure gives rise to a morphism

$$\operatorname{Spec} \mathbf{Z}[[q, \zeta_e]] \to \mathcal{M}_{\Gamma_1(n)}.$$

The pull-back of $\omega_{\Gamma_1(n)}$ via this map is canonically trivialised by α. For any Abelian group A, this gives an injective map

$$\mathrm{M}_k(\Gamma_1(n), A) = \mathrm{H}^0(\mathcal{M}_{\Gamma_1(n)}, A \otimes \omega_{\Gamma_1(n)}^{\otimes k}) \rightarrowtail A \otimes_{\mathbf{Z}} \mathbf{Z}[[q, \zeta_e]],$$

called the *q-expansion map* relative to $\underline{\mathrm{Tate}}(q^d)$ with the given $\Gamma_1(n)$-structure. As an important special case, we consider the $\Gamma_1(n)$-structure ϕ on $\underline{\mathrm{Tate}}(q^n)$ over $\operatorname{Spec} \mathbf{Z}[[q]]$ given by $\phi(i) = q^i$ for $i \in \mathbf{Z}/n\mathbf{Z}$. We call the corresponding q-expansion the *q-expansion at 0* (because of the connection with complex modular curves). For any $f \in \mathrm{M}_k(\Gamma_1(n), A)$ we define $a_m(f)$ to be the m-th coefficient in this q-expansion.

Via a calculation on $\underline{\mathrm{Tate}}(q^n)$, we can express the action of the *duals* of the Hecke operators, as defined in § 2.2, in terms of the q-expansion at 0 by the well-known formula

$$a_m(T_p^\vee f) = a_{pm}(f) + p^{k-1} a_{m/p}(\langle p\rangle^\vee f) \quad \text{for } m \geq 1 \text{ and } p \text{ prime},$$

where the rightmost term is omitted if p divides n or if p does not divide m; see for example Diamond and Im [25, equation 12.4.1].

Remark. The reason for using the duals of the Hecke operators is that the cusp ∞, which is the more traditional choice for q-expansions, is not $\mathbf{Z}$-rational, but only $\mathbf{Z}[\zeta_n]$-rational. This follows from the moduli interpretation of this cusp: it corresponds to a Néron 1-gon, whose smooth locus is isomorphic to the multiplicative group, with an n-th root of unity as the distinguished n-torsion point. We therefore consider the q-expansion at the "dual" cusp 0, which is $\mathbf{Z}$-rational.

Another calculation using $\underline{\mathrm{Tate}}(q^n)$ shows that the effect of the maps $b_e^{n,d}$ on the q-expansion at the cusp 0 is given by

$$a_i((b_d^{n,m})^* f) = a_{i/e}(f) \quad \text{if } n/m = de,$$

where the right-hand side is to be interpreted as 0 if $e \nmid i$. (Again this is different from the effect on q-expansions at the cusp ∞, where the correct expression on the right-hand side is $a_{i/d}(f)$.)

Let p be a prime number. We write 0 for the cusp of $\mathcal{M}_{\Gamma_1(n;p)}$ corresponding to the Néron n-gon obtained by n copies of $\mathbf{P}^1$ indexed by $\mathbf{Z}/n\mathbf{Z}$, where the distinguished point of order n is the point 1 on the copy indexed by 1, and the distinguished subgroup of order p is the subgroup μ_p of the copy indexed by 0. For every non-negative integer k and every Abelian group A, the maps

$$q_1, q_2 \colon \mathcal{M}_{\Gamma_1(n;p)} \to \mathcal{M}_{\Gamma_1(n)}$$

defining the Hecke correspondence (1.1) induce morphisms

$$q_1^*, q_2^* \colon \mathrm{M}_k(\Gamma_1(n), A) \to \mathrm{M}_k(\Gamma_1(n;p), A).$$

A calculation on the Tate curve $\underline{\mathrm{Tate}}(q^n)$ shows that

$$a_i(q_1^* f) = a_i(f) \quad \text{and} \quad a_i(q_2^* f) = \begin{cases} p^k a_{i/p}(f) & \text{if } i \mid p, \\ 0 & \text{if } i \nmid p \end{cases}$$

for all $f \in \mathrm{M}_k(\Gamma_1(n), A)$ and all $i \geq 0$.

The following basic but very useful fact shows how many coefficients of the q-expansion are needed to determine a modular form uniquely. This is a simple case of a more general result proved by Sturm [105].

Lemma 2.1. *Let Γ be one of the groups $\Gamma_1(n)$ or $\Gamma_1(n;p)$ with $n \geq 1$ and p prime. Let f be a modular form of weight k for Γ over a field whose characteristic does not divide n. If the q-expansion $\sum_{m=0}^{\infty} a_m q^m$ of f at some cusp satisfies*

$$a_r = 0 \text{ for } r \leq \frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma],$$

then $f = 0$.

Proof. This follows from the fact that the line bundle of modular forms of weight k on $\mathcal{M}_\Gamma$ has degree $\frac{k}{24}[\mathrm{SL}_2(\mathbf{Z}) : \Gamma]$, together with the fact that the automorphism group of a Néron polygon with Γ-structure has order 1 if $-1 \notin \Gamma$ and order 2 if $-1 \in \Gamma$. $\square$

Let n and k be positive integers. The q-expansion principle gives us a canonical $\mathbf{Z}$-bilinear pairing

$$\begin{aligned} \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \times \mathrm{S}_k(\Gamma_1(n),\mathbf{Z}) &\longrightarrow \mathbf{Z} \\ (t,f) &\longmapsto a_1(t^\vee f) \end{aligned} \tag{2.3}$$

between the Hecke algebra and the space of cusp forms. Moreover, this pairing is bilinear over $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$ in the sense that $(tt',f) = (t,t'f)$ for all $f \in \mathrm{S}_k(\Gamma_1(n),\mathbf{Z})$ and all $t,t' \in \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$; this follows immediately from the definition. After changing the base to $\mathbf{Z}[1/n]$, the above pairing becomes perfect.

In addition to $\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})$, we will also be interested in the $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$-module

$$\mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n)) = \{f \in \mathrm{S}_k(\Gamma_1(n),\mathbf{Q}) \mid \text{ the } q\text{-expansion of } f \text{ at } 0 \text{ has coefficients in } \mathbf{Z}\}.$$

The advantage of this module is that the pairing

$$\begin{aligned} \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \times \mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n)) &\longrightarrow \mathbf{Z} \\ (t,f) &\longmapsto a_1(t^\vee f) \end{aligned}$$

is perfect over $\mathbf{Z}$; see Ribet [87, Theorem 2.2].

Now let K be a field of characteristic not dividing n. Then we have

$$K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z}) = K \otimes \mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n)),$$

and the pairing (2.3) induces a perfect K-bilinear pairing

$$\bigl(K \otimes \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))\bigr) \times \bigl(K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z})\bigr) \longrightarrow K \tag{2.4}$$

This pairing gives rise to a canonical bijection between the set of ring homomorphisms $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \to K$ and the set of lines in the K-vector space $K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z})$ that are stable under the action of $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$. More precisely, this bijection is given as follows. We identify $K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z})$ with a K-linear subspace of $\mathrm{S}_k(\Gamma_1(n),K)$; this is in fact the whole space, except possibly when $k=1$ and K is of non-zero characteristic. For any eigenform

$$f \in K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z}),$$

there is a corresponding ring homomorphism

$$\mathrm{ev}_f\colon \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \to K$$

sending each Hecke operator to its eigenvalue on f. Conversely, given a ring homomorphism

$$\phi\colon \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \to K,$$

the kernel of the induced homomorphism

$$1 \otimes \phi\colon K \otimes \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \to K$$

is a K-linear subspace of codimension 1, so its annihilator in $K \otimes \mathrm{S}_k(\Gamma_1(n),\mathbf{Z})$ with respect to the pairing (2.4) is a one-dimensional K-linear subspace spanned by some eigenform.

3. Modular Galois representations

In this section we introduce certain representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ associated to Hecke eigenforms, called *modular Galois representations*. Such representations can be defined over finite extensions of either $\mathbf{F}_l$ or $\mathbf{Q}_l$, where l is a prime number. All representations will be assumed to be continuous without further mention.

We will describe how to associate to a given Hecke eigenform f over a field of characteristic 0 a family of λ-adic representations, where λ runs over all primes of the number field K_f obtained by adjoining the Hecke eigenvalues of f to $\mathbf{Q}$. This gives an example of a *compatible family* of l-adic representations [97, chapitre I, n° 2.3]. In the case where f is of weight 1, all such l-adic representations can moreover be obtained from a representation defined over K_f via extension of scalars to the completions of K_f at its finite places. The claim that this representation is defined over K_f is somewhat subtle and relies the fact that the representation is odd. The existence of this representation over K_f will, however, not be used in this thesis.

References for this section include Deligne [20], Deligne and Serre [24], Serre [99], Gross [41], Edixhoven [31], and Couveignes, Edixhoven et al. [17].

3.1. Modular Galois representations over fields of characteristic 0

In [98], Serre conjectured that for every cusp form f that is an eigenform of the Hecke operators, there should be an associated family of l-adic representations with certain properties that we will give below. For cusp forms of weight 2, the existence of such representations follows from work of Eichler [33], Shimura [102] and Igusa [47]. Using the étale cohomology of powers of the universal elliptic curve over a certain modular curve, Deligne [20] generalised their construction to cusp forms of weight at least 2. In [20], the construction is only described in the case of cusp forms for $\mathrm{SL}_2(\mathbf{Z})$, but Deligne certainly knew how to generalise this to cusp forms for congruence subgroups. Conrad's book [15] contains a complete construction of the representations attached to cuspidal eigenforms of weight at least 2. Finally, a construction for cusp forms of weight 1 was given by Deligne and Serre [24]. Their construction actually uses the existence of l-adic representations associated to cuspidal eigenforms of weight ≥ 2 in order to associate to any cuspidal eigenform of weight 1 a family of representations over various finite fields; these are then shown to be the reductions of a two-dimensional representation over the field K_f having the desired properties.

With all of the above results put together, the precise statement on l-adic Galois representations associated to modular forms is as follows.

Theorem 3.1. *Let n and k be positive integers, and let f be a modular form of weight k for $\Gamma_1(n)$ over a field of characteristic 0. Assume that f is a (non-zero) eigenvector of the Hecke operators T_p (p prime) and $\langle d\rangle$ ($d \in (\mathbf{Z}/n\mathbf{Z})^\times$) for $\Gamma_1(n)$, with corresponding eigenvalues a_p (p prime) and $\epsilon(d)$ ($d \in (\mathbf{Z}/n\mathbf{Z})^\times$). Let K_f be the number field generated by these eigenvalues. Let l be a prime number, let λ be a prime of K_f over l, and let $K_{f,\lambda}$ denote the completion of K_f at λ. There exists a two-dimensional representation*

$$\rho_{f,\lambda}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \mathrm{Aut}_{K_{f,\lambda}} V_{f,\lambda}$$

that is unramified outside nl and such that for every prime number $p \nmid nl$ the characteristic polynomial of a Frobenius element at p (defined up to conjugation in $\operatorname{Aut}_{K_{f,\lambda}} V_{f,\lambda}$) equals $X^2 - a_pX + \epsilon(p \bmod n)p^{k-1}$.

It follows from the properties characterising such a representation $\rho_{f,\lambda}$ that if it exists, it must be *odd*, i.e. the determinant of a complex conjugation c equals -1; equivalently, the characteristic polynomial of c equals $X^2 - 1$. Also, $\rho_{f,\lambda}$ is unique up to Jordan–Hölder equivalence, i.e., after semi-simplification it is unique up to (non-unique) isomorphism. Moreover, it is known that $\rho_{f,\lambda}$ is irreducible if and only if f is a cusp form; see Ribet [86, Theorem 2.3].

3.2. Modular Galois representations over finite fields

Let f be a Hecke eigenform of weight k for $\Gamma_1(n)$, and consider the associated family of l-adic representations as in Theorem 3.1. The following basic result allows us to reduce these representations to representations over the residue fields of the number field K_f.

Lemma 3.2. *Let A be a discrete valuation ring with maximal ideal $\mathfrak{m}$ and field of fractions K. Let G be a group, and let V be a finite-dimensional representation of G over K. The following are equivalent:*

(1) there exists a G-stable lattice in V;

(2) there exists a basis of V with respect to which G acts via matrices with coefficients in A;

(3) the image of G in $\operatorname{Aut}_K V$ is bounded, i.e. for some (hence any) basis of V, the matrices giving the action of G have coefficients in $\mathfrak{m}^{-N}$ for some sufficiently large integer N.

Proof. The implications (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) are clear. Now assume (3), and choose any lattice L in V. By assumption, there exists an integer $N \geq 0$ such that $gL \subseteq \mathfrak{m}^{-N}L$ for all $g \in G$. Therefore we have inclusions of A-modules

$$L \subseteq \sum_{g \in G} gL \subseteq \mathfrak{m}^{-N}L,$$

so $\sum_{g\in G} gL$ is a lattice, and it is clearly G-stable. □

Let λ be a finite place of the number field K_f, let $K_{f,\lambda}$ be the completion of K_f with respect to λ, and let $\mathbf{F}$ be the residue field. Since $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is compact, applying Lemma 3.2 to the representation

$$\rho_{f,\lambda}\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \longrightarrow \operatorname{Aut}_{K_{f,\lambda}} V_{f,\lambda}$$

given by Theorem 3.1 shows that $\rho_{f,\lambda}$ can be obtained by base change from a representation over the valuation ring of $K_{f,\lambda}$. Reducing modulo the maximal ideal of this valuation ring, we obtain a two-dimensional representation of $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\mathbf{F}$, unramified outside nl and such that the characteristic polynomial of a Frobenius element at a prime number $p \nmid nl$ is the reduction modulo λ of $X^2 - a_pX + \epsilon(p \bmod n)p^{k-1}$.

By the Brauer–Nesbitt theorem, such a representation is unique up to Jordan–Hölder equivalence. Replacing the reduced representation by its semi-simplification, we therefore get a unique *semi-simple* representation with these properties. We note that although l-adic representations associated to cusp forms are irreducible, the corresponding representations over finite fields may be reducible. We will take a more precise look at this below.

In view of the above construction, one may ask whether a representation as above can be constructed starting from any eigenform over a finite field, not necessarily one obtained by reducing an eigenform in characteristic 0 modulo a prime. In fact, this can be done by lifting eigenforms over finite fields to forms in characteristic 0 that are eigenforms "modulo the given prime". In this way one can prove the following analogue of Theorem 3.1; see Deligne and Serre [24, théorème 6.7] or Gross [41, Proposition 11.1].

Theorem 3.3. *Let n and k be positive integers, let l be a prime number, and let f be a modular form of weight k for $\Gamma_1(n)$ over a finite field $\mathbf{F}$ of characteristic l. Assume f is an eigenvector of the Hecke operators T_p (p prime) and $\langle d\rangle$ ($d \in (\mathbf{Z}/n\mathbf{Z})^\times$), with corresponding eigenvalues a_p (p prime) and $\epsilon(d)$ ($d \in (\mathbf{Z}/n\mathbf{Z})^\times$). There exists a unique semi-simple two-dimensional representation*

$$\rho_f\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}}\, V$$

that is unramified outside nl and with the property that for every prime number p not dividing nl, the characteristic polynomial of a Frobenius element at p equals $X^2 - a_pX + \epsilon(p \bmod n)(p \bmod l)^{k-1}$. Moreover, this ρ_f is odd.

The above theorem associates to a Hecke eigenform f over a finite field $\mathbf{F}$ an isomorphism class of two-dimensional, odd, semi-simple representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\mathbf{F}$. Any representation in this isomorphism class is said to *arise from* f, and any representation arising from an eigenform is said to be *modular*.

3.3. Distinguishing between modular Galois representations

Let n be a positive integer, let l be a prime number, and let $\mathbf{F}$ be a finite field of characteristic l. It is a natural question to ask when two eigenforms for $\Gamma_1(n)$ over $\mathbf{F}$ give rise to isomorphic Galois representations. We will use this later to decide whether the Galois representation associated to a given modular form is reducible.

Let us first mention that if $n = l^a m$ with $a \geq 0$ and m not divisible by l, then the representation attached to an eigenform for $\Gamma_1(n)$ over $\mathbf{F}$ also arises from an eigenform (of possibly different weight) for $\Gamma_1(m)$; see Serre [99, page 195, remarque], Ribet [89, § 2], Buzzard [12], Wiese [114], and Khare and Wintenberger [54, Theorem 1.2(2)].

From now on we assume that $l \nmid n$. Below we will give a criterion that allows us to decide whether two eigenforms for $\Gamma_1(n)$ over $\mathbf{F}$ give rise to isomorphic Galois representations. In the case $n = 1$, this criterion is proved (in a slightly different form) in [17, Proposition 2.5.16]. Before giving the criterion, we state some results on modular forms in characteristic l.

There exists a unique modular form

$$A_l \in \mathrm{M}_{l-1}(\Gamma_1(1), \mathbf{F}_l)$$

that has q-expansion 1 at all cusps; it is called the *Hasse invariant* in characteristic l. If f is an eigenform of some weight $k \ge 2$ for some $\Gamma_1(n)$ over a field of characteristic l, then the product $A_l f$ is an eigenform of weight $k+l-1$ for $\Gamma_1(n)$.

There exists a derivation θ_l on modular forms over fields of characteristic l; see Katz [52]. It increases weights by $l+1$ and acts on q-expansions at all cusps as $q\,d/dq$. This implies that if f is an eigenform whose q-expansion at some cusp is non-constant, $\theta_l f$ is an eigenform with $T_l f = 0$, and the Galois representations are related by

$$\rho_{\theta_l f} \cong \chi_l \otimes_{\mathbf{F}_l} \rho_f,$$

where χ_l is the l-cyclotomic character.

Let f be a form whose q-expansion at some cusp is constant. Using the derivation θ_l, one can show that the weight of f is a multiple of $l-1$; see Katz [52, § 1]. This implies that f is a scalar multiple of a power of A_l, so the q-expansion of f at *every* cusp is constant. Therefore we can simply say that f has *constant q-expansion* without causing confusion. Furthermore, each Hecke operator T_p with p a prime number not dividing nl acts on f as multiplication by $(p+1)/p$. This implies that the Galois representation ρ_f is isomorphic to $1 \oplus \chi_l^{-1}$.

Let f be an eigenform of weight k for $\Gamma_1(n)$ over $\mathbf{F}$ with non-constant q-expansion, let p be a prime number, and let a_p be the eigenvalue of T_p on f. We define an eigenform

$$\eta_p f \in \begin{cases} \mathrm{M}_k(\Gamma_1(np), \mathbf{F}) & \text{if } p \mid n \\ \mathrm{M}_k(\Gamma_1(np^2), \mathbf{F}) & \text{if } p \nmid n \end{cases}$$

with the property that $T_p(\eta_p f) = 0$ by the formula

$$\eta_p f = \begin{cases} (b_1^{np,n})^* f - a_p (b_p^{np,n})^* f & \text{if } p \mid n; \\ (b_1^{np^2,n})^* f - a_p (b_p^{np^2,n})^* f + p^{k-1} (b_{p^2}^{np^2,n})^* f & \text{if } p \nmid n. \end{cases} \tag{3.1}$$

We now first give a special case of the criterion for distinguishing modular Galois representations. The general case is deduced from this in the theorem below.

Lemma 3.4. *Let $f, g \in \mathrm{S}_k(\Gamma_1(n), \mathbf{F})$ be eigenforms with non-constant q-expansions, with eigenvalues given by*

$$T_p f = a_p(f) f \quad \text{and} \quad T_p g = a_p(g) g \quad \text{for all prime numbers } p.$$

If $a_p(f) = a_p(g)$ for all prime numbers $p \le \frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma_1(n)]$, then ρ_f and ρ_g are isomorphic.

Proof. By the recurrence relations for the eigenvalues of Hecke operators, the condition implies that for some $\lambda \in \mathbf{F}^\times$, the q-expansions of f and λg are equal up to order $\frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma_1(n)]$. By Lemma 2.1, this implies $f = \lambda g$. In particular, ρ_f and ρ_g are isomorphic. $\square$

Theorem 3.5. *Let n be a positive integer, let l be a prime number not dividing n, and let $\mathbf{F}$ be a finite field of characteristic l. Let f be an eigenform of weight k_f for $\Gamma_1(n)$ over $\mathbf{F}$, and let g be an eigenform of weight k_g for $\Gamma_1(n)$ over $\mathbf{F}$, with eigenvalues given by*

$$T_p f = a_p(f) f \quad \text{and} \quad T_p g = a_p(g) g \quad \text{for all prime numbers } p.$$

We define

$$m = n \prod_{p \mid n \text{ prime}} p$$

and

$$N = [\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma_1(m)] \cdot \begin{cases} \max\{k_f, k_g, 11\} + 3 & \text{if } l = 2, \\ \max\{k_f, k_g, 10\} + 4 & \text{if } l = 3, \\ \max\{k_f, k_g\} + l^2 - 1 & \text{if } l \geq 5. \end{cases}$$

For any integer i with $0 \leq i \leq l-2$, the following are equivalent:

(1) the representations ρ_f and $\chi_l^i \rho_g$ are isomorphic;

(2) we have $k_f \equiv k_g + 2i \pmod{l-1}$, and $a_p(f) = p^i a_p(g)$ for all prime numbers p with $p \nmid nl$ and $p \leq N$;

(3) we have $k_f \equiv k_g + 2i \pmod{l-1}$, and one of the following three situations occurs:

(a) the q-expansions of f and g are constant (so $k_f \equiv k_g \equiv 0 \pmod{l-1}$), and $i = 0$;

(b) the q-expansion of f (but not that of g) is constant (so $k_f \equiv 0 \pmod{l-1}$), and $a_p(g) = p^{-i}(1 + p^{-1})$ for all prime numbers p with $p \nmid nl$ and $p \leq N$;

(b′) the q-expansion of g (but not that of f) is constant (so $k_g \equiv 0 \pmod{l-1}$), and $a_p(f) = p^i(1 + p^{-1})$ for all prime numbers p with $p \nmid nl$ and $p \leq N$;

(c) the q-expansions of f and g are non-constant, and $a_p(f) = p^i a_p(g)$ for all prime numbers p with $p \nmid nl$ and $p \leq N$.

Proof. The implication (1) ⇒ (2) follows from the properties characterising ρ_f and ρ_g. The implication (2) ⇒ (3) follows from the standard formula for the action of Hecke operators on q-expansions, together with the fact that the only forms f with constant q-expansion are multiples of powers of A_l. To prove the implication (3) ⇒ (1), we treat the three cases (a), (b), (c) separately. The proof in case (b′) is almost identical to that in case (b) and is therefore omitted.

Case (a). This is trivial since both ρ_f and ρ_g are isomorphic to $1 \oplus \chi_l^{-1}$.

Case (b). We have $\rho_f \cong 1 \oplus \chi_l^{-1}$, and we have to show that $\rho_g \cong \chi_l^{-i}(1 \oplus \chi_l^{-1})$. We distinguish the cases $l = 2$, $l = 3$ and $l \geq 5$.

If $l = 2$, the assumption means that $a_p(g) = 0$ for all prime numbers p with $p \nmid 2n$ and $p \leq N$, and we have to show that ρ_g is the trivial representation. We define an integer k and an eigenform $h_1 \in \mathrm{S}_k(\Gamma_1(1), \mathbf{F}_2)$ by

$$\begin{cases} k = 12 \text{ and } h_1 = \Delta \bmod 2 & \text{if } k_g \equiv 0 \pmod 3, \\ k = 13 \text{ and } h_1 = A_2(\Delta \bmod 2) & \text{if } k_g \equiv 1 \pmod 3, \\ k = 14 \text{ and } h_1 = A_2^2(\Delta \bmod 2) & \text{if } k_g \equiv 2 \pmod 3, \end{cases}$$

where Δ is the discriminant modular form. In particular, we have $k \equiv k_g \pmod 3$, and ρ_{h_1} is trivial. Applying the operators η_p defined in (3.1) to h_1 for $p \mid n$, we construct an eigenform

$$h_m \in \mathrm{S}_k(\Gamma_1(m), \mathbf{F}_2)$$

such that ρ_{h_m} is trivial and $T_p h_m = 0$ for every p. Similarly, applying the η_p with $p \mid n$ to g, we construct an eigenform

$$g_m \in \mathrm{S}_{k_g}(\Gamma_1(m), \mathbf{F})$$

such that $\rho_{g_m} \cong \rho_g$ and $T_p g_m = 0$ for all prime numbers p with $p \neq 2$ and $p \leq N$. If $k_g < k$, we define

$$g'_m = \theta_2^{(k-k_g)/3} g_m \quad \text{and} \quad h'_m = h_m \quad \text{in} \quad \mathrm{S}_k(\Gamma_1(m), \mathbf{F})$$

Similarly, if $k_g \geq k$, we define

$$g'_m = \theta_2 g_m \quad \text{and} \quad h'_m = \theta_2^{(k_g-k)/3+1} h_m \quad \text{in} \quad \mathrm{S}_{k_g+3}(\Gamma_1(m), \mathbf{F}).$$

In either case, g'_m and h'_m are eigenforms, $\rho_{g'_m} \cong \rho_g$ and $\rho_{h'_m}$ is trivial. Furthermore, the q-expansions of g'_m and h'_m are non-constant, and we have

$$a_p(g'_m) = 0 = a_p(h'_m) \quad \text{for all primes } p \leq N.$$

Lemma 3.4 now implies that ρ_g is trivial, which is what we had to prove.

If $l = 3$, then k_g is even, the assumption means that $a_p(g) = 1 + p$ for all prime numbers p with $p \nmid 3n$ and $p \leq N$, and we have to show that $\rho_g \cong 1 \oplus \chi_3$. We define an integer k and an eigenform $h_1 \in \mathrm{S}_k(\Gamma_1(1), \mathbf{F}_3)$ by

$$\begin{cases} k = 12 \text{ and } h_1 = \Delta \bmod 3 & \text{if } k_g \equiv 0 \pmod 4, \\ k = 14 \text{ and } h_1 = A_3(\Delta \bmod 3) & \text{if } k_g \equiv 2 \pmod 4, \end{cases}$$

where Δ is the discriminant modular form. In particular, we have $k \equiv k_g \pmod 4$, and $\rho_{h_1} \cong 1 \oplus \chi_3$. Applying the operators η_p defined in (3.1) to h_1 for $p \mid n$, we construct an eigenform

$$h_m \in \mathrm{S}_k(\Gamma_1(m), \mathbf{F}_3)$$

with $\rho_{h_m} \cong 1 \oplus \chi_3$, $T_3 h_m = 0$ and $T_p h_m = (1+p) h_m$ for all $p \neq 3$. Similarly, applying the η_p with $p \mid n$ to f_2, we construct an eigenform

$$g_m \in \mathrm{S}_{k_g}(\Gamma_1(m), \mathbf{F})$$

such that $\rho_{g_m} \cong \rho_{f_2}$ and $T_p g_m = 0$ for all prime numbers p with $p \mid m$. If $k_g < k$, we define

$$g'_m = \theta_3^{(k-k_g)/4} g_m \quad \text{and} \quad h'_m = h_m \quad \text{in} \quad \mathrm{S}_k(\Gamma_1(m), \mathbf{F}).$$

Similarly, if $k_g \geq k$, we define

$$g'_m = \theta_3 g_m \quad \text{and} \quad h'_m = \theta_3^{(k_g-k)/4+1} h_m \quad \text{in} \quad \mathrm{S}_{k_g+4}(\Gamma_1(m), \mathbf{F}).$$

In either case, g_m and h'_m are eigenforms, $\rho_{g'_m}$ is isomorphic to either ρ_g or $\chi_3\rho_g$, and $\rho_{h'_m} \cong 1 \oplus \chi_3$. Furthermore, the q-expansions of g'_m and h'_m are non-constant, and for all prime numbers $p \leq N$ we have

$$a_p(g'_m) = a_p(h'_m) = \begin{cases} 0 & \text{if } p \mid 3n; \\ 1+p & \text{if } p \nmid 3n. \end{cases}$$

Lemma 3.4 now implies that $\rho_{g'_m}$ is isomorphic to $1 \oplus \chi'_3$. Therefore the same holds for ρ_g, which is what we had to prove.

If $l \geq 5$, then again k_g is even. We define an eigenform $h_1 \in \mathrm{M}_{l+1}(\Gamma_1(1), \mathbf{F}_l)$ by

$$h_1 = E_{l+1} \bmod l,$$

where E_{l+1} is the Eisenstein series of weight $l+1$. Since $l-1$ does not divide $l+1$, the q-expansion of E_{l+1} is non-constant. We have

$$\rho_{h_1} = 1 \oplus \chi_l.$$

Applying the operators η_p to h_1 for $p \mid n$, we construct an eigenform

$$h_m \in \mathrm{M}_{l+1}(\Gamma_1(m), \mathbf{F})$$

such that $\rho_{h_m} \cong 1 \oplus \chi_l$ and $T_p h_m = 0$ for all prime numbers $p \mid m$. Similarly, we construct an eigenform

$$g_m \in \mathrm{M}_{k_g}(\Gamma_1(m), \mathbf{F})$$

such that $\rho_{g_m} \cong \rho_g$ and $T_p g_m = 0$ for all prime numbers p with $p \mid m$. We define

$$g'_m = A_l^{\max\{-j,0\}} \theta_l g_m \in \mathrm{S}_{k'}(\Gamma_1(m), \mathbf{F})$$

and

$$h'_m = A_l^{\max\{j,0\}} \theta_l^{l-1-i} h_m \in \mathrm{S}_{k'}(\Gamma_1(m), \mathbf{F}_l).$$

where

$$j = \frac{(k_g + l + 1) - (l-i)(l+1)}{l-1} \in \mathbf{Z} \quad \text{and} \quad k' = \max\{k_g + l + 1, (l-i)(l+1)\}.$$

Then g'_m and h'_m are eigenforms with

$$\rho_{g'_m} \cong \chi_l \rho_g \quad \text{and} \quad \rho_{h'_m} \cong \chi_l^{-i}(1 \oplus \chi_l).$$

Furthermore, the q-expansions of g'_m and h'_m are non-constant and agree up to order N. Lemma 3.4 now implies that $\rho_{g'_m} \cong \rho_{h'_m}$, and we conclude that $\rho_g \cong \chi_l^{-i}(1 \oplus \chi_l^{-1})$, which is what we had to prove.

Case (c). Applying the operators η_p for $p \mid n$ prime to f and g, we construct eigenforms

$$f_m \in \mathrm{S}_{k_f}(\Gamma_1(m), \mathbf{F}) \quad \text{and} \quad g_m \in \mathrm{S}_{k_g}(\Gamma_1(m), \mathbf{F})$$

with $\rho_{f_m} \cong \rho_f$, $\rho_{g_m} \cong \rho_g$, and $T_p f_m = 0$ and $T_p g_m = 0$ for all prime numbers $p \mid n$. We define

$$f'_m = A_l^{\max\{-j,0\}} \theta_l f_m \quad \text{and} \quad g'_m = A_l^{\max\{j,0\}} \theta_l^{i+1} g_m \quad \text{in} \quad \mathrm{S}_{k'}(\Gamma_1(m), \mathbf{F}),$$

where

$$j = \frac{(k_f + l + 1) - (k_g + (i+1)(l+1))}{l-1} \in \mathbf{Z}$$

and

$$k' = \max\{k_f + l + 1, k_g + (i+1)(l+1)\}.$$

Then f'_m and g'_m are eigenforms with

$$\rho_{f'_m} \cong \chi_l \rho_f \quad \text{and} \quad \rho_{g'_m} \cong \chi_l^{i+1} \rho_g.$$

Furthermore, the q-expansions of f and g are non-constant, and we have

$$a_p(f'_m) = a_p(g'_m) \quad \text{for all prime numbers } p \leq N.$$

Lemma 3.4 now implies that $\rho_{f'_m} \cong \rho_{g'_m}$. From this it follows that $\rho_f \cong \chi_l^i \rho_g$, which is what we had to prove. □

It is known that if f is an eigenform of weight k for $\Gamma_1(n)$ over $\mathbf{F}$, then there exist integers i and $\tilde{k}$ with

$$0 \leq i \leq l-2, \quad 1 \leq \tilde{k} \leq l+1 \quad \text{and} \quad \tilde{k} \equiv k + 2i \pmod{l-1}$$

and an eigenform $\tilde{f}$ of weight $\tilde{k}$ such that if the eigenvalues of the Hecke operators on f are given by

$$T_p f = a_p f \text{ for } p \text{ prime} \quad \text{and} \quad \langle d \rangle f = \epsilon(d) f \text{ for } d \in (\mathbf{Z}/n\mathbf{Z})^\times,$$

then the eigenvalues on $\tilde{f}$ are given by

$$T_p \tilde{f} = (p \bmod l)^i a_p \tilde{f} \text{ for } p \neq l \text{ prime} \quad \text{and} \quad \langle d \rangle \tilde{f} = \epsilon(d) \tilde{f} \text{ for } d \in (\mathbf{Z}/n\mathbf{Z})^\times.$$

For a proof of the existence of such an $\tilde{f}$, we refer to Edixhoven [31, Theorem 3.4]. The Galois representation $\rho_{\tilde{f}}$ associated to such an $\tilde{f}$ is isomorphic to $\chi_l^i \otimes_{\mathbf{F}_l} \rho_f$.

3.4. Reducible representations

Let n and k be positive integers, and let l be a prime number not dividing n.

Let f be an eigenform of weight k for $\Gamma_1(n)$ over $\overline{\mathbf{F}}_l$, and let $\epsilon\colon (\mathbf{Z}/n\mathbf{Z})^\times \to \overline{\mathbf{F}}_l^\times$ be the character such that $\langle d\rangle f = \epsilon(d)f$ for all $d \in (\mathbf{Z}/n\mathbf{Z})^\times$. If ρ_f is reducible, we can write

$$\rho_f = \epsilon_1\chi_l^i \oplus \epsilon_2\chi_l^j$$

with characters ϵ_1 and ϵ_2 whose conductors are coprime to l, and with $i, j \in \mathbf{Z}/(l-1)\mathbf{Z}$; here χ_l is the l-cyclotomic character. The defining properties of ρ_f imply that

$$\epsilon_1\epsilon_2 = \epsilon \quad \text{and} \quad i + j = k - 1 \in \mathbf{Z}/(l-1)\mathbf{Z}.$$

Furthermore, it follows from work of Carayol that the product of the conductors of ϵ_1 and ϵ_2 divides n; see Livné [71, Proposition 0.1].

Conversely, it is well known that given characters

$$\epsilon_1\colon (\mathbf{Z}/n_1\mathbf{Z})^\times \to \overline{\mathbf{F}}_l^\times, \quad \epsilon_2\colon (\mathbf{Z}/n_2\mathbf{Z})^\times \to \overline{\mathbf{F}}_l^\times$$

such that

$$n_1n_2 \mid n \quad \text{and} \quad \epsilon_1(-1)\epsilon_2(-1) = (-1)^k,$$

there exists a Hecke eigenform f of some weight k' for $\Gamma_1(n)$ over $\overline{\mathbf{F}}_l$ such that the Galois representation ρ_f is isomorphic to $\epsilon_1 \oplus \epsilon_2\chi_l^{k-1}$. In fact, we can take the unique integer k' satisfying

$$3 \le k' \le l + 1 \quad \text{and} \quad k' \equiv k \pmod{l-1},$$

and take f to be the reduction of a suitable multiple of the Eisenstein series $E_{k'}^{\epsilon_1,\epsilon_2}$, which will be defined in § II.2.3.

3.5. Serre's conjecture

In 1973, Serre made the conjecture that all two-dimensional, odd, irreducible representations

$$\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}} V$$

are modular. He published his conjecture in 1987, stating it in a sharper form [99] that predicted the minimal level and weight of the modular form that should give rise to ρ. To this end Serre associated to such a ρ two invariants, called its *level* $n(\rho)$ and its *weight* $k(\rho)$. The level is defined in terms of the local behaviour of ρ at the primes different from the characteristic of $\mathbf{F}$, whereas the weight is defined using the local behaviour at the characteristic of $\mathbf{F}$. For a general introduction to Serre's conjecture, we refer to Edixhoven [32] or Ribet and Stein [90].

It is known, by the work of many people, that the weak form of Serre's conjecture implies the strong form, i.e. that if a representation ρ as above is modular, it arises from a modular form of level $n(\rho)$ and weight $k(\rho)$. In fact, a modular representation ρ is even known to arise from a form of *minimal weight* (which equals $k(\rho)$ in most cases, but is sometimes smaller; see Edixhoven [31, Definition 4.3]), except possibly if $\mathbf{F}$ is of characteristic 2 and the restriction of ρ to a decomposition group at 2 is an extension of an unramified character by itself. For the proof of this result, we refer to Kisin's overview article [60, Theorem 1.1.4].

Starting with the proof of the modularity theorem for elliptic curves (previously the Shimura–Taniyama conjecture) by Wiles [115] and Taylor and Wiles [107], important developments in the theory of deformations of Galois representations have been made by many authors. These developments, including a crucial result of Kisin [61], have finally allowed Khare and Wintenberger [54], [55] to prove Serre's conjecture.

3.6. Galois representations on torsion subgroups of Jacobians of modular curves

Let $n \geq 1$ be an integer, let l be a prime number not dividing n, and let k be an integer such that

$$2 \leq k \leq l+1.$$

We write

$$n' = \begin{cases} n & \text{if } k = 2; \\ nl & \text{if } k > 2. \end{cases}$$

We saw in § 2.3 that there exists a canonical surjective ring homomorphism

$$\mathbf{T}_1(n') \longrightarrow \mathbf{T}(\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l)),$$

where $\mathbf{T}_1(n')$ is the subring of $\operatorname{End} \mathrm{J}_1(n')_{\mathbf{Z}[1/n']}$ generated by the Hecke operators.

Let us now consider a cusp form

$$f \in \mathrm{S}_k(\Gamma_1(n), \overline{\mathbf{F}}_l)$$

that is an eigenvector for all the Hecke operators. Let $\mathbf{F}_f$ denote the finite extension of $\mathbf{F}_l$ generated by the corresponding eigenvalues. We define a surjective ring homomorphism

$$e_f \colon \mathbf{T}_1(n') \to \mathbf{F}_f$$

as the composed map

$$\mathbf{T}_1(n') \longrightarrow \mathbf{T}(\mathrm{S}_k(\Gamma_1(n), \mathbf{F}_l)) \xrightarrow{\mathrm{ev}_f} \mathbf{F}_f,$$

where ev_f denotes the ring homomorphism from § 2.4 that sends each Hecke operator to its eigenvalue of f. We define a maximal ideal $\mathfrak{m}_f$ of $\mathbf{T}_1(n')$ by

$$\mathfrak{m}_f = \ker e_f.$$

Note that giving a form f as above up to scalar multiplication is equivalent to specifying the ring homomorphism e_f, and that giving the homomorphism e_f up to Galois conjugacy comes down to specifying the maximal ideal $\mathfrak{m}_f$.

Let $\mathrm{J}_1(n')_{\mathbf{Z}[1/nl]}[\mathfrak{m}_f]$ be the largest closed subscheme of $\mathrm{J}_1(n')_{\mathbf{Z}[1/nl]}$ annihilated by $\mathfrak{m}_f$. Since e_f induces an isomorphism $\mathbf{T}_1(n')/\mathfrak{m}_f \xrightarrow{\sim} \mathbf{F}_f$, the action of $\mathbf{T}_1(n')$ makes $\mathrm{J}_1(n')_{\mathbf{Z}[1/nl]}[\mathfrak{m}_f]$ into a finite-dimensional $\mathbf{F}_f$-vector space scheme over $\operatorname{Spec} \mathbf{Z}[1/nl]$. By Lemma 1.2, this vector space scheme is non-zero and finite étale.

It follows from the Eichler–Shimura congruence relation (see § 1.4), the Čebotarev density theorem and the Brauer–Nesbitt theorem that if ρ_f is irreducible, then the semi-simplification of the $\mathbf{F}_f[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is a direct sum of copies of the Galois representation ρ_f attached to f; see Mazur [76, Chapter II, Proposition 14.2]. In the case that ρ_f is *absolutely* irreducible, we can invoke the theorem of Boston, Lenstra and Ribet [10]. This gives a stronger result than the Brauer–Nesbitt theorem, namely that $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is a direct sum of copies of ρ_f, i.e. it is already semi-simple. If $l > 2$, then ρ_f is absolutely irreducible as soon as it is irreducible.

Remark. Another place where modular Galois representations over finite fields occur is in étale cohomology of modular curves. We refer to Wiese [113] for details.

3.7. Simplicity

We complement the results of § 3.6 with a result on *simplicity* (or *multiplicity one*, as it is usually called) of the $\mathbf{F}_f[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ in the case that the $\mathbf{F}_f$-linear representation

$$\rho_f\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \operatorname{Aut}_{\mathbf{F}_f} W_f$$

associated to f is absolutely irreducible. As we have just seen, in this situation the $\mathbf{F}_f[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is a direct sum of copies of W_f.

In many cases, $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is in fact simple as a $\mathbf{F}_f[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module, which is to say that it is isomorphic to W_f. It is known precisely under which conditions this occurs. Sufficient conditions for this simplicity phenomenon follow from work of Mazur [76], Mazur and Ribet [77, Theorem 1], Gross [41, Theorem 12.10(1)], Edixhoven [31, Theorem 9.2] and Buzzard [90, Appendix]. Wiese proved in [112] that these conditions are also necessary, under an extra assumption in the case $l = 2$. (This assumption is automatically fulfilled if the strong form of Serre's conjecture, as described in § 3.5, is true.)

Theorem 3.6. *Suppose that $2 \le k \le l+1$ and that ρ_f is absolutely irreducible.*

(1) If ρ_f is ramified at l, or if ρ_f is unramified at l and a Frobenius element at l does not act as a scalar, then the $\mathbf{F}_f[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is isomorphic to W_f.

(2) If ρ_f is unramified at l, a Frobenius element at l does act as a scalar, and ρ_f arises from a form of weight one (this last condition is implied by the preceding ones if $l > 2$), then the semi-simplification of $\mathrm{J}_1(n')[\mathfrak{m}_f](\overline{\mathbf{Q}})$ is a direct sum of at least two copies of W_f.

Chapter II

Analytic results on modular curves

This chapter is devoted to certain analytic results that will be used in the next chapters. Most importantly, we give explicit bounds on Petersson norms and supremum norms of cusp forms, and on Green functions of Fuchsian groups.

1. Fuchsian groups

In this section we collect some results that will be useful later on when we bound analytic quantities related to modular curves. We will begin by summarising some basic facts about the hyperbolic plane and about Fuchsian groups. After that we describe some material concerning harmonic analysis on the quotient of the hyperbolic plane by a Fuchsian group.

The author has found Iwaniec's book [49] to be a very valuable reference. Other references are Beardon [5] and Terras [108]. Selberg's foundational article [94], in which he develops this material (and much more) in a general context, is also highly recommended. Hejhal's two volumes [45] and [46] contain in-depth proofs of Selberg's results, as well as a lot of useful background material.

1.1. Hyperbolic geometry

The hyperbolic plane $\mathbf{H}$ is the unique two-dimensional, complete, connected and simply connected Riemannian manifold with constant Gaussian curvature -1. We will always identify $\mathbf{H}$ with the complex upper half-plane; this gives $\mathbf{H}$ the structure of a (non-compact) Riemann surface. The Riemannian metric is given in terms of the standard coordinate $z = x + iy$ by

$$\frac{dz\,d\bar{z}}{(\Im z)^2} = \frac{dx^2 + dy^2}{y^2},$$

and the associated volume form is

$$\mu_{\mathbf{H}} = \frac{i\,dz \wedge d\bar{z}}{2(\Im z)^2} = \frac{dx \wedge dy}{y^2}.$$

In the hyperbolic plane, the circumference of a circle of radius r equals $2\pi\sinh(r)$, and the area of a disc of radius r equals $2\pi(\cosh(r)-1)$.

The group $\mathrm{SL}_2(\mathbf{R})$ acts on $\mathbf{H}$ by isometries. Under the identification of $\mathbf{H}$ with the complex upper half-plane, this action on $\mathbf{H}$ is the restriction of the action on $\mathbf{P}^1(\mathbf{C})$ by Möbius transformations. The elements of $\mathrm{SL}_2(\mathbf{R})$ can be classified according to their fixed points in $\mathbf{P}^1(\mathbf{C})$. For any element $\gamma \in \mathrm{SL}_2(\mathbf{R})$ that is not the identity, there are three possibilities, depending on the trace of γ:

(E) $|\mathrm{tr}\,\gamma| < 2$: two conjugate fixed points in $\mathbf{P}^1(\mathbf{C}) \setminus \mathbf{P}^1(\mathbf{R})$;

(P) $|\mathrm{tr}\,\gamma| = 2$: a unique fixed point in $\mathbf{P}^1(\mathbf{R})$;

(H) $|\mathrm{tr}\,\gamma| > 2$: two distinct fixed points in $\mathbf{P}^1(\mathbf{R})$.

The element γ is called *elliptic*, *parabolic* or *hyperbolic*, acccording to its place in this classification. This terminology also applies to conjugacy classes.

Instead of the usual geodesic distance $r(z,w)$ between two points of $\mathbf{H}$, the function

$$u(z,w) = \cosh r(z,w)$$

turns out to be a more convenient measure of distance for computations. Clearly, any function on $\mathbf{H}\times\mathbf{H}$ depending only on the hyperbolic distance between its arguments can be expressed as a function of u. For z and w in the upper half-plane, $u(z,w)$ can be expressed as

$$u(z,w) = 1 + \frac{|z-w|^2}{2(\Im z)(\Im w)}.$$

This is easily checked for z and w on the imaginary axis. For arbitrary z and w, the identity follows from this case after translation by a suitable element $\gamma \in \mathrm{SL}_2(\mathbf{R})$, using the fact that both sides are invariant under replacing (z,w) by $(\gamma z, \gamma w)$.

We denote by

$$\Delta = -y^2(\partial_x^2 + \partial_y^2)$$

the Laplace–Beltrami operator on $\mathbf{H}$. It turns out to be useful to write the eigenvalues of Δ as $\frac{1}{4}+t^2$ with t a complex number (defined up to sign).

An *invariant integral operator* on $\mathbf{H}$ is an integral operator of the form

$$L_k\colon f \mapsto \left(z \mapsto \int_{w\in\mathbf{H}} k(u(z,w))f(w)\mu_{\mathbf{H}}(w)\right)$$

for some function $k\colon (1,\infty) \to \mathbf{R}$. The domain of definition depends on the function k; for example, if k is smooth, then we can take f to range over the smooth functions with compact support.

The Laplace operator Δ commutes with all invariant integral operators; see Selberg [94, pages 51–52] or Iwaniec [49, Theorem 1.9]. In fact, every eigenfunction of Δ is also an eigenfunction of all invariant integral operators, and conversely; see Selberg [94, page 55] or Iwaniec [49, Theorems 1.14 and 1.15].

Let $k\colon [1,\infty) \to \mathbf{R}$ be a smooth function with compact support, and let L_k be the invariant integral operator defined by k. The relation between the eigenvalues of Δ and those of L_k is given by the *Selberg–Harish-Chandra transform* of k. This is a function

$$h\colon \mathbf{R} \cup [-1/2, 1/2]i \to \mathbf{C}$$

defined by the following property. Let $f\colon \mathbf{H} \to \mathbf{C}$ be an eigenfunction of the Laplace operator with eigenvalue $\lambda = 1/4 + t^2$. Then f is also an eigenfunction of L_k, and the eigenvalue depends only on λ; we can therefore define $h(t)$ uniquely such that

$$\int_{w \in \mathbf{H}} k(u(z, w)) f(w) \mu_{\mathbf{H}}(w) = h(t) f(z). \tag{1.1}$$

In particular, taking $f = 1$ we see that

$$h(\pm i/2) = 2\pi \int_1^\infty k(u) du. \tag{1.2}$$

The Selberg–Harish-Chandra transform can be defined for general symmetric spaces; see Selberg [94, page 55]. In the case of $\mathbf{H}$ it can be identified with the classical *Mehler–Fock transform*, which is defined as follows (see Iwaniec [49, equation 1.62′]):

$$h(t) = 2\pi \int_1^\infty k(u) P_{-1/2+it}(u) du. \tag{1.3}$$

Here P_ν is the Legendre function of the first kind of degree ν (see Iwaniec [49, equation 1.43] or any book on special functions, such as Erdélyi et al. [34, § 3.6.1]). The function k can be recovered from h by means of the *Mehler–Fock inversion formula* (see Iwaniec [49, equation 1.42], Erdélyi et al. [34, § 3.15.1, equations 8 and 9], or Mehler [78, page 192]):

$$k(u) = \frac{1}{2\pi} \int_0^\infty P_{-1/2+it}(u) h(t) \tanh(\pi t) t\, dt. \tag{1.4}$$

We call k the *inverse Selberg–Harish-Chandra transform* of h.

The identity (1.1) holds more generally than just for smooth functions k with compact support; see Selberg [94, pages 60–61]. It will be enough for us to state a slightly weaker, but more convenient sufficient condition (cf. Selberg [94, page 72] or Iwaniec [49, equation 1.63]). Let h be a function with the following properties:

(H1) For some $\alpha > 1/2$, the function h is even and holomorphic on $\{t \in \mathbf{C} \mid |\Im t| < \alpha\}$.

(H2) For some $\beta > 2$, the function $t \mapsto |h(t)||t|^\beta$ is bounded in this strip.

Then the inverse Selberg–Harish-Chandra transform k of h (as defined by (1.4)) exists, and (1.1) is valid for the pair (h, k).

There is an alternative way to compute the Selberg–Harish-Chandra transform and its inverse which is sometimes useful. In fact, this is the formula originally given by Selberg [94, page 72]. Writing h as the Fourier transform of a function g, we can compute h in two steps using the following formula (cf. Iwaniec [49, equation 1.62]):

$$\begin{aligned} g(r) &= \sqrt{2} \int_{\cosh r}^\infty \frac{k(u) du}{\sqrt{u - \cosh r}}, \\ h(t) &= 2 \int_0^\infty \cos(rt) g(r) dr. \end{aligned}$$

The inverse can be computed as follows (cf. Iwaniec [49, equation 1.64]):

$$g(r) = \frac{1}{\pi}\int_0^\infty \cos(rt)h(t)dt,$$
$$k(u) = -\frac{1}{\pi\sqrt{2}}\int_{\operatorname{acosh} u}^\infty \frac{g'(r)dr}{\sqrt{\cosh r - u}}.$$

1.2. Fuchsian groups

A *Fuchsian group* is a discrete subgroup of $\mathrm{SL}_2(\mathbf{R})$. For any Fuchsian group Γ, the quotient space $\Gamma\backslash\mathbf{H}$ is a connected Hausdorff space and can be made into a Riemann surface in a natural way. However, the $\Gamma\backslash\mathbf{H}$ "inherits" the structure of Riemannian manifold from $\mathbf{H}$ only outside the set of fixed points of elliptic elements of Γ.

The hyperbolic metric on $\mathbf{H}$ induces a measure on $\Gamma\backslash\mathbf{H}$, given by a smooth volume form outside the elliptic points. If the volume of $\Gamma\backslash\mathbf{H}$ with respect to this measure is finite, we call Γ a *cofinite Fuchsian group*. In this case we define

$$\mathrm{vol}_\Gamma = \int_{\Gamma\backslash\mathbf{H}} \mu_{\mathbf{H}}.$$

Let Γ be a cofinite Fuchsian group. The Riemann surface $\Gamma\backslash\mathbf{H}$ is in general not compact, but can always be compactified by adding a finite number of points, called *cusps*. These correspond to the conjugacy classes of non-trivial *maximal parabolic subgroups* in Γ, i.e. non-trivial subgroups that are maximal among the subgroups containing only parabolic elements. Every such subgroup has a unique fixed point under the natural action of Γ on $\mathbf{P}^1(\mathbf{R})$. For every conjugacy class $\mathfrak{c}$ we choose one representative, which we denote by $\Gamma_\mathfrak{c}$. We fix an element $\sigma_\mathfrak{c} \in \mathrm{SL}_2(\mathbf{R})$ such that $\sigma_\mathfrak{c}\infty \in \mathbf{P}^1(\mathbf{R})$ is the unique fixed point of $\Gamma_\mathfrak{c}$ and such that

$$\sigma_\mathfrak{c}^{-1}\Gamma_\mathfrak{c}\sigma_\mathfrak{c} = (\Gamma\cap\{\pm 1\})\bigl\{\bigl(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\bigr) \bigm| a \in \mathbf{Z}\bigr\}.$$

Such a $\sigma_\mathfrak{c}$ exists and is unique up to multiplication from the right by a matrix of the form $\pm\bigl(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\bigr)$ with $b \in \mathbf{R}$; see Iwaniec [49, § 2.2].

Let $\mathfrak{c}$ be a cusp of Γ. We define

$$\begin{aligned} q_\mathfrak{c}\colon \mathbf{H} &\to \mathbf{C}\\ z &\mapsto \exp(2\pi i\sigma_\mathfrak{c}^{-1}z)\end{aligned}$$

and

$$\begin{aligned} y_\mathfrak{c}\colon \mathbf{H} &\to (0,\infty)\\ z &\mapsto \Im\sigma_\mathfrak{c}^{-1}z = -\frac{\log|q_\mathfrak{c}(z)|}{2\pi}.\end{aligned}$$

For ϵ a positive real number, we let $B_\mathfrak{c}(\epsilon)$ denote the open subset of $\Gamma\backslash\mathbf{H}$ that is the image of the strip

$$\{x + iy \mid 0 \le x < 1 \text{ and } y > 1/\epsilon\} \subset \mathbf{H}$$

under the quotient map

$$\begin{aligned} \mathbf{H} &\to \Gamma\backslash\mathbf{H}\\ z &\mapsto \Gamma\sigma_\mathfrak{c} z.\end{aligned}$$

For ϵ sufficiently small, $B_{\mathfrak{c}}(\epsilon)$ is an open disc of area ϵ around $\mathfrak{c}$, and the map $q_{\mathfrak{c}}$ induces a chart on $\Gamma\backslash\mathbf{H}$ with domain $B_{\mathfrak{c}}(\epsilon)$ and with image equal to the punctured disc $\{z \in \mathbf{C} \mid 0 < |z| < \exp(-2\pi/\epsilon)\}$. A compactification of $\Gamma\backslash\mathbf{H}$ can be obtained by adding a point for every cusp $\mathfrak{c}$, corresponding to the point $0 \in \mathbf{C}$ in the chart $q_{\mathfrak{c}}$, and defining the topology such that $q_{\mathfrak{c}}$ extends to a chart with image equal to the disc $\{z \in \mathbf{C} \mid |z| < \exp(-2\pi/\epsilon)\}$.

Next we look at the non-trivial *maximal elliptic subgroups* of Γ. These correspond bijectively to the points of $\mathbf{H}$ with non-trivial stabiliser in Γ, and the conjugacy classes correspond to the images of these points in $\Gamma\backslash\mathbf{H}$. By an *elliptic point* of Γ we mean a point of $\Gamma\backslash\mathbf{H}$ as above. For each elliptic point $\mathfrak{e}$, we choose a representative of the corresponding conjugacy class and denote it by $\Gamma_{\mathfrak{e}}$. We write

$$m_{\mathfrak{e}} = \frac{\#\Gamma_{\mathfrak{e}}}{\#(\Gamma \cap \{\pm 1\})}.$$

If w is the point of $\mathbf{H}$ stabilised by $\Gamma_{\mathfrak{e}}$, then for all $\epsilon > 0$ the open disc

$$\{z \in \mathbf{H} \mid 2\pi(u(z,w) - 1) < m_{\mathfrak{e}}\epsilon\}$$

of area $m_{\mathfrak{e}}\epsilon$ maps to a disc of area ϵ in $\Gamma\backslash\mathbf{H}$ if ϵ is sufficiently small. We denote this disc by $B_{\mathfrak{e}}(\epsilon)$. For $r > 0$ sufficiently small, the map

$$\begin{aligned} q_{\mathfrak{e}}\colon \left\{ w \in \mathbf{H} \;\middle|\; \left|\frac{w-z}{w-\bar{z}}\right| < r \right\} &\longrightarrow \mathbf{C} \\ w &\longmapsto \left(\frac{z-w}{z-\bar{w}}\right)^{m_{\mathfrak{e}}} \end{aligned}$$

induces a chart around z on $\Gamma\backslash\mathbf{H}$ with image equal to the disc $\{z \in \mathbf{C} \mid |z| < r^{m_{\mathfrak{e}}}\}$. The map $q_{\mathfrak{e}}$ is related to the distance function u on $\mathbf{H}$ by

$$u(z,w) = \frac{1 + |q_{\mathfrak{e}}(w)|^{2/m_{\mathfrak{e}}}}{1 - |q_{\mathfrak{e}}(w)|^{2/m_{\mathfrak{e}}}}.$$

This means that the image of $B_{\mathfrak{e}}(\epsilon)$ under $q_{\mathfrak{e}}$ equals the disc $\{q \in \mathbf{C} \mid |q| < \delta\}$, where δ is chosen such that

$$\frac{4\pi\delta^{2/m_{\mathfrak{e}}}}{1 - \delta^{2/m_{\mathfrak{e}}}} = m_{\mathfrak{e}}\epsilon.$$

For later use, we note that the function

$$\begin{aligned} \mathbf{H} \times \mathbf{H} &\to [1, \infty) \\ (z,w) &\mapsto \min_{\gamma\in\Gamma} u(z, \gamma w) \end{aligned}$$

is Γ-invariant in both variables and hence induces a function

$$d\colon \Gamma\backslash\mathbf{H} \times \Gamma\backslash\mathbf{H} \to [1, \infty).$$

It can be viewed as the hyperbolic cosine of a distance function on $\Gamma\backslash\mathbf{H}$.

Finally, we introduce a *point counting function* which will be useful several times in this chapter. For any two points z, w in $\mathbf{H}$ and $U \geq 1$, we denote by $N_\Gamma(z,w,U)$ the number of translates of w by elements of Γ lying in a disc around z of radius r given by $\cosh(r) = U$, i.e.

$$N_\Gamma(z,w,U) = \#\{\gamma \in \Gamma \mid u(z, \gamma w) \leq U\} \tag{1.5}$$

2. Modular curves and modular forms over the complex numbers

Let us now describe how the notions of modular curves and modular forms as described in Chapter I are related to the classical view of modular curves as quotients of the hyperbolic plane $\mathbf{H}$ by congruence subgroups of $\mathrm{SL}_2(\mathbf{Z})$ and of modular forms as functions on $\mathbf{H}$ satisfying a certain transformation property with respect to these groups. For proofs, we refer to Deligne and Rapoport [23, IV, § 5; VII, § 4].

Over the hyperbolic plane, which we identify as always with the upper half-plane in $\mathbf{C}$, there is a (complex analytic) elliptic curve

$$f\colon E \to \mathbf{H}.$$

This E can be defined as the cokernel of the closed embedding

$$\begin{aligned} \mathbf{Z}_{\mathbf{H}}^2 &\rightarrowtail \mathbf{C}_{\mathbf{H}} \\ \left(\begin{pmatrix} n \\ m \end{pmatrix}, \tau\right) &\mapsto n + m\tau \end{aligned}$$

of topological groups over $\mathbf{H}$. The fibre over a point $\tau \in \mathbf{H}$ can also be described as

$$\begin{aligned} E_\tau = \mathbf{C}/(\mathbf{Z} + \mathbf{Z}\tau) &\xrightarrow{\sim} \mathbf{C}^\times/q^{\mathbf{Z}} \\ z &\longmapsto \exp(2\pi i z), \end{aligned}$$

where

$$q = \exp(2\pi i \tau).$$

The curve E has a global relative differential

$$\alpha_E = 2\pi i\, dz,$$

with z the standard coordinate on $\mathbf{C}$. Via the isomorphism $\mathbf{C}/(\mathbf{Z} + \mathbf{Z}\tau) \xrightarrow{\sim} \mathbf{C}^\times/q^{\mathbf{Z}}$, this corresponds to the differential dt/t, with t the standard coordinate on $\mathbf{C}^\times$.

Let n be a positive integer. We write

$$\Gamma_1(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \;\middle|\; \begin{array}{r} a \equiv d \equiv 1 \pmod{n}, \\ c \equiv 0 \pmod{n} \end{array} \right\}.$$

There is a canonical isomorphism

$$\Gamma_1(n)\backslash\mathbf{H} \xrightarrow{\sim} \mathrm{X}_1(n)^\circ(\mathbf{C}).$$

of (non-compact) Riemann surfaces. Similarly, for every prime number p we define

$$\Gamma_1(n;p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \;\middle|\; \begin{array}{r} a \equiv d \equiv 1 \pmod{n}, \\ c \equiv 0 \pmod{np} \end{array} \right\}.$$

Then there is a canonical isomorphism

$$\Gamma_1(n;p)\backslash\mathbf{H} \xrightarrow{\sim} \mathrm{X}_1(n;p)^\circ(\mathbf{C}).$$

Let Γ be one of the above groups, and let k be a positive integer. Via the canonical differential α_E on E over $\mathbf{H}$, the space $\mathrm{M}_k(\Gamma,\mathbf{C})$ can be identified with the space of modular forms of weight k for Γ in the classical sense, i.e. as holomorphic functions

$$f\colon \mathbf{H} \to \mathbf{C}$$

that satisfy

$$f(\gamma\tau) = (c\tau+d)^k f(\tau) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$$

and are holomorphic at the cusps.

Remark. One can also consider an arbitrary congruence subgroup Γ of $\mathrm{SL}_2(\mathbf{Z})$, i.e. a group containing the kernel of the group homomorphism $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/n\mathbf{Z})$ for some positive integer n. To such a Γ one can associate a moduli stack classifying (generalised) elliptic curves with level structure, with a corresponding coarse moduli scheme $\mathrm{X}(\Gamma)$, such that $\Gamma\backslash\mathbf{H}$ is isomorphic to $\mathrm{X}(\Gamma)(\mathbf{C})$. Since we only need the special cases of $\Gamma_1(n)$ and $\Gamma_1(n;p)$, we refer to Deligne and Rapoport [23] for this more general theory.

2.1. The Petersson inner product

Let f and g be complex analytic modular forms of weight k for a congruence subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbf{Z})$ such that at least one of them is a cusp form. Then the function $\tau \mapsto (\Im\tau)^k f(\tau)\bar{g}(\tau)$ on $\mathbf{H}$ is Γ-invariant and bounded on $\Gamma\backslash\mathbf{H}$, so the integral

$$\langle f,g\rangle_\Gamma = \int_{\tau\in\Gamma\backslash\mathbf{H}} (\Im\tau)^k f(\tau)\bar{g}(\tau)\mu_{\mathbf{H}}(\tau)$$

converges. In particular, this defines a Hermitean inner product

$$\langle\ ,\ \rangle_\Gamma\colon \mathrm{S}_k(\Gamma,\mathbf{C}) \times \mathrm{S}_k(\Gamma,\mathbf{C}) \to \mathbf{C}.$$

This is called the *Petersson inner product* on $\mathrm{S}_k(\Gamma,\mathbf{C})$.

For every positive integer k, we equip the line bundle of cusp forms of weight k with the *Petersson metric*

$$|f|_{k,\mathrm{Pet}}(\tau) = (\Im\tau)^{k/2}|f(\tau)|.$$

This metric vanishes at the cusps.

2.2. Newforms

We briefly describe the theory of newforms, developed by Atkin and Lehner [3] and extended by Li [69]. Let n and k be positive integers. For every divisor d of n and every divisor e of n/d, the map

$$b_e^{n,d}\colon \mathcal{M}_{\Gamma_1(n)} \to \mathcal{M}_{\Gamma_1(d)}$$

defined in § I.1.2 induces an injective $\mathbf{C}$-linear map

$$(b_e^{n,d})^*\colon \mathrm{S}_k(\Gamma_1(d),\mathbf{C}) \to \mathrm{S}_k(\Gamma_1(n),\mathbf{C}).$$

We define a subspace

$$\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C}) \subseteq \mathrm{S}_k(\Gamma_1(n),\mathbf{C})$$

as the orthogonal complement of the subspace of $\mathrm{S}_k(\Gamma_1(n),\mathbf{C})$ spanned by the images of all the $(b_e^{n,d})^*$ with d a strict divisor of n. The Hecke operators form a family of normal commuting operators on $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$. This implies that $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$ admits an orthogonal basis of eigenforms. It follows from the formulae for the action of the Hecke operators on the q-expansion of an eigenform f given in § I.2.4 that the first coefficient $a_1(f)$ of the q-expansion of f at the cusp 0 does not vanish. A *primitive cusp form* of weight k for $\Gamma_1(n)$ is an eigenform in $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$ that is normalised such that $a_1(f) = 1$. We write $P_k(\Gamma_1(d))$ for the set of primitive cusp forms for $\Gamma_1(d)$. The $\mathbf{C}$-vector space $\mathrm{S}_k(\Gamma_1(n),\mathbf{C})$ has a canonical basis

$$B_k(\Gamma_1(n)) = \bigsqcup_{d\mid n}\bigsqcup_{e\mid n/d} (b_e^{n,d})^* P_k(\Gamma_1(d)).$$

The matrix of the Petersson inner product with respect to the basis $B_k(\Gamma_1(n))$ is not in general diagonal. For positive integers

$$d \mid n, \quad d' \mid n, \quad e \mid n/d, \quad e' \mid n/d'$$

and primitive forms f and f' for $\Gamma_1(d)$ and $\Gamma_1(d')$, respectively, the inner product

$$\langle (b_e^{n,d})^* f, (b_{e'}^{n,d'})^* f'\rangle_{\Gamma_1(n)}.$$

vanishes in all cases except possibly when $d = d'$ and $f = f'$.

2.3. Eisenstein series

Let k and n be positive integers. The orthogonal complement of the subspace of cusp forms in $\mathrm{M}_k(\Gamma_1(n),\mathbf{C})$ is called the space of *Eisenstein series*. In [44, § 10], Hecke gave an explicit description of this orthogonal complement. We briefly state the result; for details, we refer to Miyake [80, Chapter 7] or Stein [104, § 5.3].

For every primitive character

$$\epsilon\colon (\mathbf{Z}/n\mathbf{Z})^\times \to \mathbf{C}^\times,$$

we extend ϵ to a function $\mathbf{Z}/n\mathbf{Z} \to \mathbf{C}$ by putting $\epsilon(d) = 0$ if $d \in \mathbf{Z}/n\mathbf{Z}$ is not invertible. Furthermore, we define the *generalised Bernoulli numbers* B_k^ϵ for $k \geq 0$ by the formula

$$\sum_{a=1}^{n} \frac{\epsilon(a)\exp(ax)x}{\exp(nx)-1} = \sum_{k=0}^{\infty} B_k^\epsilon \frac{x^k}{k!}.$$

Let us now consider primitive characters

$$\epsilon_1\colon (\mathbf{Z}/n_1\mathbf{Z})^\times \to \mathbf{C}^\times, \quad \epsilon_2\colon (\mathbf{Z}/n_2\mathbf{Z})^\times \to \mathbf{C}^\times$$

such that $\epsilon_1(-1)\epsilon_2(-1) = (-1)^k$ and $n_1 n_2 \mid n$. We define the formal power series

$$E_k^{\epsilon_1,\epsilon_2}(q) = -\delta_{n_1,1}\frac{B_k^{\epsilon_2}}{2k} + \sum_{m=1}^{\infty}\left(\sum_{d\mid m}\epsilon_1(m/d)\epsilon_2(d)d^{k-1}\right)q^m \in \mathbf{C}[[q]].$$

Assume first that $k \neq 2$. Then for every positive integer t dividing $n/(n_1 n_2)$, there is a modular form $E_k^{\epsilon_1,\epsilon_2,t}$ of weight k for $\Gamma_1(n)$ whose q-expansion at the cusp ∞ is $E_k^{\epsilon_1,\epsilon_2}(q^t)$. When $k = 2$, the same holds for all ϵ_1 and ϵ_2 that are not both the trivial character. As for the case where $k = 2$ and both ϵ_1 and ϵ_2 are trivial, for every divisor $t \mid n$ with $t > 1$ there is a modular form of weight 2 for $\Gamma_1(n)$ whose q-expansion is $E_2(q) - tE_2(q^t)$, where $E_2(q)$ is the power series

$$E_2(q) = -\frac{1}{24} + \sum_{m=1}^{\infty}\sigma_1(m)q^m;$$

here $\sigma_1(m)$ denotes the sum of the positive divisors of m. Moreover, the modular forms mentioned above are eigenforms for the Hecke operators, and they form a basis for the space of Eisenstein series of weight k for $\Gamma_1(n)$.

2.4. Petersson norms of cusp forms

Let f be a primitive form of weight k for $\Gamma_1(n)$. Iwaniec proved in [48] that for all $\epsilon > 0$, the (squared) Petersson norm $\langle f, f\rangle_{\Gamma_1(n)}$ of f satisfies

$$\frac{\langle f, f\rangle_{\Gamma_1(n)}}{\mathrm{vol}_{\Gamma_1(n)}} \leq A_{k,\epsilon} n^\epsilon$$

for some positive real number $A_{k,\epsilon}$ independent of n and f. Below we will give such an $A_{k,\epsilon}$ explicitly, using the Rankin–Selberg L-function attached to f, the functional equation proved by Li [70], and the Ramanujan–Petersson bound proved by Deligne in [20] and [21]. If the q-expansion of f is given by

$$f = \sum_{m=1}^{\infty} a_m(f)q^m,$$

then the Rankin–Selberg L-function attached to f is defined for $s \in \mathbf{C}$ with $\Re s > 1$ by the series

$$L_{f,\bar{f}}(s) = \sum_{m=1}^{\infty}\frac{|a_m(f)|^2}{m^{s+k-1}}$$

Rankin proved in [84] that $L_{f,\bar f}$ can be continued to a meromorphic function on $\mathbf{C}$, having a simple pole at $s = 1$ with residue

$$\operatorname{res}_1 L_{f,\bar f} = \frac{(4\pi)^k \langle f, f\rangle_{\Gamma_1(n)}}{\Gamma(k)\operatorname{vol}_{\Gamma_1(n)}},$$

where Γ is the usual gamma function; see also Li [70, Theorem 3.2].

We now assume for simplicity that f is *p-primitive* at every prime number $p \mid n$ for which $a_p(f) = 0$ in the sense of [70, page 139]. For forms f that do not satisfy this condition, we actually get a slightly sharper bound by "lowering the level". To state the functional equation, we introduce the following notation: if the prime factorisation of n is

$$n = \prod_{p \text{ prime}} p^{r(p)},$$

then we write

$$c_f = \prod_{p \text{ prime}} p^{2\lfloor (r(p)+1)/2\rfloor},$$

and we define S as the set of prime numbers p dividing n except those for which $a_p(f) = 0$ and $r(p)$ is even. We consider the "completed" L-function

$$\tilde L_{f,\bar f}(s) = c_f^{s/2}(2\pi)^{-2s}\Gamma(s)\Gamma(s+k-1)\zeta(2s)\prod_{p\in S}(1+p^{-s})L_{f,\bar f}(s),$$

where ζ is the Riemann zeta function. The function $\tilde L_{f,\bar f}$ has a simple pole at $s = 1$ with residue

$$\operatorname{res}_1 \tilde L_{f,\bar f} = c_f^{1/2}(2\pi)^{-2}\zeta(2)\prod_{p\in S}(1+p^{-1})\frac{(4\pi)^k\langle f, f\rangle_{\Gamma_1(n)}}{\operatorname{vol}_{\Gamma_1(n)}}.$$

In [70, Theorem 2.2], Li proved the functional equation

$$\tilde L_{f,\bar f}(s) = \tilde L_{f,\bar f}(1-s).$$

Lemma 2.1. *Let n and k be positive integers, and let f be a primitive form of weight k for $\Gamma_1(n)$. Then for all $\epsilon > 0$ the Petersson norm of f satisfies*

$$\frac{\langle f, f\rangle_{\Gamma_1(n)}}{\operatorname{vol}_{\Gamma_1(n)}} \le A_{k,\epsilon} c_f^{\epsilon/2} \le A_{k,\epsilon} n^{\epsilon},$$

where

$$A_{k,\epsilon} = (4\pi)^{-k}(2\pi)^{-\epsilon}\frac{\zeta(2+2\epsilon)}{\zeta(2)}\sum_{m=1}^{\infty}\frac{\sigma_0(m)^2}{m^{1+\epsilon}}\sup_{\Re s = 1+\epsilon}\bigl|s(1-s)\Gamma(s)\Gamma(s+k-1)\bigr|.$$

Here $\sigma_0(m)$ denotes the number of positive divisors of m.

Proof. We apply the Phragmén–Lindelöf principle (see for example Markushevich [74, Volume II, Theorem 7.7]) to the function $s(1-s)\tilde{L}_{f,\bar{f}}(s)$ on the strip $\{s \in \mathbf{C} \mid -\epsilon \leq \Re s \leq 1+\epsilon\}$; this is permitted by Stirling's estimate for the Γ-function. This shows that $\left|s(1-s)\tilde{L}_{f,\bar{f}}(s)\right|$ is bounded by its values on the boundary of the strip. The functional equation now implies

$$\operatorname{res}_1 \tilde{L}_{f,\bar{f}} \leq \sup_{\Re s=1+\epsilon} \left|s(1-s)\tilde{L}_{f,\bar{f}}(s)\right|.$$

Plugging in the definition of $\tilde{L}_{f,\bar{f}}$ and the expression for the residue at $s=1$ given above and using the inequalities

$$|\zeta(2s)| \leq \zeta(2+2\epsilon), \quad |1+p^{-s}| < 1+p^{-1}, \quad |L_{f,\bar{f}}(s)| \leq \sum_{m=1}^{\infty} \frac{|a_m(f)|^2}{m^{k+\epsilon}}$$

for $\Re s = 1+\epsilon$, we deduce that

$$\frac{\langle f,f\rangle_{\Gamma_1(n)}}{\operatorname{vol}_{\Gamma_1(n)}} \leq \frac{c_f^{\epsilon/2}}{(4\pi)^k(2\pi)^\epsilon}\frac{\zeta(2+2\epsilon)}{\zeta(2)} \sup_{\Re s=1+\epsilon}\left|s(1-s)\Gamma(s)\Gamma(s+k-1)\right| \sum_{m=1}^{\infty}\frac{|a_m(f)|^2}{m^{k+\epsilon}}.$$

The first inequality in the statement of the lemma now follows from Deligne's bound

$$|a_m(f)| \leq \sigma_0(m)m^{\frac{k-1}{2}}.$$

The second is a result of the easily verified inequality $c_f \leq n^2$. □

In addition to the above upper bound for primitive forms, we note the following "trivial" lower bound for the Petersson norm of a cusp form with integral q-expansion.

Lemma 2.2. *Let $n \geq 1$ and $k \geq 2$ be integers, and let f be a non-zero element of $\mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n))$. Then we have*

$$\langle f,f\rangle_{\Gamma_1(n)} \geq \frac{\exp(-4\pi(d(k,n)+1))}{4\pi(d(k,n)+1)},$$

where $d(k,n)$ is the degree of the line bundle $\omega^{\otimes k}(-\text{cusps})$ on $\mathcal{M}_{\Gamma_1(n)}$.

Proof. The open subset

$$\{z \in \mathbf{C} \mid -\tfrac{1}{2} < \Re(-1/z) < 1/2 \text{ and } \Im(-1/z) > 1\}$$

maps injectively to $\Gamma_1(n)\backslash\mathbf{H}$. This implies that if the q-expansion of f at the cusp 0 is given by

$$f(z) = \sum_{m=1}^{\infty} a_m q_0(z)^m \quad \text{with } q_0(z) = \exp(-2\pi i/z),$$

then

$$\langle f,f\rangle_{\Gamma_1(n)} \geq \int_1^\infty y^{k-2}\sum_{m=1}^{\infty} a_m(f)^2\exp(-4\pi m y)dy.$$

By definition, all the a_m are in $\mathbf{Z}$, and at least one of $a_1, \ldots, a_{d(k,n)+1}$ is non-zero since otherwise f would be the zero form. This implies that

$$\begin{aligned}\langle f, f\rangle_{\Gamma_1(n)} &\geq \int_1^\infty y^{k-2}\exp(-4\pi(d(k,n)+1)y)dy \\ &\geq \int_1^\infty \exp(-4\pi(d(k,n)+1)y)dy \\ &= \frac{\exp(-4\pi(d(k,n)+1))}{4\pi(d(k,n)+1)},\end{aligned}$$

which proves the lemma. □

3. Spectral theory of Fuchsian groups

Let Γ be a cofinite Fuchsian group. We denote by $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$ the Hilbert space of square-integrable complex-valued functions on $\Gamma\backslash\mathbf{H}$ (with respect to the measure given by $\mu_{\mathbf{H}}$), and by $\langle\ ,\ \rangle$ the standard inner product on this Hilbert space.

The Laplace operator Δ on the space of smooth Γ-invariant functions with compact support on $\mathbf{H}$ can be extended to an (unbounded) self-adjoint operator on the Hilbert space $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$, defined on a dense subspace; we denote this extension by Δ as well. The spectrum of Δ consists of a discrete part and a continuous part.

3.1. Automorphic forms of weight 0

The discrete spectrum consists of eigenvalues of Δ and is of the form $\{\lambda_j\}_{j=0}^\infty$ with

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots, \quad \lambda_j \to \infty \text{ as } j \to \infty.$$

Let $\{\phi_j\}_{j=0}^\infty$ be a corresponding set of eigenfunctions; these are called *automorphic forms of Maaß* (of weight 0). We may (and do) assume that they are orthonormal with respect to the inner product on $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$. For each $j \geq 0$, we define complex numbers s_j and t_j by

$$\lambda_j = s_j(1-s_j) \quad \text{and} \quad s_j = \frac{1}{2} + it_j$$

with $s_j \in [1/2, 1]$ if $\lambda_j \leq 1/4$. For $\lambda_j > 1/4$, the s_j are only determined up to $s_j \leftrightarrow 1 - s_j$ and the t_j are only determined up to sign.

3.2. Eisenstein–Maaß series of weight 0

The continuous part of the spectrum of the Laplace operator on $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$ is the interval $[1/4, \infty)$, with multiplicity equal to the number of cusps of Γ. In particular, the continuous spectrum is absent if Γ has no cusps. The continuous spectrum does not consist of eigenvalues, but corresponds to "wave packets" that can be constructed from *non-holomorphic Eisenstein series* or *Eisenstein–Maaß series* (introduced by Maaß in [72]). These series are defined as follows: for every cusp $\mathfrak{c}$ of Γ the series

$$E_{\mathfrak{c}}(z,s) = \sum_{\gamma\in\Gamma_{\mathfrak{c}}\backslash\Gamma} (\Im\sigma_{\mathfrak{c}}^{-1}\gamma z)^s \quad (z \in \mathbf{H}, s \in \mathbf{C} \text{ with } \Re s > 1)$$

converges uniformly on sets of the form $K \times \{s \in \mathbf{C} \mid \Re s \geq \delta\}$ with K a compact subset of $\mathbf{H}$ and $\delta > 1$. In particular, $E_{\mathfrak{c}}(z,s)$ is a holomorphic function of s.

A crucial ingredient in the spectral theory of automorphic forms is the *meromorphic continuation of Eisenstein series*, due to Selberg [95]. The functions $E_{\mathfrak{c}}(z,s)$ can be continued to functions of the form $E_{\mathfrak{c}}(z,s) = H(z,s)/G(s)$, where H is a smooth function of $(z,s) \in \mathbf{H}\times\mathbf{C}$ and both G and H are entire functions of s. These meromorphic continuations have a finite number of simple poles on the segment $(1/2, 1]$ and no other poles in $\{s \in \mathbf{C} \mid \Re s \geq 1/2\}$. Furthermore, the meromorphically continued Eisenstein series satisfy a functional equation of the form

$$E_{\mathfrak{c}}(z,s) = \sum_{\mathfrak{d}} \phi_{\mathfrak{c},\mathfrak{d}}(s) E_{\mathfrak{d}}(z, 1-s),$$

for certain meromorphic functions $\phi_{\mathfrak{c},\mathfrak{d}}$, which we will not write down. We refer to Hejhal [46, Chapter VI, § 11] for a construction of the meromorphic continuation of the $E_{\mathfrak{c}}(z,s)$ and proofs of the functional equation and the other properties stated here; see Faddeev [36, § 4], Hejhal [46, Appendix F] or Iwaniec [49, Chapter 6] for different constructions.

For each $s \in \mathbf{C}$ such that $E_{\mathfrak{c}}(z,s)$ is holomorphic in s for all $z \in \mathbf{H}$, the function $z \mapsto E_{\mathfrak{c}}(z,s)$ is Γ-invariant and satisfies the differential equation

$$\Delta E_{\mathfrak{c}}(\ ,s) = s(1-s) E_{\mathfrak{c}}(\ ,s).$$

If s is a complex number with $\Re s = 1/2$, the Eisenstein–Maaß series $E_{\mathfrak{c}}(\ ,s)$ are integrable, but not square-integrable, as functions on $\Gamma\backslash\mathbf{H}$. In contrast, the "wave packets" mentioned above are square-integrable. They are constructed as follows: if $g\colon [0,\infty) \to \mathbf{C}$ is a smooth function with compact support, then the function

$$\begin{aligned} \mathcal{E}_{\mathfrak{c}} g\colon \mathbf{H} &\longrightarrow \mathbf{C} \\ z &\longmapsto \frac{1}{2\pi} \int_0^\infty g(t) E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt \end{aligned}$$

is in $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$, and by extension we get an embedding of Hilbert spaces

$$\mathcal{E}_{\mathfrak{c}}\colon \mathrm{L}^2\Bigl([0,\infty), \frac{1}{2\pi} dt\Bigr) \longrightarrow \mathrm{L}^2(\Gamma\backslash\mathbf{H}).$$

The orthogonal projection on the image of $\mathcal{E}_{\mathfrak{c}}$, which we denote by $\Pi_{\mathfrak{c}}$, is given by the following formula (valid for smooth and bounded Γ-invariant functions $f\colon \mathbf{H} \to \mathbf{C}$):

$$\begin{aligned} \Pi_{\mathfrak{c}} f\colon [0,\infty) &\to \mathbf{C} \\ t &\mapsto \int_{z \in \Gamma\backslash\mathbf{H}} f(z) \bar{E}_{\mathfrak{c}}(z, \tfrac{1}{2} + it) \mu_{\mathbf{H}}(z). \end{aligned}$$

3.3. Spectral theory for automorphic forms of weight 0

The following result is fundamental in the theory of automorphic forms of weight 0.

Theorem 3.1 (see Iwaniec [49, Theorems 4.7 and 7.3]; cf. Faddeev [36, Theorem 4.1]). *Every smooth and bounded* Γ*-invariant function* $f\colon \mathbf{H}\to\mathbf{C}$ *has the spectral representation*

$$f(z)=\sum_{j=0}^{\infty}\langle f,\phi_j\rangle\phi_j(z)+\sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^{\infty}(\Pi_{\mathfrak{c}}f)(t)\,E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)dt,$$

where $\mathfrak{c}$ *runs over the cusps of* Γ, *in the sense that the right-hand side converges to* f *in the Hilbert space* $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$. *If in addition the smooth* Γ*-invariant function* $\Delta f\colon\mathbf{H}\to\mathbf{C}$ *is bounded, the convergence is uniform on compacta in* $\mathbf{H}$.

There is an analogous result (Theorem 3.2 below) for functions on $\mathbf{H}\times\mathbf{H}$ that are of the form $\sum_{\gamma\in\Gamma}k(u(z,\gamma w))$, where $k\colon[1,\infty)\to\mathbf{R}$ is a function satisfying certain conditions. To state the conditions and the result, we need the Selberg–Harish-Chandra transform of k, introduced in § 1.1. We also have to explain the type of convergence provided by the theorem below. Let A be a filtered set, and let $\{K_a\}_{a\in A}$ be a family of continuous functions on $\Gamma\backslash\mathbf{H}\times\Gamma\backslash\mathbf{H}$, square-integrable in the second variable. If K is a function such that for all compact subsets C of $\Gamma\backslash\mathbf{H}$ we have

$$\lim_{a\in A}\biggl(\sup_{z,w\in C}|K_a(z,w)-K(z,w)|+\sup_{z\in C}\int_{w\in\Gamma\backslash\mathbf{H}}|K_a(z,w)-K(z,w)|^2\mu_{\mathbf{H}}(w)\biggr)=0,$$

we say that the family of functions $\{K_a\}_{a\in A}$ *converges to* K *in the* $(\mathrm{L}^\infty_{\mathrm{loc}},\mathrm{L}^2\cap\mathrm{L}^\infty_{\mathrm{loc}})$*-topology*. In other words, this condition means that the family converges uniformly on compacta in $\mathbf{H}\times\mathbf{H}$, and also with respect to the L^2-norm in the variable w, uniformly for z in compacta of $\Gamma\backslash\mathbf{H}$.

Theorem 3.2 (see Iwaniec [49, Theorem 7.4]). *Let* $k\colon[1,\infty)\to\mathbf{R}$ *be a function that is the inverse Selberg–Harish-Chandra transform of a function* h *satisfying the conditions (H1) and (H2) of § 1.1. Then the function*

$$\begin{aligned}K\colon\mathbf{H}\times\mathbf{H}&\longrightarrow\mathbf{R}\\ (z,w)&\longmapsto\frac{1}{\#(\Gamma\cap\{\pm1\})}\sum_{\gamma\in\Gamma}k(u(z,\gamma w))\end{aligned}$$

is Γ*-invariant with respect to both variables and admits the spectral representation*

$$K(z,w)=\sum_{j=0}^{\infty}h(t_j)\phi_j(z)\bar\phi_j(w)+\sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^{\infty}h(t)E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt.\quad(3.1)$$

More precisely, the right-hand side converges to $K(z,w)$ *in the following sense. For* J *a positive integer and* T *a positive real number, we define*

$$K^{J,T}(z,w)=\sum_{j=0}^{J}h(t_j)\phi_j(z)\bar\phi_j(w)+\sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^{T}h(t)E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt.$$

Then as J *and* T *tend to infinity,* $K^{J,T}$ *converges to* K *in the* $(\mathrm{L}^\infty_{\mathrm{loc}},\mathrm{L}^2\cap\mathrm{L}^\infty_{\mathrm{loc}})$*-topology.*

3.4. Bounds on eigenfunctions

The convergence of the spectral representation (3.1) can be deduced from suitable bounds on the values of the Γ-invariant function

$$\begin{aligned} \mathbf{H} &\longrightarrow [0,\infty) \\ z &\longmapsto \sum_{j\colon \lambda_j\le T} |\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^{\sqrt{T-1/4}} \bigl|E_{\mathfrak{c}}(z,\tfrac{1}{2}+it)\bigr|^2 dt \end{aligned}$$

as $T\to\infty$. In Lemma 3.4 below, we will provide such bounds in a way that applies at the same time to all the subgroups of finite index in a given Fuchsian group Γ_0. This will turn out to be useful later for bounding suprema of Green functions in a uniform way.

We define a function $k_U\colon [1,\infty)\to\mathbf{R}$ by

$$k_U(u) = \begin{cases} 1 & \text{if } u\le U;\\ 0 & \text{if } u>U.\end{cases} \tag{3.2}$$

From (1.3) and the formula for $\int_1^z P_\nu(w)dw$ found in Erdélyi et al. [34, § 3.6.1, equation 8], we see that the Selberg–Harish-Chandra transform of k_U is

$$h_U(t) = 2\pi\sqrt{U^2-1}\,P^{-1}_{-1/2+it}(U).$$

Here P^μ_ν is the associated Legendre function of degree ν and order μ; see [34, § 3.2].

Lemma 3.3. *Suppose $U\in[1,3]$ and $t\in\mathbf{R}\cup\bigl[-\frac{1}{2},\frac{1}{2}\bigr]i$ are such that*

$$\bigl(\tfrac{1}{4}+t^2\bigr)(U-1)\le\frac{1}{2}.$$

Then the real number $h_U(t)$ satisfies the inequalities

$$(4\pi-8)(U-1)\le h_U(t)\le 8(U-1).$$

Proof. We start by expressing the Legendre function P^μ_ν in terms of Gauß's hypergeometric function $F(a,b;c;z)$. (This function is described in Erdélyi et al. [34, Chapter II].) Because of the many transformation identities satisfied by the hypergeometric function, there are lots of ways to do this. We use [34, § 3.2, equation 3]; this gives

$$h_U(t) = 2\pi(U-1)\,F\biggl(\frac{1}{2}+it,\frac{1}{2}-it;2;\frac{1-U}{2}\biggr).$$

Next we use the hypergeometric series for $F(a,b;c;z)$ with convergence radius 1 (see [34, § 2.1, equation 2]):

$$F(a,b;c;z) = \sum_{n=0}^\infty \frac{(a)_n(b)_n}{(c)_n n!}z^n, \tag{3.3}$$

where $(\ )_n$ is Pochhammer's rising factorial symbol, defined by

$$(a)_n = \Gamma(a+n)/\Gamma(a) = a(a+1)\cdots(a+n-1).$$

Putting $x = \frac{U-1}{2}$ for a moment, using the series expansion (3.3) and applying the triangle inequality, we get the bound

$$\left|F\left(\tfrac{1}{2}+it, \tfrac{1}{2}-it; 2; -x\right) - 1\right| \le \sum_{n\ge 1} \left| \frac{\left(\frac{1}{2}+it\right)_n \left(\frac{1}{2}-it\right)_n}{(2)_n n!} (-x)^n \right|$$

The assumption $\left(\frac{1}{4}+t^2\right)(U-1) \le \frac{1}{2}$ is equivalent to $\left(\frac{1}{4}+t^2\right) x \le \frac{1}{4}$. Therefore the n-th term in the series on the right-hand side can be bounded as follows:

$$\begin{aligned}
\left| \frac{\left(\frac{1}{2}+it\right)_n \left(\frac{1}{2}-it\right)_n}{(2)_n n!} (-x)^n \right| &= \frac{\prod_{k=0}^{n-1} \left(\left(\frac{1}{4}+t^2\right)x + k(k+1)x\right)}{(2)_n n!} \\
&\le \frac{\prod_{k=0}^{n-1} \left(\frac{1}{4}+k(k+1)\right)}{(2)_n n!} \\
&= \frac{\left(\frac{1}{2}\right)_n \left(\frac{1}{2}\right)_n}{(2)_n n!}.
\end{aligned}$$

This implies that

$$\begin{aligned}
\left|F\left(\tfrac{1}{2}+it, \tfrac{1}{2}-it; 2; -x\right) - 1\right| &\le F\left(\tfrac{1}{2}, \tfrac{1}{2}; 2; 1\right) - 1 \\
&= 4/\pi - 1,
\end{aligned}$$

where the last equality follows from the formula

$$F(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-b-a)}{\Gamma(c-a)\Gamma(c-b)} \quad \text{for } \Re c > 0 \text{ and } \Re c > \Re(a+b)$$

(see Erdélyi et al. [34, § 2.1.3, equation 14] or Iwaniec [49, equation B.20]) and the fact that $\Gamma(3/2) = \sqrt{\pi}/2$. We conclude that

$$\begin{aligned}
\left|h_U(t) - 2\pi(U-1)\right| &= 2\pi(U-1)\left|F\left(\tfrac{1}{2}+it, \tfrac{1}{2}-it; 2; -x\right) - 1\right| \\
&\le 2\pi(U-1)(4/\pi - 1) \\
&= (8-2\pi)(U-1),
\end{aligned}$$

which is equivalent to the inequalities in the statement of the lemma. □

The following result can be used to show that the spectral representation in Theorem 3.2 converges; however, we will also apply it in Section 5 below in order to find upper bounds for Green functions of Fuchsian groups. To state the result, we introduce the notation

$$N'_\Gamma(z,U) = \#\{\gamma \in \Gamma \mid u(z,\gamma z) \le 2U^2 - 1\} \quad \text{for } z \in \mathbf{H} \text{ and } U \ge 1;$$

this defines a Γ-invariant function of z.

Lemma 3.4. *Let Γ be a cofinite Fuchsian group. For all $z \in \mathbf{H}$ and all $T \geq 1/4$, the Maaß forms ϕ_j and the Eisenstein–Maaß series $E_{\mathfrak{c}}$ for Γ corresponding to eigenvalues less than or equal to T satisfy the inequality*

$$\sum_{j\colon \lambda_j \leq T} |\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\sqrt{T-1/4}} \bigl|E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2}+it\bigr)\bigr|^2 dt \leq \frac{\pi T}{(2\pi-4)^2} N'_\Gamma\left(z, 1+\frac{1}{2T}\right).$$

Proof. For a given $T \geq 1/4$, we put

$$U = 1 + \frac{1}{2T} \in (1,3],$$

so that $T(U-1) = 1/2$. We note that

$$\sum_{\gamma \in \Gamma} k_U(u(z,w)) = N_\Gamma(z,w,U),$$

where N_Γ is the point counting function defined in (1.5). From Bessel's inequality one can deduce that

$$\begin{aligned}\sum_{j\colon \lambda_j \leq T} |h_U(t_j)\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\sqrt{T-1/4}} \bigl|h_U(t)E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2}+it\bigr)\bigr|^2 dt& \\ \leq \int_{w \in \Gamma\backslash\mathbf{H}} N_\Gamma(z,w,U)^2 \mu_{\mathbf{H}}(w)&;\end{aligned}$$

see Iwaniec [49, § 7.2]. The inequality $h_U(t) \geq (2\pi-4)/T$ given by Lemma 3.3 implies

$$\begin{aligned}\sum_{j\colon \lambda_j \leq T} |\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\sqrt{T-1/4}} \bigl|E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2}+it\bigr)\bigr|^2 dt& \\ \leq \frac{T^2}{(2\pi-4)^2} \int_{w \in \Gamma\backslash\mathbf{H}} N_\Gamma(z,w,U)^2 \mu_{\mathbf{H}}(w)&\end{aligned}$$

for $z \in \Gamma\backslash\mathbf{H}$ and all $T \geq \frac{1}{4}$.

It remains to bound the integral on the right-hand side of the above inequality. For this we rewrite it as follows (cf. Iwaniec [49, page 109]):

$$\begin{aligned}\int_{w \in \Gamma\backslash\mathbf{H}} N_\Gamma(z,w,U)^2 \mu_{\mathbf{H}}(w) &= \sum_{\gamma,\gamma' \in \Gamma} \int_{w \in \Gamma\backslash\mathbf{H}} k_U(z,\gamma' w) k_U(\gamma z, \gamma' w) \mu_{\mathbf{H}}(w) \\ &= \sum_{\gamma \in \Gamma} \int_{w \in \mathbf{H}} k_U(z,w) k_U(\gamma z, w) \mu_{\mathbf{H}}(w).\end{aligned}$$

The last integral can be interpreted as the area of the intersection of the discs of radius r around the points z and γz of $\mathbf{H}$, where $\cosh r = U$. By the triangle inequality for the hyperbolic distance, this intersection is empty unless

$$u(z,\gamma z) \leq \cosh(2r) = 2U^2 - 1;$$

furthermore, the area of this intersection is at most $2\pi(U-1) = \pi/T$. From this we deduce that

$$\int_{w\in\Gamma\backslash\mathbf{H}} N_\Gamma(z,w,U)^2\mu_{\mathbf{H}}(w) \le \frac{\pi}{T} N_\Gamma(z,z,2U^2-1)$$

By the definition of $N'_\Gamma(z,U)$, this proves the lemma. □

Corollary 3.5. *Let Γ_0 be a cofinite Fuchsian group, fix a compact subset Y_0 of $\Gamma_0\backslash\mathbf{H}$, and write*

$$\nu(Y_0,T) = \frac{\pi T}{(2\pi-4)^2}\sup_{z\in Y_0} N'_{\Gamma_0}\left(z, 1+\frac{1}{2T}\right),$$

with $N'_{\Gamma_0}(z,U)$ as in Lemma 3.4. Let Γ be a subgroup of finite index in Γ_0, and let Y be the inverse image of Y_0 in $\Gamma\backslash\mathbf{H}$. Then the Maaß forms ϕ_j and the Eisenstein–Maaß series $E_{\mathfrak{c}}$ for Γ corresponding to eigenvalues $\le T$ satisfy the inequality

$$\sum_{j\colon \lambda_j\le T} |\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^{\sqrt{T-1/4}} \left|E_{\mathfrak{c}}\left(z,\tfrac{1}{2}+it\right)\right|^2 dt \le \nu(Y_0,T)$$

for all $z\in Y$ and all $T\ge 1/4$.

We note for later use that the function $\nu(Y_0,T)$ in the preceding result is bounded by a linear function of T.

3.5. The hyperbolic lattice point problem

Let Γ be a cofinite Fuchsian group. By the *hyperbolic lattice point problem* for Γ we mean the following question: what is the asymptotic behaviour of the point counting function $N_\Gamma(z,w,U)$ (defined in (1.5)) as $U\to\infty$? The Euclidean analogue of this question (about the number of points in $\mathbf{Z}^2$ lying inside a given disc in $\mathbf{R}^2$) was first treated by Gauß using an elementary packing method, and the error term was later improved using spectral theory on $\mathbf{R}^2/\mathbf{Z}^2$. In the hyperbolic setting, no packing method is known to even give the dominant term of $N_\Gamma(z,w,U)$ as $U\to\infty$; the difficulty here is that the circumference of a circle in the hyperbolic plane grows as fast as its area as the radius goes to infinity. To produce estimates for $N_\Gamma(z,w,U)$, we will use a more sophisticated tool, namely spectral theory on $\Gamma\backslash\mathbf{H}$.

The strategy is to take suitable functions

$$k_U^+, k_U^-\colon [1,\infty)\to\mathbf{R}$$

with compact support, and to define functions K_U^+ and K_U^- on $\mathbf{H}\times\mathbf{H}$, invariant with respect to the action of Γ on each of the two variables, by

$$K_U^\pm(z,w) = \sum_{\gamma\in\Gamma} k_U^\pm(u(z,\gamma w)).$$

Notice that the sum is finite because the functions $k_U^\pm$ have compact support. We take the functions $k_U^\pm$ such that the inequality

$$K_U^-(z,w) \le N_\Gamma(z,w,U) \le K_U^+(z,w) \tag{3.4}$$

holds for all $z, w \in \mathbf{H}$ and $U > 1$. Provided the Selberg–Harish-Chandra transforms $h_U^{\pm}$ of $k_U^{\pm}$, defined in § 1.1, satisfy the conditions of Theorem 3.2, the functions $K_U^{\pm}$ have spectral representations

$$\begin{aligned} K_U^{\pm}(z,w) = \sum_{j=0}^{\infty} h_U^{\pm}(t_j)\phi_j(z)\bar{\phi}_j(w) \\ + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^{\infty} h_U^{\pm}(t) E_{\mathfrak{c}}\bigl(z, \tfrac12 + it\bigr)\bar{E}_{\mathfrak{c}}\bigl(w, \tfrac12 + it\bigr)dt \end{aligned} \tag{3.5}$$

for all $z, w \in \mathbf{H}$. These spectral representations can then be used to find the asymptotic behaviour of $N_\Gamma(z, w, U)$ as $U \to \infty$.

A reasonable choice at first sight would be to take for both k_U^+ and k_U^- the function k_U defined by (3.2), so that the inequalities in (3.4) become equalities. Unfortunately, the Selberg–Harish-Chandra transform h_U of k_U does not decay quickly enough as $t \to \infty$ to give a spectral representation of $N_\Gamma(z, w, U)$ as in Theorem 3.2. Following Iwaniec [49, Chapter 12] (cf. Patterson [82]), we therefore take

$$k_U^+(u) = \begin{cases} 1 & \text{if } 1 \le u \le U, \\ \frac{V-u}{V-U} & \text{if } U \le u \le V, \\ 0 & \text{if } V \le u \end{cases}$$

and

$$k_U^-(u) = \begin{cases} 1 & \text{if } 1 \le u \le T, \\ \frac{U-u}{U-T} & \text{if } T \le u \le U, \\ 0 & \text{if } U \le u \end{cases}$$

for certain T, V, depending on U, with $1 \le T < U < V$. It turns out that a suitable choice is

$$V - U \sim U - T \sim \beta U^{2/3} \quad \text{as } U \to \infty, \text{ for some } \beta > 0. \tag{3.6}$$

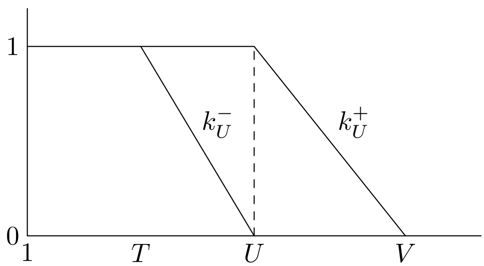

Figure 1: The functions k_U^+ and k_U^-.

Using (1.3), integrating by parts and applying the integral relation between the Legendre functions P_ν and P_ν^{-2} given in Erdélyi et al. [34, § 3.6.1, equation 8], we get

$$h_U^+(t) = 2\pi \frac{(V^2-1)P_{-1/2+it}^{-2}(V) - (U^2-1)P_{-1/2+it}^{-2}(U)}{V-U}.$$

Replacing (U, V) by (T, U) we get a similar formula for $h_U^-(t)$.

Theorem 3.6. *Let Γ be a cofinite Fuchsian group. For all $z, w \in \mathbf{H}$, the point counting function N_Γ satisfies*

$$N_\Gamma(z,w,U) = \sum_{j\colon 2/3<s_j\le 1} 2^{s_j}\sqrt{\pi}\frac{\Gamma(s_j-\frac{1}{2})}{\Gamma(s_j+1)}\phi_j(z)\bar{\phi}_j(w)U^{s_j} + O(U^{2/3}) \quad \text{as } U\to\infty,$$

with an implied constant depending on Γ and the points z and w.

Proof. See Iwaniec [49, Theorem 12.1]. □

In particular, since $|\phi_0|^2$ is the constant function $1/\mathrm{vol}_\Gamma$, where vol_Γ is the volume of $\Gamma\backslash\mathbf{H}$, this shows that

$$N_\Gamma(z,w,U) \sim \frac{2\pi(U-1)}{\mathrm{vol}_\Gamma} \quad \text{as } U\to\infty.$$

The main term in the estimate comes from the eigenvalue $\lambda_0 = 0$, corresponding to $t_0 = \pm i/2$. It follows from (1.2) that

$$h_U^+(\pm i/2) = 2\pi(U-1) + \pi(V-U). \tag{3.7}$$

Since $2\pi(U-1)$ is the area of a disc of radius r with $\cosh r = U$, Theorem 3.6 is the result that one would intuitively expect, in the sense that it shows that this area is asymptotically equivalent to the number of lattice points inside the disc times the area of a fundamental domain for the action of Γ.

For future reference, we also derive an estimate for the derivatives of the functions $K_U^\pm(z,w)$ with respect to U. For this we assume that T and V are differentiable and satisfy

$$T'(U) = 1 + O(U^{-\delta}) \text{ and } V'(U) = 1 + O(U^{-\delta}) \quad \text{as } U\to\infty \tag{3.8}$$

for some $\delta > 0$. By differentiating k_U with respect to U, applying the definition of K_U^+ and estimating the sum using Theorem 3.6, it is straightforward to prove that

$$\frac{d}{dU}K_U^+(z,w) = \frac{2\pi}{\mathrm{vol}_\Gamma} + O(U^{-\epsilon}) \tag{3.9}$$

for some $\epsilon > 0$.

3.6. The Green function of a Fuchsian group

We fix a cofinite Fuchsian group Γ, and we write vol_Γ for the volume of $\Gamma\backslash\mathbf{H}$. The Laplace operator on $\Gamma\backslash\mathbf{H}$ is invertible on the orthogonal complement of the constant functions in the following sense: there exists a unique bounded self-adjoint operator R on $\mathrm{L}^2(\Gamma\backslash\mathbf{H}, \mu_\mathbf{H})$ such that for all smooth and bounded functions f on $\Gamma\backslash\mathbf{H}$ the function Rf satisfies

$$\Delta Rf = f - \frac{1}{\mathrm{vol}_\Gamma}\int_{\Gamma\backslash\mathbf{H}} f\mu_\mathbf{H} \quad \text{and} \quad \int_{\Gamma\backslash\mathbf{H}} Rf\mu_\mathbf{H} = 0.$$

With regard to the spectral representations provided by Theorem 3.1, the effect of R is as follows: if f has the spectral representation

$$f(z) = \sum_{j=0}^{\infty} b_j \phi_j(z) + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\infty} b_{\mathfrak{c}}(t) E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt,$$

then Rf has the corresponding spectral representation

$$Rf(z) = \sum_{j=1}^{\infty} \frac{b_j}{\lambda_j} \phi_j(z) + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\infty} \frac{b_{\mathfrak{c}}(t)}{1/4 + t^2} E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt.$$

(Note the absence of the eigenvalue $\lambda_0 = 0$.)

There exists a unique function

$$\mathrm{gr}_\Gamma\colon \{(z, w) \in \mathbf{H} \times \mathbf{H} \mid z \notin \Gamma w\} \to \mathbf{R}$$

with the following properties:

(1) gr_Γ is smooth and Γ-invariant in both variables;

(2) $\mathrm{gr}_\Gamma(z, w) = \mathrm{gr}_\Gamma(w, z)$;

(3) for fixed $w \in \Gamma\backslash\mathbf{H}$ and z near a cusp $\mathfrak{c}$ of Γ, the behaviour of $\mathrm{gr}_\Gamma(z, w)$ is

$$\mathrm{gr}_\Gamma(z, w) = \log(\Im\sigma_{\mathfrak{c}}^{-1} z) + O(1) \text{ as } \Im\sigma_{\mathfrak{c}}^{-1} z \to \infty;$$

(4) if f is a smooth and bounded Γ-invariant function on $\mathbf{H}$, then the function Rf is given by

$$Rf(z) = -\int_{w \in \Gamma\backslash\mathbf{H}} \mathrm{gr}_\Gamma(z, w) f(w) \mu_{\mathbf{H}}(w).$$

The function gr_Γ is called the *Green function* of the Fuchsian group Γ. In § 5.1 below, we will give a construction of gr_Γ that will allow us to study it quantitatively.

Remark. Different normalisations of gr_Γ occur in the literature; for example, our gr_Γ is $1/4\pi$ times the function defined by Gross in [40, § 9].

3.7. Automorphic forms of general weight

We recall the definition of automorphic forms of arbitrary real weight. We also describe q-expansions of holomorphic forms, and we define the Petersson inner product.

Definition. (Cf. Rankin [85], § 3.1.) Let Γ be a cofinite Fuchsian group and let k be a real number. An *automorphy factor* of weight k for Γ is a function

$$\nu\colon \Gamma \times \mathbf{H} \to \mathbf{C}^\times$$

satisfying the following conditions:

(1) the map $\gamma \mapsto \nu(\gamma, z)$ is a 1-cocycle with values in the right Γ-module of holomorphic functions $\mathbf{H} \to \mathbf{C}^\times$ (with pointwise multiplication), i.e. the function $z \mapsto \nu(\gamma, z)$ is holomorphic for all $\gamma \in \Gamma$ and

$$\nu(\gamma\delta, z) = \nu(\gamma, \delta z)\nu(\delta, z)$$

for all $\gamma, \delta \in \Gamma$ and $z \in \mathbf{H}$;

(2) for all $\gamma \in \Gamma$ and $z \in \mathbf{H}$, we have

$$|\nu(\gamma, z)| = \left(\frac{\Im z}{\Im \gamma z}\right)^{k/2} \quad \left(= |cz+d|^k \quad \text{if } \gamma = \begin{pmatrix} * & * \\ c & d \end{pmatrix}\right);$$

(3) if $-1 \in \Gamma$, then $\nu(-1, z) = 1$.

Definition. (Cf. Roelcke [91], Definition 1.1.) Let Γ be a cofinite Fuchsian group, let k be a real number, and let ν an automorphy factor of weight k for Γ. An *automorphic form (of Maaß)* of type ν for Γ is a smooth function $f\colon \mathbf{H} \to \mathbf{C}$ with the following properties:

(1) for all $\gamma = \left(\begin{smallmatrix} * & * \\ c & d \end{smallmatrix}\right) \in \Gamma$ and $z \in \mathbf{H}$, the transformation formula

$$f(\gamma z) = \frac{\nu(\gamma, z)}{|cz+d|^k} f(z)$$

holds;

(2) for every cusp $\mathfrak{c}$ of Γ, there is a real number κ such that $|\sigma_{\mathfrak{c}}^* f(x+iy)| = O(y^\kappa)$ as $y \to \infty$.

A *cusp form* of type ν for Γ is a function f satisfying (1) and the following condition (which is stronger than (2)):

(2′) for every cusp $\mathfrak{c}$ of Γ there exists $\epsilon > 0$ such that $|\sigma_{\mathfrak{c}}^* f(x+iy)| = O(\exp(-\epsilon y))$ as $y \to \infty$.

Remark. If we write

$$\tilde{f}(z) = \frac{f(z)}{(\Im z)^{k/2}},$$

then condition (1) in the definition of automorphic forms is equivalent to

$$\tilde{f}(\gamma z) = \nu(\gamma, z)\tilde{f}(z) \quad \text{for all } \gamma \in \Gamma,\ z \in \mathbf{H}.$$

For every cusp $\mathfrak{c}$ of Γ and every automorphic form f of type ν, we define

$$f_{\mathfrak{c}}(z) = \frac{|cz+d|^k}{(cz+d)^k} f(\sigma_{\mathfrak{c}} z),$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is some lift of $\sigma_{\mathfrak{c}}$ to $\mathrm{SL}_2(\mathbf{R})$, so that

$$f(z) = \frac{|-cz+a|^k}{(-cz+a)^k} f_{\mathfrak{c}}(\sigma_{\mathfrak{c}}^{-1} z).$$

The definition of $f_{\mathfrak{c}}$ implies that f comes from a holomorphic form (i.e. the function $(\Im z)^{-k/2} f(z)$ is holomorphic) if and only if $(\Im z)^{-k/2} f_{\mathfrak{c}}(z)$ is holomorphic.

We now assume that the automorphy factor ν is *singular* at the cusp $\mathfrak{c}$, by which we mean the following. We fix a specific branch of the k-th power map by

$$z^k = |z|^k \exp(ik \arg z), \quad -\pi < \arg z \le \pi.$$

We recall from § 1.2 that $\Gamma_{\mathfrak{c}}$ is generated by $\Gamma \cap \{\pm 1\}$ and the element

$$\gamma_{\mathfrak{c}} = \sigma_{\mathfrak{c}} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \sigma_{\mathfrak{c}}^{-1} \in \Gamma_{\mathfrak{c}}.$$

The automorphy factor ν is said to be singular at the cusp $\mathfrak{c}$ if the restriction of ν to $\Gamma_{\mathfrak{c}} \times \mathbf{H}$ is given by the particular formula

$$\nu(\gamma_{\mathfrak{c}}, z) = (cz + d)^k \quad \text{if } \gamma_{\mathfrak{c}} = \begin{pmatrix} * & * \\ c & d \end{pmatrix}.$$

Under the assumption that ν is singular at $\mathfrak{c}$, we have

$$f_{\mathfrak{c}}(z+1) = f_{\mathfrak{c}}(z) \quad \text{for every cusp } \mathfrak{c}.$$

Let us assume furthermore that f, and hence also $f_{\mathfrak{c}}$, is holomorphic. Then $f_{\mathfrak{c}}$ has a q-expansion of the form

$$f_{\mathfrak{c}}(z) = y^{k/2} \sum_{n=0}^{\infty} a_{\mathfrak{c},n}(f) q^n \quad \text{with } q = \exp(2\pi i z).$$

This can be rewritten as

$$f(z) = \frac{|-cz+a|^k}{(-cz+a)^k} y_{\mathfrak{c}}(z)^{k/2} \sum_{n=0}^{\infty} a_{\mathfrak{c},n}(f) q_{\mathfrak{c}}(z)^n, \tag{3.10}$$

where $y_{\mathfrak{c}}(z) = \Im \sigma_{\mathfrak{c}}^{-1} z$ and $q_{\mathfrak{c}} = \exp(2\pi i \sigma_{\mathfrak{c}}^{-1} z)$ as in § 1.2.

If f and g are automorphic forms of type ν for Γ, the function $f\bar{g}$ is Γ-invariant, and hence can be viewed as a function on $\Gamma \backslash \mathbf{H}$. We let $\mathrm{L}^2_\nu(\Gamma \backslash \mathbf{H})$ denote the Hilbert space obtained by completing the space of smooth and bounded automorphic forms with respect to the *Petersson inner product*

$$\langle f, g \rangle = \int_{\Gamma \backslash \mathbf{H}} f \bar{g}\, \mu_{\mathbf{H}}.$$

3.8. Spectral theory for automorphic forms

Let Γ be a cofinite Fuchsian group, and let k be a real number. The *Laplace operator* of weight k is the differential operator

$$\begin{aligned}\Delta_k &= -y^2(\partial_x^2 + \partial_y^2) + iky\partial_x \\ &= \Delta + iky\partial_x\end{aligned}$$

on the space of twice continuously differentiable functions on $\mathbf{H}$. For any automorphy factor ν of weight k, the operator Δ_k can be extended to an (unbounded) self-adjoint operator on a dense subspace of $\mathrm{L}^2_\nu(\Gamma\backslash\mathbf{H})$; see Roelcke [91, § 3].

The spectrum of Δ_k is contained in the interval $[(|k|/2)(1-|k|/2), \infty)$, and the eigenspace corresponding to the eigenvalue $(k/2)(1-k/2)$ is equal to the space of functions $y^{k/2}f$ with f a holomorphic form of type ν that is square-integrable with respect to the Petersson inner product; see Roelcke [91, Sätze 5.2 und 5.5]. In particular, this implies the well-known fact that there are no holomorphic forms of negative weight, since $(|k|/2)(1-|k|/2) > (k/2)(1-k/2)$ if $k < 0$. For $k \geq 1$, the only square-integrable holomorphic forms of weight k are the cusp forms; see Roelcke [92, Satz 13.1]. This means that the map $f \mapsto y^{k/2}f$ gives an identification of the space $\mathrm{S}_\nu(\Gamma)$ of holomorphic cusp forms of type ν with the subspace of $\mathrm{L}^2_\nu(\Gamma\backslash\mathbf{H})$ on which Δ_k acts with eigenvalue $(k/2)(1-k/2)$.

As in the case of automorphic forms of weight 0, the discrete spectrum of Δ_k on $\mathrm{L}^2_\nu(\Gamma\backslash\mathbf{H})$ consists of eigenvalues. Let $\{\lambda_j\}_{j=0}^\infty$ be this discrete spectrum, ordered such that

$$\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots,$$

and let $\{\phi_j^\nu\}_{j=0}^\infty$ be a corresponding orthonormal set of eigenfunctions. We write

$$\lambda_j = s_j(1-s_j) \quad \text{and} \quad s_j = \frac{1}{2} + it_j$$

with $s_j \in [1/2, 1]$ if $\lambda_j \leq 1/4$.

Apart from the discrete spectrum, there is also a continuous spectrum, which can again be described in terms of suitably defined Eisenstein–Maaß series. These functions are defined for the cusps $\mathfrak{c}$ at which ν is singular in the sense of § 3.7. For such a cusp $\mathfrak{c}$, the Eisenstein series $E_\mathfrak{c}^\nu$ is defined for $\Re s > 1$ by

$$E_\mathfrak{c}^\nu(z,s) = (\Im z)^{k/2} \sum_{\gamma\in\Gamma_\mathfrak{c}\backslash\Gamma} \frac{(\Im \sigma_\mathfrak{c}^{-1}\gamma z)^{s-k/2}}{\nu(\gamma, z)(c\gamma z + d)^k};$$

cf. Roelcke [92, § 10] or Hejhal [46, Chapter 9, Definition 5.3]. Here $(c\ d)$ denotes the bottom row of the unique lift of $\sigma_\mathfrak{c}^{-1}$ to $\mathrm{SL}_2(\mathbf{R})$ such that either $c > 0$, or $c = 0$ and $d > 0$. The functions $E_\mathfrak{c}^\nu(z,s)$ can be meromorphically continued in the variable s in the same sense as the Eisenstein–Maaß series $E_\mathfrak{c}(z,s)$ defined in § 3.2, and they satisfy a similar functional equation. For any smooth and bounded function f on $\Gamma\backslash\mathbf{H}$, we define

$$\begin{aligned}\Pi_\mathfrak{c} f\colon [0,\infty) &\to \mathbf{C} \\ t &\mapsto \int_{z\in\Gamma\backslash\mathbf{H}} f(z)\bar{E}_\mathfrak{c}^\nu(z, \tfrac{1}{2}+it)\mu_\mathbf{H}(z).\end{aligned}$$

The generalisation of Theorem 3.1 is now as follows. Every smooth and bounded automorphic form f of type ν has the spectral decomposition

$$f(z) = \sum_{j=0}^{\infty} \langle f, \phi_j^\nu \rangle \phi_j^\nu(z) + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^\infty (\Pi_{\mathfrak{c}} f)(t)\, E_{\mathfrak{c}}^\nu\bigl(z, \tfrac{1}{2} + it\bigr) dt,$$

where $\mathfrak{c}$ runs over the cusps at which ν is singular. Similarly, we have the following generalisation of 3.2. Let $c > \max\{|k|/2, 1\}$, and let $\phi\colon [1,\infty) \to \mathbf{R}$ be a continuous function such that $\phi(u) = O(u^{-c})$ as $u \to \infty$. Then the sum

$$K^\nu(z,w) = \frac{\exp(ik\pi/2)}{\#(\Gamma \cap \{\pm 1\})} \sum_{\gamma \in \Gamma} \nu(\gamma, w) \left(\frac{\Im \gamma w}{\Im w} \right)^{k/2} \left(\frac{|z - \gamma \bar{w}|}{z - \gamma \bar{w}} \right)^k \phi(u(z, \gamma w)) \quad (3.11)$$

converges uniformly on compact subsets of $\mathbf{H} \times \mathbf{H}$; cf. Faddeev [36, Theorem 4.1]. It can be shown that this function satisfies

$$K^\nu(w,z) = \overline{K^\nu(z,w)}$$

and

$$K^\nu(\gamma z, \gamma' w) = \frac{\nu(\gamma, z)(\Im \gamma z)^{k/2}}{(\Im z)^{k/2}} K^\nu(z,w) \frac{(\Im w)^{k/2}}{\nu(\gamma', w)(\Im \gamma' w)^{k/2}}$$

The last equation implies that $K^\nu(z,w)$ defines an invariant integral operator on the space $\mathrm{L}^2_\nu(\Gamma \backslash \mathbf{H})$.

To state the generalisation of Theorem 3.2, we need an analogue of the Selberg–Harish-Chandra transform in higher weights. For this we introduce the following variants of the Legendre functions (see Fay [38, § 1]; note that our k would be $2k$ in Fay's notation):

$$P_{s,k}(u) = \left(\frac{2}{1+u} \right)^s F\left(s-k, s+k; 1; \frac{u-1}{u+1} \right).$$

Let $\phi\colon [1,\infty) \to \mathbf{R}$ be a continuous function satisfying $\phi(u) = O(u^{-\delta})$ for some real number $\delta > \max\{|k|/2, 1-|k|/2\}$. We have the following generalisation of the Selberg–Harish-Chandra transform (see Fay [38, (34)]):

$$h(t) = 2\pi \int_1^\infty \phi(u) P_{1/2+it,k}(u) du.$$

This transform can also be computed as follows (cf. Hejhal [46, pages 385-386]):

$$\begin{aligned} q(w) &= \sqrt{2} \int_{-\infty}^{\infty} \phi(w + v^2) \left[\frac{\sqrt{w+1} + iv}{\sqrt{w+1} - iv} \right]^{k/2} dv, \\ g(r) &= q(\cosh r) \\ h(t) &= 2 \int_0^\infty \cos(rt) g(r) dr. \end{aligned}$$

Its inverse can be computed by means of the formulae

$$\begin{aligned} g(r) &= \frac{1}{\pi}\int_0^\infty \cos(rt)h(t)dt, \\ q(w) &= g(\operatorname{acosh} w) \\ \phi(u) &= -\frac{1}{\pi\sqrt{2}}\int_{-\infty}^\infty q'(u+t^2)\left[\frac{\sqrt{u+1+t^2}-t}{\sqrt{u+1+t^2}+t}\right]^{k/2} dt. \end{aligned}$$

By means of the change of variables $u+t^2 = \cosh r$, the last two equations can be rewritten as

$$\begin{aligned} \phi(u) = -\frac{1}{2^{3/2}\pi(u+1)^{k/2}}\int_{\operatorname{acosh} u}^\infty g'(r)\Big\{\big(\sqrt{\cosh r+1}+\sqrt{\cosh r-u}\big)^k \\ + \big(\sqrt{\cosh r+1}-\sqrt{\cosh r-u}\big)^k\Big\}\frac{dr}{\sqrt{\cosh r-u}}. \end{aligned}$$

We are now in a position to describe the spectral decomposition of a function of the form (3.11). Let h be the Selberg–Harish-Chandra transform of weight k of the function ϕ. Assume h is even and holomorphic on the strip $\{t\in\mathbf{C} \mid |\Im t|<\alpha\}$ for some $\alpha > \max\{(|k|-1)/2, 1/2\}$, and $|h(t)||t|^\beta$ is bounded on this strip for some $\beta > 2$; cf. the conditions (H1) and (H2) of § 1.1. Then K^ν has the spectral representation

$$\begin{aligned} K^\nu(z,w) = \sum_{j=0}^\infty h(t_j)\phi_j^\nu(z)\bar\phi_j^\nu(w) \\ + \sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^\infty h(t)E_{\mathfrak{c}}^\nu\big(z,\tfrac12+it\big)\bar E_{\mathfrak{c}}^\nu\big(w,\tfrac12+it\big)dt, \end{aligned} \tag{3.12}$$

where the second sum is taken over the cusps at which ν is singular; see Hejhal [46, Chapter 8, (4.1), and Chapter 9, § 6].

4. Bounds on cusp forms

Let Γ be a cofinite Fuchsian group, let k be a real number with $k\ge 1$, and let ν be an automorphy factor of weight k for Γ. We assume for simplicity that ν is singular at all cusps of Γ (in the sense of § 3.8). We define a smooth and Γ-invariant function $F_{\Gamma,\nu}$ on $\mathbf{H}$ by

$$F_{\Gamma,\nu}(z) = \sum_{f\in B}(\Im z)^k|f(z)|^2,$$

where B is an orthonormal basis for the space $\mathrm{S}_\nu(\Gamma)$ of holomorphic cusp forms of type ν for Γ. The function $F_{\Gamma,\nu}$ is independent of the choice of B. In this section we give explicit bounds on the values of $F_{\Gamma,\nu}$. These results are due to Jorgenson and Kramer [50]; we have written them down here in a slightly more explicit form.

4.1. The heat kernel for automorphic forms

Following Jorgenson and Kramer [50], we are going to apply the spectral theory described in § 3.8 to the function

$$h_{k,\chi}(t) = \exp\Bigl(-\Bigl(\frac{(k-1)^2}{4} + t^2\Bigr)\chi\Bigr),$$

where χ is a fixed positive real number. This is the spectral function for the heat kernel associated to the operator Δ_k. We compute the kernel function ϕ_χ corresponding to $h_{k,\chi}$ via the auxiliary function g_χ introduced in § 3.8 (the Fourier transform of $h_{k,\chi}$):

$$\begin{aligned} g_\chi(r) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(irt)\exp\Bigl(-\Bigl(\frac{(k-1)^2}{4} + t^2\Bigr)\chi\Bigr)dt \\ &= \frac{1}{2\sqrt{\pi}}\exp\Bigl(-\frac{(k-1)^2\chi}{4} - \frac{r^2}{4\chi}\Bigr). \end{aligned}$$

It now follows from the formula for the inverse Selberg–Harish-Chandra transform given at the end of § 3.8 that

$$\begin{aligned} \phi_\chi(u) = \frac{\exp\bigl(-(k-1)^2\chi/4\bigr)}{4(2\pi\chi)^{3/2}(u+1)^{k/2}} \int_{\operatorname{acosh} u}^{\infty} r\exp\Bigl(-\frac{r^2}{4\chi}\Bigr) \\ \Bigl\{\bigl(\sqrt{\cosh r+1} + \sqrt{\cosh r - u}\bigr)^k + \bigl(\sqrt{\cosh r+1} - \sqrt{\cosh r - u}\bigr)^k\Bigr\}\frac{dr}{\sqrt{\cosh r - u}}. \end{aligned}$$

This does not look very enlightening, but the only property of ϕ_χ that we will need is that it is non-negative. The properties of $h_{k,\chi}$ imply that the function

$$K_\chi^{\Gamma,\nu}(z,w) = \frac{\exp(ik\pi/2)}{\#(\Gamma\cap\{\pm1\})(\Im w)^{k/2}}\sum_{\gamma\in\Gamma}\nu(\gamma,w)(\Im\gamma w)^{k/2}\left(\frac{z-\gamma\bar w}{|z-\gamma\bar w|}\right)^k\phi_\chi(u(z,\gamma w))$$

has a spectral representation given by (3.12). In particular, we have the identity

$$K_\chi^{\Gamma,\nu}(z,z) = \sum_{j=0}^{\infty} h_{k,\chi}(t_j)|\phi_j(z)|^2 + \sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^\infty h_{k,\chi}(t)\bigl|E_{\mathfrak{c}}^\nu\bigl(z,\tfrac12+it\bigr)\bigr|^2. \qquad (4.1)$$

4.2. Bounds on cusp forms

As we saw in § 3.8, the assumption that $k \ge 1$ implies that the space $\mathrm{S}_\nu(\Gamma)$ can be identified with the eigenspace of the Laplace operator Δ_k on $\mathrm{L}^2_\nu(\Gamma)$ associated to the eigenvalue

$$\lambda_{\mathrm{h}} = (k/2)(1-k/2) = \tfrac14 + t_{\mathrm{h}}^2,$$

where

$$t_{\mathrm{h}} = \pm\frac{k-1}{2}i.$$

It is clear from the definition of $h_{k,\chi}$ that

$$h_{k,\chi}(t_{\mathrm{h}}) = 1$$

and that $h_{k,\chi}(t)$ is non-negative for all values of t involved in the spectral representation (4.1). Therefore (4.1) implies the bound

$$F_{\Gamma,\nu}(z) \le K_\chi^{\Gamma,\nu}(z,z) \quad \text{for all } z \in \mathbf{H} \text{ and } \chi > 0.$$

By the triangle inequality, the fact that

$$|\nu(\gamma, w)| = \left(\frac{\Im w}{\Im \gamma w}\right)^{k/2}$$

and the non-negativity of ϕ_χ, we deduce from the above definition of $K_\chi^{\Gamma,\nu}$ the inequality

$$K_\chi^{\Gamma,\nu}(z,z) \le \frac{1}{\#(\Gamma \cap \{\pm 1\})} \sum_{\gamma \in \Gamma} \phi_\chi(u(z, \gamma z)).$$

We now take Γ to be an arbitrary subgroup of finite index in a fixed cofinite Fuchsian group Γ_0. We fix a compact subset Y_0 of $\Gamma_0\backslash\mathbf{H}$, and we let Y_Γ denote the inverse image of Y_0 in $\Gamma\backslash\mathbf{H}$. Since there is an injective map $\Gamma/(\Gamma \cap \{\pm 1\}) \to \Gamma_0/(\Gamma_0 \cap \{\pm 1\})$, we get the inequality

$$\sup_{Y_\Gamma} F_{\Gamma,\nu} \le C(Y_0, k), \tag{4.2}$$

where

$$C(Y_0, k) = \frac{1}{\#(\Gamma_0 \cap \{\pm 1\})} \sup_{w \in Y_0} \sum_{\gamma \in \Gamma_0} \phi_\chi(u(z, \gamma z)).$$

4.3. Extension to neighbourhoods of the cusps

We now take the compact subset Y_0 to be of the following specific form. For every cusp $\mathfrak{c}_0$ of Γ_0, we choose a real number $\epsilon_{\mathfrak{c}_0} > 0$ such that the disc $B_{\mathfrak{c}_0}(\epsilon_{\mathfrak{c}_0})$ of area $\epsilon_{\mathfrak{c}_0}$ around $\mathfrak{c}_0$ as in § 1.2 is well-defined. We define a compact subset Y_0 of $\Gamma_0\backslash\mathbf{H}$ as the complement of the discs $B_{\mathfrak{c}_0}(\epsilon_{\mathfrak{c}_0})$, with $\mathfrak{c}_0$ running over the cusps of Γ_0.

The inverse image in $\Gamma\backslash\mathbf{H}$ of the disc $B_{\mathfrak{c}_0}(\epsilon_{\mathfrak{c}_0}) \subset \Gamma_0\backslash\mathbf{H}$ equals the union of the discs $B_\mathfrak{c}(\epsilon_\mathfrak{c})$, where $\mathfrak{c}$ runs over the cusps of Γ lying over $\mathfrak{c}_0$, and where

$$\epsilon_\mathfrak{c} = m_\mathfrak{c} \epsilon_{\mathfrak{c}_0},$$

with $m_\mathfrak{c}$ the ramification index at $\mathfrak{c}$ of the map from the compactification $\Gamma\backslash\mathbf{H}$ to that of $\Gamma_0\backslash\mathbf{H}$. We write Y_Γ for the inverse image of Y_0 in $\Gamma\backslash\mathbf{H}$; then Y_Γ equals the complement of the discs $B_\mathfrak{c}(\epsilon_\mathfrak{c})$, with $\mathfrak{c}$ running over the cusps of Γ.

Because the forms in our basis B are holomorphic cusp forms, they have q-expansions of the form (3.10) with $a_{\mathfrak{c},0} = 0$. In particular, we see that every $f \in B$ satisfies

$$|f(z)|^2 = y_\mathfrak{c}(z)^k \left| \sum_{n=1}^\infty a_{\mathfrak{c},n}(f) q_\mathfrak{c}(z)^n \right|^2.$$

Therefore the function

$$y_\mathfrak{c}(z)^{-k} \exp(4\pi y_\mathfrak{c}(z)) F_{\Gamma,\nu}(z) = \sum_{f \in B} \left| \frac{f(z)}{q_\mathfrak{c}(z)} \right|^2$$

extends to a subharmonic function on the compactification

$$\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}) = \{z \in \Gamma\backslash\mathbf{H} \mid y_{\mathfrak{c}}(z) \geq 1/\epsilon_{\mathfrak{c}}\} \cup \{\mathfrak{c}\}.$$

of $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$. By the maximum principle for subharmonic functions, the function takes its maximum on the boundary, so that

$$F_{\Gamma,\nu}(z) \leq (\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z))^k \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}}(z)) \sup_{y_{\mathfrak{c}}(z')=1/\epsilon_{\mathfrak{c}}} F_{\Gamma,\nu}(z') \tag{4.3}$$

for all $z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$.

Lemma 4.1. *Let Γ_0 be a cofinite Fuchsian group, and let k be a real number. There is a real number $D(\Gamma_0, k)$ such that for any subgroup Γ of finite index in Γ_0 and any automorphy factor ν of weight k for Γ that is singular at all cusps, we have*

$$\sup_{z\in\Gamma\backslash\mathbf{H}} F_{\Gamma,\nu}(z) \leq (\max_{\mathfrak{c}} m_{\mathfrak{c}})^k D(\Gamma_0, k).$$

Proof. We fix a positive real number χ and define $\phi_\chi\colon [1,\infty) \to \mathbf{R}$ as in § 4.1. We choose $\epsilon > 0$ small enough such that the discs $B_{\mathfrak{c}_0}(\epsilon)$ around the cusps of $\Gamma_0\backslash\mathbf{H}$ are disjoint. We write

$$Y_0 = \Gamma_0\backslash\mathbf{H} \setminus \bigsqcup_{\mathfrak{c}_0} B_{\mathfrak{c}_0}(\epsilon),$$

and we let Y_Γ denote the inverse image of Y_0 in $\Gamma\backslash\mathbf{H}$. An elementary calculation shows that

$$y^k \exp(-4\pi y) \leq \left(\frac{k}{4\pi}\right)^k \exp(-k),$$

with equality if and only if $y = k/4\pi$. Combining this with (4.2) and (4.3) gives

$$F_{\Gamma,\nu}(z) \leq \begin{cases} C(Y_0, k) & \text{if } z \in Y_\Gamma; \\ \left(\frac{k\epsilon_{\mathfrak{c}}}{4\pi}\right)^k \exp(4\pi/\epsilon_{\mathfrak{c}} - k) C(Y_0, k) & \text{if } z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}), \end{cases}$$

which implies the lemma. □

5. Bounds on Green functions of Fuchsian groups

In this section we will apply the spectral theory described in § 3.3 and the solution to the hyperbolic lattice point problem from § 3.5 in order to obtain bounds on Green functions of Fuchsian groups. We start with an elementary "counting lemma" that we will need later.

Lemma 5.1. *Let $\{x_\lambda\}_{\lambda\in\Lambda}$ and $\{w_\lambda\}_{\lambda\in\Lambda}$ be two families of real numbers, indexed by a set Λ, such that for every $x \in \mathbf{R}$ there are only finitely many $\lambda \in \Lambda$ with $x_\lambda \le x$. (In particular, Λ is countable.) For all $x \in \mathbf{R}$ we define*

$$W(x) = \sum_{\lambda\in\Lambda\colon x_\lambda\le x} w_\lambda.$$

Suppose a is a real number such that $x_\lambda > a$ for all $\lambda \in \Lambda$, and $A, B\colon (a,\infty) \to \mathbf{R}$ are continuous functions such that $A(x) \le W(x) \le B(x)$ for all $x \ge a$. Let $f\colon (a,\infty) \to [0,\infty)$ be a decreasing, continuously differentiable function such that the sum

$$S = \sum_{\lambda\in\Lambda} f(x_\lambda) w_\lambda$$

converges absolutely. Then the inequality

$$-\int_a^\infty f'(x)A(x)dx + \lim_{x\to\infty} f(x)A(x) \le S \le -\int_a^\infty f'(x)B(x)dx + \lim_{x\to\infty} f(x)B(x),$$

holds, provided the integrals and limits exist. If in addition A and B are piecewise continuously differentiable, then S satisfies the inequality

$$\int_a^\infty f(x)A'(x)dx + \lim_{x\searrow a} f(x)A(x) \le S \le \int_a^\infty f(x)B'(x)dx + \lim_{x\searrow a} f(x)B(x),$$

again provided the integrals and limits exist.

Proof. By assumption, the subset $\{x_\lambda \mid \lambda \in \Lambda\}$ of (a,∞) is discrete. We write y_1, y_2, ..., for its elements in increasing order, and we put $y_0 = a$. Using the absolute convergence of the sum S, we can rewrite it as

$$\begin{aligned} S &= \sum_{i=1}^\infty f(y_i) \sum_{\lambda\in\Lambda\colon x_\lambda=y_i} w_\lambda \\ &= \lim_{I\to\infty}\left(\sum_{i=1}^I f(y_i)(W(y_i) - W(y_{i-1}))\right) \\ &= \lim_{I\to\infty}\left(\sum_{i=1}^{I-1}(f(y_i) - f(y_{i+1}))W(y_i) + f(y_I)W(y_I)\right); \end{aligned}$$

the last equality is gotten by partial summation and the fact that $W(y_0) = 0$. Because W is constant on each $[y_i, y_{i+1})$ and zero on (y_0, y_1), we may rewrite this as

$$S = \lim_{I\to\infty}\left(-\sum_{i=0}^{I-1}\int_{y_i}^{y_{i+1}} f'(x)W(x)dx + f(y_I)W(y_I)\right).$$

Together with the inequality $A(x) \le W(x) \le B(x)$ and the assumption that f is decreasing and non-negative, we now get

$$\lim_{I\to\infty}\biggl(-\int_{y_0}^{y_I} f'(x)A(x)dx + f(y_I)A(y_I)\biggr) \le S$$

$$\le \lim_{I\to\infty}\biggl(-\int_{y_0}^{y_I} f'(x)B(x)dx + f(y_I)B(y_I)\biggr),$$

from which the first inequality follows. If A and B are piecewise continuously differentiable, the second is equivalent to the first via integration by parts. □

5.1. A construction of the Green function

We will now give a construction of the Green function for Γ that will allow us to find explicit bounds on its values. For this we use a family of auxiliary functions

$$k_a\colon (1,\infty) \to [0,\infty) \quad (a \in A)$$

parametrised by a filtered set A, that converges in a suitable sense (made precise in Lemma 5.3 below) to the function k_1 defined by

$$k_1(u) = \frac{1}{4\pi}\log\frac{u+1}{u-1}. \tag{5.1}$$

We will take the k_a such that their Selberg–Harish-Chandra transforms are of the kind described in the following definition.

Definition. An *admissible spectral function* is an even and holomorphic function

$$h\colon D \to \mathbf{C}$$

where D is an open subset of $\mathbf{C}$ containing the strip $\bigl\{t \in \mathbf{C} \bigm| |\Im t| \le \alpha\bigr\}$ for some $\alpha > 1/2$, such that for some $\beta > 1$ the function

$$\bigl|\tfrac{1}{4}+t^2\bigr|^\beta \left|h(t) - \frac{1}{\frac{1}{4}+t^2}\right| = \bigl|\tfrac{1}{4}+t^2\bigr|^{\beta-1}\bigl|\bigl(\tfrac{1}{4}+t^2\bigr)h(t) - 1\bigr|$$

is bounded on this strip.

Lemma 5.2. *Let D be an open subset of $\mathbf{C}$ containing the strip $\bigl\{t \in \mathbf{C} \bigm| |\Im t| \le \alpha\bigr\}$ for some $\alpha > 1/2$, and let $h\colon D \to \mathbf{C}$ be an admissible spectral function. Then h satisfies the conditions (H1) and (H2) of § 1.1; in particular, the inverse Selberg–Harish-Chandra transform k of h (see (1.4)) exists. Moreover, $u^{\alpha+1/2}k(u)$ is bounded as $u \to \infty$.*

Proof. The claim that h satisfies (H1) and (H2) is straightforward to check. To compute k, we use the formulae from the end of § 1.1 relating k and h via the intermediate function

$$g(r) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(irt)h(t)dt.$$

Since h is even, we may assume $r > 0$. Because h is holomorphic, we may shift the line of integration to $\mathbf{R} + i\alpha$; this gives

$$g(r) = \frac{\exp(-\alpha r)}{2\pi} \int_{-\infty}^{\infty} \exp(irt) h(t + i\alpha) dt.$$

Next we note that

$$\int_{-\infty}^{\infty} \frac{\exp(irt)}{\frac{1}{4} + (t + i\alpha)^2} dt = 0$$

(since the integrand is holomorphic for $\Im t > 0$, we may integrate over the semi-circle $\{i\alpha + R\exp(i\theta) \mid 0 \le \theta \le \pi\}$ and let R tend to ∞; it is not hard to show that this integral tends to 0). This implies that

$$g(r) = \frac{\exp(-\alpha r)}{2\pi} \int_{-\infty}^{\infty} \exp(irt) \left(h(t + i\alpha) - \frac{1}{\frac{1}{4} + (t + i\alpha)^2} \right) dt.$$

Likewise, we have

$$g'(r) = \frac{i\exp(-\alpha r)}{2\pi} \int_{-\infty}^{\infty} \exp(irt)(t + i\alpha) \left(h(t + i\alpha) - \frac{1}{\frac{1}{4} + (t + i\alpha)^2} \right) dt.$$

Applying the triangle inequality and the assumption on h, we deduce from this that

$$|g(r)| \le D\exp(-\alpha r) \quad \text{and} \quad |g'(r)| \le D'\exp(-\alpha r),$$

where D and D' are certain positive real numbers. The formula

$$k(u) = -\frac{1}{\pi\sqrt{2}} \int_{\operatorname{acosh} u}^{\infty} \frac{g'(r)dr}{\sqrt{\cosh r - u}}$$

now implies that

$$\begin{aligned} |k(u)| &\le \frac{D'}{\pi\sqrt{2}} \int_{\operatorname{acosh} u}^{\infty} \frac{\exp(-\alpha r)dr}{\sqrt{\cosh r - u}} \\ &= \frac{D'}{\pi} Q_{\alpha-1/2}(u), \end{aligned}$$

where Q_ν denotes the Legendre function of the second kind of degree ν; see Erdélyi et al. [34, § 3.6.1 and § 3.7, equation 4]. Since $Q_{\alpha-1/2}(u) = O(u^{-1/2-\alpha})$ as $u \to \infty$ (see [34, § 3.92, equation 21]), this proves the claim. □

Let $h\colon D \to \mathbf{C}$ be an admissible spectral function as defined above, and let k be its inverse Selberg–Harish-Chandra transform. It follows from Lemma 5.2 that the sum $\sum_{\gamma\in\Gamma} k(u(z, \gamma w))$ converges uniformly on compact subsets of $\mathbf{H} \times \mathbf{H}$ not containing any points of the form $(z, \gamma z)$; see Faddeev [36, pages 363–364 of the English translation]. (See also Lang [62, Chapter XIV], which is an explanation of [36], filling in many details.) We can therefore define a continuous, symmetric function

$$\begin{aligned} K^\Gamma\colon \{(z, w) \in \mathbf{H} \times \mathbf{H} \mid z \notin \Gamma w\} &\longrightarrow \mathbf{R} \\ (z, w) &\longmapsto \frac{1}{\#(\Gamma \cap \{\pm 1\})} \sum_{\gamma\in\Gamma} k(u(z, \gamma w)) - c, \end{aligned} \tag{5.2}$$

where the constant c is chosen such that the integral of K^Γ over $\Gamma\backslash\mathbf{H}$ with respect to each of the variables vanishes:

$$\begin{aligned}
c &= \mathrm{vol}_\Gamma^{-1} \int_{w\in\mathbf{H}} k(u(z,w))\mu_{\mathbf{H}}(w) \quad \text{for any } z\in\mathbf{H} \\
&= \frac{2\pi}{\mathrm{vol}_\Gamma}\int_1^\infty k(u)du \\
&= \mathrm{vol}_\Gamma^{-1}\, h(\pm i/2).
\end{aligned}$$

Note that the last equality is just (1.2).

Lemma 5.3. *Consider a family $\{h_a\}_{a\in A}$ of admissible spectral functions, with A a filtered set, and let $\{k_a\}_{a\in A}$ and $\{K_a^\Gamma\}_{a\in A}$ be the corresponding functions defined by (1.4) and (5.2). Suppose that the following two conditions are satisfied:*

(1) Each of the functions k_a is bounded pointwise from above by the function k_1 defined by (5.1), and the family of functions $\{k_a\}_{a\in A}$ converges pointwise to k_1.

(2) There is a real number $\beta > 1$ such that the family of functions

$$\left|\tfrac{1}{4}+t^2\right|^\beta \left|h_a(t) - \frac{1}{\frac{1}{4}+t^2}\right|$$

converges to 0 (with respect to the filtration of the set A), uniformly on the strip $\bigl\{t\in\mathbf{C} \bigm| |\Im t| \le 1/2\bigr\}$.

Then the family of functions $\{-K_a^\Gamma\}_{a\in A}$ converges to the Green function gr_Γ in the $(\mathrm{L}^\infty_{\mathrm{loc}}, \mathrm{L}^2\cap\mathrm{L}^\infty_{\mathrm{loc}})$-topology.

(The existence of families of admissible spectral functions $\{h_a\}_{a\in A}$ satisfying the above conditions will be proved in § 5.2 below.)

Proof of Lemma 5.3. For all $a, b\in A$, it follows from condition (1) that the function $k_a - k_b$ satisfies the conditions of Theorem 3.2. This implies that the function

$$K_a^\Gamma - K_b^\Gamma = \frac{1}{\#(\Gamma\cap\{\pm1\})}\sum_{\gamma\in\Gamma}\bigl(k_a(u(z,\gamma w)) - k_b(u(z,\gamma w))\bigr) - c_a + c_b$$

has the spectral representation

$$\begin{aligned}
(K_a^\Gamma - K_b^\Gamma)(z,w) &= \sum_{j=1}^\infty (h_a(t_j) - h_b(t_j))\phi_j(z)\bar\phi_j(w) \\
&\quad + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^\infty (h_a(t) - h_b(t)) E_{\mathfrak{c}}\bigl(z, \tfrac12 + it\bigr)\bar E_{\mathfrak{c}}\bigl(w, \tfrac12+it\bigr)dt,
\end{aligned} \tag{5.3}$$

where the right-hand side converges to $K_a^\Gamma - K_b^\Gamma$ in the $(\mathrm{L}^\infty_{\mathrm{loc}}, \mathrm{L}^2\cap\mathrm{L}^\infty_{\mathrm{loc}})$-topology. (Note that the eigenvalue $\lambda_0 = 0$ has disappeared because of the definition of c_a.) In particular, $K_a^\Gamma - K_b^\Gamma$ extends to a continuous function on $\mathbf{H}\times\mathbf{H}$ that is Γ-invariant with respect to both variables.

We claim that the right-hand side of (5.3) converges to 0, with respect to A, in the $(\mathrm{L}^\infty_{\mathrm{loc}}, \mathrm{L}^2 \cap \mathrm{L}^\infty_{\mathrm{loc}})$-topology. In particular, this implies that $\{K_a^\Gamma\}_{a\in A}$ converges to a symmetric continuous function on $\Gamma\backslash\mathbf{H}\times\Gamma\backslash\mathbf{H}$ that is square-integrable with respect to each variable separately.

First we show that $\{K_a^\Gamma - K_b^\Gamma\}_{a,b\in A}$ converges to zero uniformly on compact subsets of $\mathbf{H}\times\mathbf{H}$. For this we write

$$\begin{aligned}\bigl|K_a^\Gamma - K_b^\Gamma\bigr|(z,w) \le \sum_{j=1}^\infty |h_a(t_j) - h_b(t_j)|\cdot|\phi_j(z)\bar\phi_j(w)|\\ + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^\infty |h_a(t) - h_b(t)|\cdot\bigl|E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)\bigr|dt\end{aligned}$$

It follows from the triangle inequality and condition (2) on the family $\{h_a\}$ that

$$\begin{aligned}|h_a(t) - h_b(t)| &\le \left|h_a(t) - \frac{1}{\frac14+t^2}\right| + \left|h_b(t) - \frac{1}{\frac14+t^2}\right|\\ &\le (C_a + C_b)\bigl|\tfrac14+t^2\bigr|^{-\beta}\end{aligned}$$

for some family of positive real numbers $\{C_a\}_{a\in A}$ such that $\lim_{a\in A} C_a = 0$. This implies

$$\begin{aligned}\bigl|K_a^\Gamma - K_b^\Gamma\bigr|(z,w) \le (C_a+C_b)\sum_{j=1}^\infty\bigl|\tfrac14+t_j^2\bigr|^{-\beta}|\phi_j(z)\bar\phi_j(w)|\\ + (C_a+C_b)\frac{1}{2\pi}\sum_{\mathfrak{c}}\int_0^\infty\bigl|\tfrac14+t^2\bigr|^{-\beta}\bigl|E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)\bigr|dt.\end{aligned}$$

Using the elementary inequality

$$|\phi_j(z)\bar\phi_j(w)| \le \tfrac12\bigl(|\phi_j(z)|^2 + |\phi_j(w)|^2\bigr), \tag{5.4}$$

and applying Lemma 5.1 and Corollary 3.5, we see that the right-hand side converges to 0 uniformly on compacta of $\mathbf{H}\times\mathbf{H}$, as claimed.

Next we show that the right-hand side of (5.3) converges with respect to the L^2-norm on $\Gamma\backslash\mathbf{H}$, uniformly for w in compacta of $\mathbf{H}$. For this we use that the orthogonality of eigenfunctions implies

$$\begin{aligned}\int_{w\in\Gamma\backslash\mathbf{H}}\bigl|K_a^\Gamma - K_b^\Gamma\bigr|^2(z,w)\mu_{\mathbf{H}}(w) &\le \sum_{j=1}^\infty\bigl|h_a(t_j) - h_b(t_j)\bigr|^2|\phi_j(z)|^2\\ &\quad + \sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^\infty\bigl|h_a(t) - h_b(t)\bigr|^2\bigl|E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bigr|^2dt\\ &\le (C_a+C_b)\sum_{j=1}^\infty\bigl|\tfrac14+t_j^2\bigr|^{-2\beta}|\phi_j(z)|^2\\ &\quad + \frac{C_a+C_b}{2\pi}\sum_{\mathfrak{c}}\int_0^\infty\bigl|\tfrac14+t^2\bigr|^{-2\beta}\bigl|E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bigr|^2dt.\end{aligned}$$

Again using Lemma 5.1 and Corollary 3.5, we see that the right-hand side converges to 0 uniformly on compacta of $\mathbf{H}$, which is what we had to prove.

The L^2-convergence that we just proved implies that if f is a smooth, bounded, Γ-invariant function on $\mathbf{H}$, with spectral representation

$$f(z) = \sum_{j=0}^{\infty} b_j \phi_j(z) + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\infty} b_{\mathfrak{c}}(t) E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt$$

(see Theorem 3.1), then

$$\int\limits_{w \in \Gamma\backslash\mathbf{H}} \lim_{a \in A} K_a^{\Gamma}(z, w) f(w) \mu_{\mathbf{H}}(w) = \lim_{a \in A} \int\limits_{w \in \Gamma\backslash\mathbf{H}} K_a^{\Gamma}(z, w) f(w) \mu_{\mathbf{H}}(w).$$

Now the defining property (1.1) of the Selberg–Harish-Chandra transform implies that

$$\int_{\Gamma\backslash\mathbf{H}} K_a^{\Gamma}(z, w) f(w) \mu_{\mathbf{H}}(w) = \sum_{j=1}^{\infty} b_j h_a(t_j) \phi_j(z) + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\infty} b_{\mathfrak{c}}(t) h_a(t) E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt.$$

Taking the limit, we get

$$\begin{aligned} \int\limits_{w \in \Gamma\backslash\mathbf{H}} \lim_{a \in A} K_a^{\Gamma}(z, w) f(w) \mu_{\mathbf{H}}(w) &= \sum_{j=1}^{\infty} \frac{b_j}{\frac{1}{4} + t_j^2} \phi_j(z) \\ &\quad + \sum_{\mathfrak{c}} \frac{1}{2\pi} \int_0^{\infty} \frac{b_{\mathfrak{c}}(t)}{\frac{1}{4} + t^2} E_{\mathfrak{c}}\bigl(z, \tfrac{1}{2} + it\bigr) dt \\ &= Rf(z) \\ &= -\int\limits_{w \in \Gamma\backslash\mathbf{H}} \mathrm{gr}_{\Gamma}(z, w) f(w) \mu_{\mathbf{H}}(w). \end{aligned}$$

Since the set of smooth and bounded functions is dense in $L^2(\Gamma\backslash\mathbf{H})$, this proves that the limit of the convergent family of functions $\{K_a^{\Gamma}\}_{a \in A}$ equals $-\mathrm{gr}_{\Gamma}$. □

5.2. Existence of families of admissible spectral functions

Of course, the construction given in § 5.1 would be futile if there were no family of functions $\{h_a\}$ fulfilling the conditions of Lemma 5.3. Let us therefore give two examples of such families.

The resolvent kernel for a parameter $a \searrow 1$. This is the function

$$k_a^{\mathrm{R}}(u) = \frac{1}{2\pi} Q_{a-1}(u),$$

where Q_ν is the Legendre function of the second kind of degree ν; see Erdélyi et al. [34, § 3.6.1]. The function Q_{a-1} has the integral representation [34, § 3.7, equation 12]

$$Q_{a-1}(u) = \int_0^{\infty} \frac{dt}{(u + \sqrt{u^2 - 1} \cosh t)^a},$$

which shows that all the Q_{a-1} with $a \geq 1$ are bounded pointwise by the function

$$Q_0(u) = \frac{1}{2} \log \frac{u+1}{u-1}.$$

(see [34, § 3.6.2, equation 20]). This shows that the family $\{k_a\}$ satisfies condition (1) of Lemma 5.3. From [34, § 3.12, equation 4] we see that the Selberg–Harish-Chandra transform of k_a^{R} is

$$\begin{aligned} h_a^{\mathrm{R}}(t) &= \int_1^\infty P_{-1/2+it}(u) Q_{a-1}(u) du \\ &= \frac{1}{(a-1/2-it)(a-1/2+it)} \\ &= \frac{1}{a(a-1)+\frac{1}{4}+t^2}. \end{aligned}$$

One can check easily that this is an admissible spectral function and that the family $\{h_a^{\mathrm{R}}\}$ satisfies condition (2) of Lemma 5.3.

The cumulative heat kernel for a parameter $T \to \infty$. The function h_T^{C} is defined for $T > 0$ by

$$\begin{aligned} h_T^{\mathrm{C}}(t) &= \int_0^T \exp(-(1/4+t^2)\chi) d\chi \\ &= \frac{1-\exp(-(1/4+t^2)T)}{1/4+t^2}. \end{aligned}$$

It is straightforward to check that these are admissible spectral functions and that the family $\{h_T^{\mathrm{C}}\}$ satisfies condition (2) of Lemma 5.3. We compute the corresponding function k_T^{C} using an intermediate function g_T^{C} as in § 1.1. This function is given by

$$\begin{aligned} g_T^{\mathrm{C}}(r) &= \frac{1}{\pi} \int_0^\infty \frac{\cos rt}{1/4+t^2} dt - \frac{\exp(-T/4)}{\pi} \int_0^\infty \frac{\cos(rt)\exp(-t^2 T)}{1/4+t^2} dt \\ &= \exp(-|r|/2) - \frac{1}{2}\left[\exp(r/2)\operatorname{erfc}\left(\frac{T+r}{2\sqrt{T}}\right) + \exp(-r/2)\operatorname{erfc}\left(\frac{T-r}{2\sqrt{T}}\right)\right]. \end{aligned}$$

The last equality follows from Erdélyi et al. [35, § 1.2, equation 11, and § 1.4, equation 15]; the *complementary error function* appearing in this formula is defined by

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt.$$

A straightforward computation gives

$$\begin{aligned} (g_T^{\mathrm{C}})'(r) = &-\frac{\operatorname{sgn}(r)\exp(-|r|/2)}{2} \\ &- \frac{1}{4}\left[\exp(r/2)\operatorname{erfc}\left(\frac{T+r}{2\sqrt{T}}\right) - \exp(-r/2)\operatorname{erfc}\left(\frac{T-r}{2\sqrt{T}}\right)\right]. \end{aligned}$$

By differentiating with respect to T, we see that the family of functions $\{g_T'(r)\}_{T>0}$ is pointwise decreasing as $T \to \infty$, with limit function

$$\lim_{T\to\infty} (g_T^{\mathrm{C}})'(r) = -\frac{\operatorname{sgn}(r)\exp(-|r|/2)}{2}.$$

This implies that the family of functions $\{k_T^{\mathrm{C}}\}_{T>0}$ given by

$$k_T^{\mathrm{C}}(u) = -\frac{1}{\pi\sqrt{2}}\int_{\operatorname{acosh} u}^{\infty} \frac{(g_T^{\mathrm{C}})'(r)dr}{\sqrt{\cosh r - u}}$$

is pointwise increasing with limit function

$$\lim_{T\to\infty} k_T^{\mathrm{C}}(u) = \frac{1}{2^{3/2}\pi}\int_{\operatorname{acosh} u}^{\infty} \frac{\exp(-r/2)dr}{\sqrt{\cosh r - u}}$$

This last integral can be evaluated using Erdélyi et al. [34, § 3.7, equation 4, and § 3.6.2, equation 20]:

$$\begin{aligned}\lim_{T\to\infty} k_T^{\mathrm{C}}(u) &= \frac{1}{2\pi} Q_0(u)\\ &= \frac{1}{4\pi}\log\frac{u+1}{u-1}.\end{aligned}$$

We conclude that the family $\{k_T^{\mathrm{C}}\}$ satisfies condition (1) of Lemma 5.3.

5.3. Bounds on Green functions

Let $\{h_a\}_{a\in A}$, $\{k_a\}_{a\in A}$ and $\{K_a^\Gamma\}_{a\in A}$ be families of functions satisfying the conditions of Lemma 5.3. We will give bounds on the values taken by the functions K_a^Γ and by the Green function gr_Γ. For this we will exploit the estimates for the hyperbolic lattice point problem given in § 3.5. Given two points $z, w \in \mathbf{H}$, we choose a real number $\delta > 1$, and we split the sum over Γ into sums over the two subsets $\Pi(z,w)$ and $\Lambda(z,w)$ consisting of those γ for which $u(z,\gamma w) \le \delta$ and $u(z,\gamma w) > \delta$, respectively, i.e.

$$\begin{aligned}\Pi(z,w) &= \{\gamma\in\Gamma \mid u(z,\gamma w)\le\delta\},\\ \Lambda(z,w) &= \{\gamma\in\Gamma \mid u(z,\gamma w)>\delta\}.\end{aligned}$$

For any $U \ge \delta$, the inequality (3.4) implies that the number of elements $\gamma \in \Lambda(z,w)$ with $u(z,\gamma w) \le U$ can be bounded as

$$A(U) \le \#\{\gamma\in\Lambda(z,w) \mid u(z,\gamma w)\le U\} \le B(U),$$

where the functions

$$A, B\colon [\delta,\infty) \to \mathbf{R}$$

are defined by

$$A(U) = K_U^-(z,w) - \frac{\#\Pi(z,w)}{\#(\Gamma\cap\{\pm1\})} \quad\text{and}\quad B(U) = K_U^+(z,w) - \frac{\#\Pi(z,w)}{\#(\Gamma\cap\{\pm1\})}.$$

Here the functions $K_U^{\pm}$ are defined as in § 3.5 via functions T and V of U satisfying (3.6) and (3.8). The functions A and B are increasing and piecewise continuously differentiable, and the estimates from § 3.5 imply that

$$A(U) = O(U) \quad \text{and} \quad B(U) = O(U) \quad \text{as } U \to \infty,$$

with implied constants depending on the group Γ, the points z and w and the functions T and V. Applying Lemma 5.1 gives

$$\begin{aligned}\int_\delta^\infty k_a(U)A'(U)dU + k_a(\delta)A(\delta) &\le \sum_{\gamma\in\Lambda(z,w)} k_a(u(z,\gamma w)) \\ &\le \int_\delta^\infty k_a(U)B'(U)dU + k_a(\delta)B(\delta).\end{aligned}$$

for all a. By the definition (5.2) of K_a^Γ, we get the lower bound

$$\begin{aligned}K_a^\Gamma(z,w) \ge &-\frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)} k_a(u(z,\gamma w)) - \int_\delta^\infty k_a(U)B'(U)dU - k_a(\delta)B(\delta) \\ &+ \frac{2\pi}{\mathrm{vol}_\Gamma}\int_1^\infty k_a(u)du.\end{aligned}$$

Using the definition of B, we can rewrite this as

$$\begin{aligned}K_a^\Gamma(z,w) \ge &\frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)} \bigl(k_a(\delta) - k_a(u(z,\gamma w))\bigr) - k_a(\delta)K_\delta^+(z,w) \\ &- \int_\delta^\infty k_a(U)\frac{d}{dU}\left(K_U^+(z,w) - \frac{2\pi}{\mathrm{vol}_\Gamma}(U-1)\right)dU + \frac{2\pi}{\mathrm{vol}_\Gamma}\int_1^\delta k_a(U)dU.\end{aligned}$$

We recall from (3.9) that $d/dU(\ldots) = O(U^{-\epsilon})$ for some $\epsilon > 0$. Furthermore, the family of functions $\{k_a\}_{a\in A}$ converges from below to k_1 because of condition (1) in Lemma 5.3. We may therefore apply the dominated convergence theorem and take the limit inside the integral. This gives

$$\begin{aligned}\mathrm{gr}_\Gamma(z,w) \ge &\frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)} \bigl(k_1(\delta) - k_1(u(z,\gamma w))\bigr) - k_1(\delta)K_\delta^+(z,w) \\ &- \int_\delta^\infty k_1(U)\frac{d}{dU}\left(K_U^+(z,w) - \frac{2\pi}{\mathrm{vol}_\Gamma}(U-1)\right)dU + \frac{2\pi}{\mathrm{vol}_\Gamma}\int_1^\delta k_1(U)dU.\end{aligned}$$

Integrating by parts and using that

$$k_1'(u) = -\frac{1}{2\pi(u^2-1)},$$

we can simplify this to

$$\begin{aligned}\mathrm{gr}_\Gamma(z,w) \ge &\frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)} \bigl(k_1(\delta) - k_1(u(z,\gamma w))\bigr) + \frac{1}{\mathrm{vol}_\Gamma}\log\frac{\delta+1}{2} \\ &- \frac{1}{2\pi}\int_\delta^\infty\left(K_U^+(z,w) - \frac{2\pi}{\mathrm{vol}_\Gamma}(U-1)\right)\frac{dU}{U^2-1}.\end{aligned}$$

Next we insert the spectral representation (3.5) of K_U^+, the formula (3.7) for $h_U^+(\pm i/2)$ and the fact that $|\phi_0|^2 = 1/\operatorname{vol}_\Gamma$. We then interchange the resulting sums and integrals with the integral over U; this is permitted because the double sums and integrals converge absolutely (this follows from Lemma 3.4 and Theorem 3.6). The result is

$$\begin{aligned}\operatorname{gr}_\Gamma(z,w) \geq \frac{1}{\#(\Gamma\cap\{\pm 1\})} &\sum_{\gamma\in\Pi(z,w)} \bigl(k_1(\delta) - k_1(u(z,\gamma w))\bigr) + \frac{1}{\operatorname{vol}_\Gamma}\log\frac{\delta+1}{2} \\ &- \frac{1}{2\operatorname{vol}_\Gamma}\int_\delta^\infty \frac{V-U}{U^2-1}dU - \sum_{j=1}^\infty I_\delta^+(t_j)\phi_j(z)\bar\phi_j(w) \\ &- \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^\infty I_\delta^+(t) E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt,\end{aligned}$$

where I_δ^+ is the function defined by

$$I_\delta^+(t) = \frac{1}{2\pi}\int_\delta^\infty \frac{h_U^+(t)}{U^2-1}dU.$$

A similar calculation gives the upper bound

$$\begin{aligned}\operatorname{gr}_\Gamma(z,w) \leq \frac{1}{\#(\Gamma\cap\{\pm 1\})} &\sum_{\gamma\in\Pi(z,w)} \bigl(k_1(\delta) - k_1(u(z,\gamma w))\bigr) + \frac{1}{\operatorname{vol}_\Gamma}\log\frac{\delta+1}{2} \\ &+ \frac{1}{2\operatorname{vol}_\Gamma}\int_\delta^\infty \frac{U-T}{U^2-1}dU - \sum_{j=1}^\infty I_\delta^-(t_j)\phi_j(z)\bar\phi_j(w) \\ &- \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^\infty I_\delta^-(t) E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt,\end{aligned}$$

where I_δ^- is defined by

$$I_\delta^-(t) = \frac{1}{2\pi}\int_\delta^\infty \frac{h_U^-(t)}{U^2-1}dU.$$

The most interesting aspect of the above bounds concerns the functions

$$R^\pm(z,w) = \sum_{j=1}^\infty I_\delta^\pm(t_j)\phi_j(z)\bar\phi_j(w) + \sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^\infty I_\delta^\pm(t)E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt.$$

Imprecisely speaking, these reflect the fact that the Green function *formally* has the spectral representation

$$\operatorname{gr}_\Gamma(z,w) \stackrel{?}{=} -\sum_{j=1}^\infty \frac{1}{\frac14+t_j^2}\phi_j(z)\bar\phi_j(w) - \sum_{\mathfrak{c}}\frac{1}{2\pi}\int_0^\infty \frac{1}{\frac14+t^2}E_{\mathfrak{c}}\bigl(z,\tfrac12+it\bigr)\bar E_{\mathfrak{c}}\bigl(w,\tfrac12+it\bigr)dt.$$

The problem is that this expansion does not converge. This is the reason why our estimates are somewhat complicated; however, it is not very surprising that a similar expression appears in the above bounds.

5.4. Uniform bounds on compact subsets

The next step is to derive bounds that are valid at the same time for all subgroups Γ of finite index in a given cofinite Fuchsian group Γ_0 such that the non-zero eigenvalues of the Laplace operator Δ on $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$ are bounded from below by a fixed positive number. For example, take $\Gamma_0 = \mathrm{SL}_2(\mathbf{Z})$ and let $\Gamma \subseteq \Gamma_0$ be a congruence subgroup. Selberg conjectured in [96] that the least non-zero eigenvalue λ_1 of Δ is at least $1/4$, and he proved that $\lambda_1 \geq 3/16$. The sharpest result known so far, due to Kim and Sarnak (see Appendix 2 of Kim [59]), is that $\lambda_1 \geq 975/4096$.

We seek upper bounds for the absolute values of the functions $R^{\pm}(z,w)$ occurring in the bounds from § 5.3. Applying the triangle inequality and the inequality (5.4), we see that

$$\left|R^{\pm}(z,w)\right| \leq \frac{1}{2}(S^{\pm}(z) + S^{\pm}(w)),$$

where S^+ and S^- are defined by

$$S^{\pm}(z) = \sum_{j=1}^{\infty} |I_\delta^{\pm}(t_j)||\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^\infty |I_\delta^{\pm}(t)|\left|E_{\mathfrak{c}}\left(z, \tfrac{1}{2}+it\right)\right|^2 dt.$$

In order to bound these functions, we use the assumption that the spectrum of the Laplace operator on $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$ is contained in $\{0\}\cup[\lambda_{\min},\infty)$. We choose decreasing, continuously differentiable functions

$$H^+_{\delta,\lambda_{\min}}, H^-_{\delta,\lambda_{\min}} \colon [\lambda_{\min},\infty) \to (0,\infty)$$

such that

(1) $H^{\pm}_{\delta,\lambda_{\min}}(\lambda) = \sup_{[\lambda_{\min},\infty)}\left|I_\delta^{\pm}\left(\sqrt{\lambda - 1/4}\right)\right|$ for all $\lambda \in [\lambda_{\min}, 1/4]$;

(2) $H^{\pm}_{\delta,\lambda_{\min}}(\lambda) \geq \left|I_\delta^{\pm}\left(\sqrt{\lambda - 1/4}\right)\right|$ for all $\lambda \geq 1/4$.

Using the properties (1) and (2) of $H^{\pm}_{\delta,\lambda_{\min}}(1/4 + t^2)$ and rewriting the result in a similar way as in Lemma 5.1 gives

$$S^{\pm}(z) \leq -\int_{\lambda=1/4}^{\infty} W_z(\lambda)\, dH^{\pm}_{\delta,\lambda_{\min}}(\lambda),$$

where

$$W_z(\lambda) = \sum_{j\colon 0<\lambda_j\leq\lambda} |\phi_j(z)|^2 + \sum_{\mathfrak{c}} \frac{1}{2\pi}\int_0^{\sqrt{\lambda-1/4}} \left|E_{\mathfrak{c}}\left(z, \tfrac{1}{4}+t^2\right)\right|^2 dt.$$

We now assume Γ is a subgroup of finite index in a fixed Fuchsian group Γ_0. Let $Y_0 \subset \Gamma_0\backslash\mathbf{H}$ be any compact subset, and let Y be its inverse image in $\Gamma\backslash\mathbf{H}$. Then it follows from the bound from Corollary 3.5 and the fact that the functions $H^{\pm}_{\delta,\lambda_{\min}}$ are decreasing that

$$S^{\pm}(z) \leq -\int_{\lambda=1/4}^{\infty} \nu(Y_0,\lambda)\, dH^{\pm}_{\delta,\lambda_{\min}}(\lambda) \quad \text{for all } z \in Y.$$

Substituting this in the bounds from § 5.3, we see that for all $z, w \in \mathbf{H}$ whose images in $\Gamma\backslash\mathbf{H}$ lie in Y, we have

$$\begin{aligned}\mathrm{gr}_\Gamma(z,w) \le{}& \frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)}\bigl(k_1(\delta)-k_1(u(z,\gamma w))\bigr)+\frac{1}{\mathrm{vol}_\Gamma}\log\frac{\delta+1}{2}\\ &+\frac{1}{2\,\mathrm{vol}_\Gamma}\int_\delta^\infty\frac{U-T}{U^2-1}dU+\int_{\lambda=1/4}^\infty \nu(Y_0,\lambda)(-dH^-_{\delta,\lambda_{\min}})(\lambda)\end{aligned}$$

and

$$\begin{aligned}\mathrm{gr}_\Gamma(z,w) \ge{}& \frac{1}{\#(\Gamma\cap\{\pm 1\})}\sum_{\gamma\in\Pi(z,w)}\bigl(k_1(\delta)-k_1(u(z,\gamma w))\bigr)+\frac{1}{\mathrm{vol}_\Gamma}\log\frac{\delta+1}{2}\\ &-\frac{1}{2\,\mathrm{vol}_\Gamma}\int_\delta^\infty\frac{V-U}{U^2-1}dU-\int_{\lambda=1/4}^\infty \nu(Y_0,\lambda)(-dH^+_{\delta,\lambda_{\min}})(\lambda).\end{aligned}$$

Theorem 5.4. *Let Γ_0 be a cofinite Fuchsian group, let Y_0 be a compact subset of $\Gamma_0\backslash\mathbf{H}$, and let $\delta > 1$ and $\lambda_{\min} > 0$ be real numbers. There exist real numbers A and B such that the following holds. Let Γ be a subgroup of finite index in Γ_0, and let Y be the inverse image of Y_0 under the map $\Gamma_0\backslash\mathbf{H}\to\Gamma\backslash\mathbf{H}$. Suppose that the least positive eigenvalue of the Laplacian on $\mathrm{L}^2(\Gamma\backslash\mathbf{H})$ is at least $\lambda_{\min}$ and that the set*

$$\{\gamma\in\Gamma \mid u(z,\gamma w)\le\delta\}$$

contains at most one element for all $z, w \in \mathbf{H}$ whose images in $\Gamma\backslash\mathbf{H}$ lie in Y. Then the inequalities

$$\mathrm{gr}_\Gamma(z,w)\le B+\min\{0,k_1(\delta)-k_1(d(z,w))\}$$

and

$$\mathrm{gr}_\Gamma(z,w)\ge A+\min\{0,k_1(\delta)-k_1(d(z,w))\}$$

hold for all z, w in Y, where $d(z,w)$ is the "distance function" defined in § 1.2.

Proof. This follows from the above inequalities. □

Remark. For simplicity, we have limited ourselves in the above theorem to groups Γ that do not contain any elliptic elements that Γ_0 may have. One way to treat the general case would be to take two compact subsets Y_0 and Y_0' such that $Y_0\cap Y_0'$ does not contain any elliptic points and δ is taken sufficiently small such that $\Pi_0(z,w)$ contains at most one element for all $z, w \in \mathbf{H}$ whose images in $\Gamma_0\backslash\mathbf{H}$ lie in Y_0 and Y_0', respectively.

5.5. Extension to neighbourhoods of the cusps

The need to choose a compact subset Y_0 of $\Gamma_0\backslash\mathbf{H}$ in § 5.4 means that we have to do some more work to find suitable bounds on the Green function $\mathrm{gr}_\Gamma(z,w)$ in the case where one or both of z and w is near a cusp of Γ.

Let Γ be a cofinite Fuchsian group, and let $\mathfrak{c}$ be a cusp of Γ. We choose a sufficiently small $\epsilon_\mathfrak{c} > 0$ such that disc $B_\mathfrak{c}(\epsilon_\mathfrak{c})$ of area $\epsilon_\mathfrak{c}$ around $\mathfrak{c}$ is well-defined. For any $w \notin B_\mathfrak{c}(\epsilon_\mathfrak{c})$, the function $\mathrm{gr}_\Gamma(z,w)$, viewed as a function of $z \in B_\mathfrak{c}(\epsilon_\mathfrak{c})$, satisfies

$$2i\partial\bar\partial\,\mathrm{gr}_\Gamma(z,w) = -\frac{1}{\mathrm{vol}_\Gamma}\mu_\mathbf{H}(z).$$

We can therefore write

$$\mathrm{gr}_\Gamma(z,w) = \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_\mathfrak{c} y_\mathfrak{c}(z)) + h_w(z) \quad \text{for all } z \in B_\mathfrak{c}(\epsilon_\mathfrak{c}),$$

where $y_\mathfrak{c}(z)$ is defined as in § 1.2 and h_w is a real-valued harmonic function defined on the compactification

$$\bar B_\mathfrak{c}(\epsilon_\mathfrak{c}) = \{z \in \Gamma\backslash\mathbf{H} \mid y_\mathfrak{c}(z) \ge 1/\epsilon_\mathfrak{c}\} \cup \{\mathfrak{c}\}.$$

By construction, $h_w(z)$ coincides with $\mathrm{gr}_\Gamma(z,w)$ for $y_\mathfrak{c}(z) = 1/\epsilon_\mathfrak{c}$; in other words, for z on the boundary of $\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})$. The maximum principle for harmonic functions now implies that

$$\mathrm{gr}_\Gamma(z,w) \le \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_\mathfrak{c} y_\mathfrak{c}(z)) + \sup_{z' \in \partial\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})} \mathrm{gr}_\Gamma(z',w) \quad \text{for all } z \in B_\mathfrak{c}(\epsilon_\mathfrak{c}),\ w \notin B_\mathfrak{c}(\epsilon_\mathfrak{c}).$$

Finally, considering the case where z and w both lie in $B_\mathfrak{c}(\epsilon_\mathfrak{c})$ for some cusp $\mathfrak{c}$, we get

$$\begin{aligned}\mathrm{gr}_\Gamma(z,w) \le \mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}(z,w) &+ \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_\mathfrak{c} y_\mathfrak{c}(z)) + \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_\mathfrak{c} y_\mathfrak{c}(w)) \\ &+ \sup_{z',w' \in \partial\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})} \mathrm{gr}_\Gamma(z',w')\end{aligned}$$

for all $z, w \in B_\mathfrak{c}(\epsilon_\mathfrak{c})$, where $\mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}$ is the Green function for the Laplace operator on the closed disc $\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})$. This Green function is defined by the differential equation

$$\left.\begin{aligned} 2i\partial\bar\partial\,\mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}(\ \ ,w) &= \delta_w, \\ \mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}(z,w) &= 0 \quad \text{if } |q_\mathfrak{c}(z)| = \exp(-2\pi/\epsilon_\mathfrak{c}) \end{aligned}\right\} \quad \text{for all } w \in \bar B_\mathfrak{c}(\epsilon_\mathfrak{c}).$$

It is given explicitly by

$$\mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}(z,w) = \frac{1}{2\pi}\log\left|\frac{(q_\mathfrak{c}(z) - q_\mathfrak{c}(w))\exp(2\pi/\epsilon_\mathfrak{c})}{1 - q_\mathfrak{c}(z)\overline{q_\mathfrak{c}(w)}\exp(4\pi/\epsilon_\mathfrak{c})}\right|,$$

where $q_\mathfrak{c}$ is the coordinate function defined in § 1.2. The function $\mathrm{gr}_{\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})}(z,w)$ is non-positive for all z and w on $\bar B_\mathfrak{c}(\epsilon_\mathfrak{c})$, and vanishes on the boundary.

Analogous lower bounds for gr_Γ hold with suprema replaced by infima.

Chapter III

Arakelov theory for modular curves

In this chapter we describe intersection theory on arithmetic surfaces, as developed by Arakelov [2] and Faltings [37]. We also give explicit bounds for canonical Green functions of modular curves, using the results for Fuchsian groups proved in the preceding chapter.

1. Analytic part

In this section we define the basic analytic concepts that are needed for Arakelov theory, namely admissible metrics on line bundles on compact Riemann surfaces. In the case of modular curves, we also compare the admissible metric on the line bundle of holomorphic differentials to the Petersson metric on the line bundle of cusp forms of weight 2.

1.1. Admissible metrics

Let X be a Riemann surface. For $n = 0$, 1 or 2, we write $\mathcal{E}_X^n$ for the sheaf of smooth complex-valued n-forms. There is a natural decomposition

$$\mathcal{E}_X^1 = \mathcal{E}_X^{(1,0)} \oplus \mathcal{E}_X^{(0,1)}.$$

Here $\mathcal{E}_X^{(1,0)}$ and $\mathcal{E}_X^{(0,1)}$ consist of differential forms that are locally of the form $f\,dz$ and $g\,d\bar{z}$, respectively, where z is a holomorphic coordinate and f and g are smooth functions. This decomposition causes each of the two differentials

$$\mathcal{E}_X^0 \xrightarrow{d} \mathcal{E}_X^1 \xrightarrow{d} \mathcal{E}_X^2$$

to split as the sum of two partial derivatives. These four partial derivatives fit in an anti-commutative diagram

$$\begin{array}{ccc} \mathcal{E}_X^0 & \xrightarrow{\partial} & \mathcal{E}_X^{(1,0)} \\ {\scriptstyle\bar\partial}\big\downarrow & & \big\downarrow{\scriptstyle\bar\partial} \\ \mathcal{E}_X^{(0,1)} & \xrightarrow{\partial} & \mathcal{E}_X^2. \end{array}$$

The *Laplace operator* on X is the $\mathbf{C}$-linear map

$$2i\partial\bar\partial\colon \mathcal{E}_X^0 \to \mathcal{E}_X^2.$$

Now let X be a compact connected Riemann surface of genus $g \geq 1$. The space $\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$ of global holomorphic 1-forms on X has a Hermitean inner product defined by

$$\langle \alpha, \beta \rangle = \frac{i}{2} \int_X \alpha \wedge \bar\beta. \tag{1.1}$$

The *canonical (1,1)-form* on X is defined as

$$\mu_X^{\mathrm{can}} = \frac{i}{2g} \sum_{j=1}^{g} \alpha_j \wedge \bar\alpha_j, \tag{1.2}$$

where $(\alpha_1, \ldots, \alpha_g)$ is any orthonormal basis of $\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$.

If $\mathcal{L}$ is a line bundle on X, an *admissible metric* on $\mathcal{L}$ is a smooth Hermitean metric $|\ \ |$ on $\mathcal{L}$ that locally on X satisfies

$$\frac{1}{\pi i} \partial\bar\partial \log|s| = (\deg \mathcal{L}) \mu_X^{\mathrm{can}}$$

for some (hence any) local generating section s of $\mathcal{L}$. An *admissible line bundle* on X is a line bundle equipped with an admissible metric.

There exists a unique smooth (i.e. infinitely differentiable) function $\mathrm{gr}_X^{\mathrm{can}}$ outside the diagonal on $X \times X$ such that

$$2i\partial\bar\partial \,\mathrm{gr}_X^{\mathrm{can}}(\ \ , y) = \delta_y - \mu_X^{\mathrm{can}} \quad \text{and} \quad \int_X \mathrm{gr}_X^{\mathrm{can}}(\ \ , y)\mu_X^{\mathrm{can}} = 0 \quad \text{for all } y \in X.$$

This function is called the *canonical Green function* of X. For a proof of the existence of $\mathrm{gr}_X^{\mathrm{can}}$, see for example Elkik [106, exposé III].

Remark. Various normalisation conventions for the Green function can be found in the literature. Our $\mathrm{gr}_X^{\mathrm{can}}$ is $\frac{1}{2\pi}$ times the Green function used by Arakelov [2] and Faltings [37].

Let D be a divisor on X. The line bundle $\mathcal{O}_X(D)$ admits a canonical admissible metric $|\ \ |_{\mathcal{O}_X(D)}$, defined by putting

$$\log|1|_{\mathcal{O}_X(D)}(y) = 2\pi \sum_{x \in X} n_x \,\mathrm{gr}_X^{\mathrm{can}}(x, y) \quad \Bigl(D = \sum_{x \in X} n_x x\Bigr) \tag{1.3}$$

for y outside the support of D, and extending by continuity. Furthermore, there is a canonical admissible metric on the line bundle $\Omega^1_{X/\mathbf{C}}$ of holomorphic differentials, defined uniquely by

$$\log|dz|_{\Omega^1_{X/\mathbf{C}}}(x) = \lim_{y \to x} \bigl(\log|z(y) - z(x)| - 2\pi \,\mathrm{gr}_X^{\mathrm{can}}(x, y)\bigr) \quad \text{for } x \in X \tag{1.4}$$

if z is a local coordinate in a neighbourhood of x.

To every admissible line bundle $\mathcal{L}$ on X, we associate a one-dimensional complex vector space

$$\lambda(\mathcal{L}) = \det \mathrm{H}^0(X, \mathcal{L}) \otimes \det \mathrm{H}^1(X, \mathcal{L})^\vee,$$

called the *determinant of cohomology* of $\mathcal{L}$. Faltings proved in [37, Theorem 1] that there is a unique way to assign metrics to the $\lambda(\mathcal{L})$ for all admissible line bundles $\mathcal{L}$ such that the following axioms are satisfied,

(1) For every isometry $f\colon \mathcal{L} \xrightarrow{\sim} \mathcal{M}$ between admissible line bundles on X, the induced isomorphism

$$\lambda(f)\colon \lambda(\mathcal{L}) \xrightarrow{\sim} \lambda(\mathcal{M})$$

is an isometry.

(2) If the metric on $\mathcal{L}$ is scaled by a factor $\alpha > 0$, the metric on $\lambda(\mathcal{L})$ changes by a factor $\alpha^{\chi(\mathcal{L})}$, where

$$\chi(\mathcal{L}) = \dim \mathrm{H}^0(X, \mathcal{L}) - \dim \mathrm{H}^1(X, \mathcal{L})$$

is the Euler characteristic of $\mathcal{L}$.

(3) For every admissible line bundle $\mathcal{L}$ on X and every point $P \in X$, the canonical exact sequence

$$0 \longrightarrow \mathcal{L}(-P) \longrightarrow \mathcal{L} \longrightarrow P_* P^* \mathcal{L} \longrightarrow 0$$

induces an isometry

$$\lambda(\mathcal{L}) \xrightarrow{\sim} \lambda(\mathcal{L}(-P)) \otimes P^* \mathcal{L}.$$

(4) The metric on

$$\lambda(\Omega^1_{X/\mathbf{C}}) \cong \det \mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$$

comes from the inner product (1.1) on $\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$.

For later use, we extend the definition of the canonical $(1,1)$-form and the canonical Green function to the case $g = 0$, i.e. to the complex projective line $\mathbf{P}^1(\mathbf{C})$. We endow $\mathbf{P}^1(\mathbf{C})$ with the volume form for the *Fubini-Study metric*. This is the $(1,1)$-form defined as

$$\mu_{\mathbf{P}^1} = \frac{i}{2\pi} \frac{dz \wedge d\bar{z}}{(1 + |z|^2)^2};$$

this depends on the choice of the coordinate z. The corresponding Green function (defined as above) is given by

$$\mathrm{gr}_{\mathbf{P}^1}(z, w) = \frac{1}{4\pi} + \frac{1}{4\pi} \log \frac{|z - w|^2}{(1 + |z|^2)(1 + |w|^2)}.$$

1.2. Comparison between admissible and Petersson metrics

Let Γ be a cofinite Fuchsian group, and let X be the compactification of $\Gamma\backslash\mathbf{H}$ obtained by adding the cusps. We assume that Γ has no elliptic elements and that $g_X \ge 1$.

We equip the line bundle $\omega^{\otimes 2}(-\text{cusps})$ of cusp forms of weight 2 for Γ with the Petersson metric

$$|f|_{2,\mathrm{Pet}}(z) = (\Im z)|f(z)| \quad \text{for } z \in \mathbf{H}$$

as in § II.2.1. Furthermore, we have the line bundle $\Omega^1_{X/\mathbf{C}}$, equipped with the admissible metric (1.4). There is a canonical isomorphism

$$\Omega^1_{X/\mathbf{C}} \cong \omega^{\otimes 2}(-\text{cusps}) \tag{1.5}$$

constructed as follows: for a local section α of $\Omega^1_{X/\mathbf{C}}$, the pull-back of α to $\mathbf{H}$ can be written as

$$\alpha = f\,dz,$$

where f is a local section of $\omega^{\otimes 2}(-\text{cusps})$. Taking global sections, we obtain an isomorphism

$$\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}}) \cong \mathrm{S}_2(\Gamma). \tag{1.6}$$

Under this isomorphism, the inner product (1.1) on $\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$ corresponds to the Petersson inner product $\langle\ ,\ \rangle_\Gamma$ on $\mathrm{S}_2(\Gamma)$. This implies that the two $(1,1)$-forms μ_X^{can} and $\mu_{\mathbf{H}}$ on X can be compared as follows. We consider the function

$$F_\Gamma\colon X \to [0,\infty)$$

defined on the open subset $\Gamma\backslash\mathbf{H}$ by

$$F_\Gamma(z) = \sum_{f\in B} |f|^2_{2,\mathrm{Pet}}, \tag{1.7}$$

where B is an orthonormal basis of the $\mathbf{C}$-vector space $\mathrm{S}_2(\Gamma)$ of cusp forms of weight 2 with respect to the Petersson inner product; we extend F_Γ by zero to the cusps. From the isomorphism (1.6) and the definition (1.2) of μ_X^{can}, we get

$$\mu_X^{\mathrm{can}} = \frac{1}{g_X} F_\Gamma \mu_{\mathbf{H}}. \tag{1.8}$$

We can now compare the metrics $|\alpha|_{\Omega^1_{X/\mathbf{C}}}$ and $|f|_{2,\mathrm{Pet}}$. As in § II.5.1, we define

$$k_1(u) = \frac{1}{4\pi}\frac{u+1}{u-1}.$$

From the formula for the function $u(z,w) = \cosh r(z,w)$ given in § II.1.1, it follows that

$$k_1(u(z,w)) = \frac{1}{4\pi}\log\biggl(1 + \frac{4\,\Im z\,\Im w}{|z-w|^2}\biggr).$$

We define

$$\begin{aligned} H_\Gamma\colon \Gamma\backslash\mathbf{H} &\to \mathbf{R} \\ z &\mapsto \lim_{w\to z}\bigl(k_1(u(z,w)) + \mathrm{gr}_X^{\mathrm{can}}(z,w)\bigr). \end{aligned} \tag{1.9}$$

We view it as a function on X with singularities at the cusps.

Lemma 1.1. *The function H_Γ satisfies*

$$2\pi H_\Gamma = -\log|\alpha|_{\Omega^1_{X/\mathbf{C}}} + \log|f|_{2,\mathrm{Pet}} + \log 2 \tag{1.10}$$

for all local sections α of $\Omega^1_{X/\mathbf{C}}$ and f of $\omega^{\otimes 2}(-\mathrm{cusps})$ corresponding to each other via the isomorphism (1.5), and

$$2i\partial\bar\partial H_\Gamma = (2g_X - 2)\mu_X^{\mathrm{can}} - \frac{1}{2\pi}\mu_{\mathbf{H}} + \sum_{\mathfrak{c}\ \mathrm{cusp}} \delta_{\mathfrak{c}}. \tag{1.11}$$

Proof. We rewrite $\log|\alpha|_{\Omega^1_{X/\mathbf{C}}}$ as

$$\begin{aligned}\log|\alpha|_{\Omega^1_{X/\mathbf{C}}} &= \log|f(z)| + \log|dz|_{\Omega^1_{X/\mathbf{C}}}\\ &= \log|f|_{2,\mathrm{Pet}} + \lim_{w\to z}\left(\frac{1}{2}\log\frac{|z-w|^2}{\Im z\,\Im w} - 2\pi\,\mathrm{gr}_X^{\mathrm{can}}(z,w)\right).\end{aligned}$$

One easily verifies that

$$\frac{1}{2}\log\frac{|z-w|^2}{\Im z\,\Im w} + 2\pi k_1(u(z,w)) \to \log 2 \quad \text{as } w\to z.$$

This implies the equality

$$\lim_{w\to z}\left(\frac{1}{2}\log\frac{|z-w|^2}{\Im z\,\Im w} - 2\pi\,\mathrm{gr}_X^{\mathrm{can}}(z,w)\right) = \log 2 - 2\pi H_\Gamma(z)$$

and hence (1.10). We are going to deduce (1.11) by applying the operator $2i\partial\bar\partial$ to (1.10) where α and f are local generating sections of $\Omega^1_{X/\mathbf{C}}$ and $\omega^{\otimes 2}(-\mathrm{cusps})$ corresponding to each other via (1.5). First we note that for any local generating section α of $\Omega^1_{X/\mathbf{C}}$, the admissibility of $|\ \ |_{\Omega^1_{X/\mathbf{C}}}$ implies that

$$2i\partial\bar\partial\log|\alpha|_{\Omega^1_{X/\mathbf{C}}} = -2\pi(2g_X-2)\mu_X^{\mathrm{can}}.$$

To prove the lemma, it remains to prove that

$$2i\partial\bar\partial\log|f|_{2,\mathrm{Pet}} = -\mu_{\mathbf{H}} + 2\pi\delta_{\mathfrak{c}}. \tag{1.12}$$

Outside the cusps, this follows from the definition of $\log|f|_{2,\mathrm{Pet}}$. Near a cusp $\mathfrak{c}$ we may write f in terms of the coordinate $q_{\mathfrak{c}}$ introduced in § II.1.2 as

$$f = a_1 q_{\mathfrak{c}} + a_2 q_{\mathfrak{c}}^2 + \cdots \quad \text{with } a_1 \neq 0,$$

This implies that (1.12) holds everywhere. ∎

2. Intersection theory on arithmetic surfaces

In this section we give a very brief overview of the results that we need from Arakelov's intersection theory on arithmetic surfaces [2], as extended by Faltings [37]. Another useful reference is Szpiro [106].

Let K be a number field. We write K_{fin} and K_{inf} for the sets of finite and infinite places of K, respectively. For every place v of K we write K_v for the completion of K at v. For every $v \in K_{\mathrm{inf}}$, we choose an algebraic closure $\bar{K}_v$ of K_v; this is (non-canonically) isomorphic to $\mathbf{C}$. We write B for the spectrum of the ring of integers of K.

A *metrised line bundle* on B is a line bundle $\mathcal{L}$ on B together with a Hermitean inner product $\langle\ ,\ \rangle_v$ on the geometric fibre $\mathcal{L}_v$ of L at each infinite place v of K, where we view $\mathcal{L}_v$ as a one-dimensional $\bar{K}_v$-vector space. We denote by $|\ \ |_v$ the corresponding norm, defined by

$$|x|_v^2 = \langle x, x\rangle_v \quad \text{for } x \in \mathcal{L}_v.$$

The *degree* of a metrised line bundle $(\mathcal{L}, |\ \ |)$ is defined as

$$\deg(\mathcal{L}, |\ \ |) = \sum_{v\in K_{\mathrm{fin}}} \mathrm{ord}_v(s)\log \#k_v + \sum_{v\in K_{\mathrm{inf}}} (-\log|s|_v)[K_v : \mathbf{R}], \tag{2.1}$$

where s is any non-zero rational section of $\mathcal{L}$ and $\mathrm{ord}_v(s)$ is the order of vanishing of s at v. This degree is well-defined by the product formula for the places of K.

An *arithmetic surface* over B is a proper flat morphism

$$\pi\colon X \to B,$$

where X is a normal integral scheme of Krull dimension 2, such that the generic fibre of π is geometrically connected. For each infinite place v of X, we let $\mathfrak{X}_v$ denote the compact connected Riemann surface $X(\bar{K}_v)$. Any line bundle $\mathcal{L}$ on X gives rise to a line bundle $\mathcal{L}_v$ on each of the $\mathfrak{X}_v$.

A *metrised line bundle* on X is a line bundle $\mathcal{L}$ on X together with a Hermitean inner product on the line bundle $\mathcal{L}_v$ for each infinite place v of K. As above, we denote by $|\ \ |_v$ the corresponding norm. If the genus of X is at least 1, an *admissible line bundle* on X is a metrised line bundle $(\mathcal{L}, (|\ \ |_v)_{v\in K_{\mathrm{inf}}})$ such that for each $v \in K_{\mathrm{inf}}$ the metric $|\ \ |_v$ on $\mathcal{L}_v$ is admissible in the sense of § 1.1.

An *Arakelov divisor* on X is a formal linear combination

$$D = D_{\mathrm{fin}} + \sum_{v\in K_{\mathrm{inf}}} a_v \mathfrak{X}_v$$

where D_{fin} is a Cartier divisor on X and the a_v are real numbers; the $\mathfrak{X}_v$ play the role of "vertical prime divisors at infinity". We say that an Arakelov divisor D is *horizontal* if all the a_v are zero and every irreducible component of D_{fin} is flat over B. We say D is *vertical* if D_{fin} is a linear combination of irreducible components of the fibres of X. The *principal Arakelov divisor* associated to a non-zero rational function f on X is

$$\mathrm{div}(f) = \mathrm{div}_{\mathrm{fin}}(f) + \sum_{v\in K_{\mathrm{inf}}} a_v(f)\,\mathfrak{X}_v,$$

where $\operatorname{div}_{\mathrm{fin}}(f)$ denotes the usual (Cartier) divisor of f and where $a_v(f)$ is defined for $f \in K_{\mathrm{inf}}$ by

$$a_v(f) = -\int_{\mathfrak{X}_v} \log|f|_v \, \mu^{\mathrm{can}}_{\mathfrak{X}_v}.$$

Let D be an Arakelov divisor on X. We make $\mathcal{O}_X(D)$ into a admissible line bundle by equipping the pull-back of $\mathcal{O}_X(D)$ to each $\mathfrak{X}_v$ with the metric $|\ |_{\mathcal{O}_{\mathfrak{X}_v}(D_v)}$ defined by (1.3).

The *Arakelov class group* on X is the group $\operatorname{Cl} X$ of Arakelov divisors modulo the subgroup of principal divisors. The *Picard group* of an arithmetic surface X is the group $\operatorname{Pic} X$ of isomorphism classes of admissible line bundles on X. There is a canonical isomorphism between these groups via the map that sends an Arakelov divisor D to the admissible line bundle $\mathcal{O}_X(D)$.

The *Arakelov intersection pairing* is the unique symmetric bilinear map

$$(\ .\ )\colon \operatorname{Cl} X \times \operatorname{Cl} X \longrightarrow \mathbf{R}$$

with the following properties. If C and D are effective Cartier divisors without common components, then

$$(C\,.\,D) = \sum_{x\in X} \log \#k(x) i_x(C,D) - \sum_{v\in K_{\mathrm{inf}}} [K_v : \mathbf{R}]\cdot 2\pi \operatorname{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(C_v, D_v),$$

where x ranges over the closed points of x, the residue field at x is denoted by $k(x)$, and i_x is the local intersection number at x. If C is a horizontal Cartier divisor of degree n over B, then $(C\,.\,\mathfrak{X}_v) = n[K_v : \mathbf{R}]$ for every infinite place v of K. Finally, $(\mathfrak{X}_v\,.\,\mathfrak{X}_w) = 0$ for all infinite places v and w of K.

The definition of the intersection pairing implies that if $S\colon B \to X$ is a section whose image is a Cartier divisor (also denoted by S), then

$$(S\,.\,D) = \deg S^*\mathcal{O}_X(D) \tag{2.2}$$

for any Arakelov divisor D on X.

To any line bundle $\mathcal{L}$ on X we associate a line bundle $\lambda_\pi\mathcal{L}$ on B, called the *determinant of cohomology* of $\mathcal{L}$. It is defined as

$$\lambda_\pi\mathcal{L} = \det \pi_*\mathcal{L} \otimes (\det \mathrm{R}^1\pi_*\mathcal{L})^\vee.$$

where the determinant of a coherent sheaf on B is defined using a resolution by locally free sheaves of finite rank; see Moret-Bailly [106, exposé II, § 1.1]. The formation of $\lambda_\pi\mathcal{L}$ is compatible with arbitrary base change on B. The line bundle λ_π is made into a metrised line bundle by equipping the fibre

$$(\lambda_\pi\mathcal{L})_{K_v} = \det \mathrm{H}^0(\mathfrak{X}_v, \mathcal{L}_{\mathfrak{X}_v}) \otimes \det \mathrm{H}^1(\mathfrak{X}_v, \mathcal{L}_{\mathfrak{X}_v})^\vee$$

for each infinite place of K with the metric given by Faltings's axioms as in § 1.1.

From now on we assume that the morphism $\pi\colon X \to B$ is semi-stable. In this case there exists a line bundle Ω_π called the *relative dualising sheaf*. On the open subscheme of X where π is smooth, Ω_π coincides with the line bundle Ω^1_π of relative differential forms. In particular, we can equip it at the infinite places of K with the canonical admissible metric defined in § 1.1.

Let S be a section of π such that the image of S lies in the regular part of X, so that S is a Cartier divisor. Then we have the *adjunction formula*

$$(S\,.\,S) = -(S\,.\,\Omega_\pi). \tag{2.3}$$

The determinant of cohomology is compatible with the relative version of Serre duality (see Hartshorne [42, III, § 11]). More precisely, if $\mathcal{L}$ is an admissible line bundle on X, then the metrised line bundles $\lambda_\pi\mathcal{L}$ and $\lambda_\pi(\mathcal{L}^\vee\otimes\Omega_\pi)$ on B are canonically isomorphic; see Moret-Bailly [106, exposé II, proposition 4.15.2]. Furthermore,

$$\deg\lambda_\pi\mathcal{L} = \deg\bigl[\det\pi_*\mathcal{L}\otimes\det\pi_*\mathcal{H}om(\mathcal{L},\Omega_\pi)\bigr] - \log\#(\mathrm{H}^1(X,\mathcal{L})_{\mathrm{tor}}),$$

where the line bundle $\det\pi_*\mathcal{L}\otimes\det\pi_*\mathcal{H}om(\mathcal{L},\Omega_\pi)$ is metrised according to Faltings's axioms and $\mathrm{H}^1(X,\mathcal{L})_{\mathrm{tor}}$ is the torsion submodule of $\mathrm{H}^1(X,\mathcal{L})$.

If $\pi\colon X\to B$ is semi-stable, we have Faltings's *arithmetic Riemann–Roch formula*

$$\deg\lambda_\pi\mathcal{L} = \frac{1}{2}(\mathcal{L}\,.\,\mathcal{L}\otimes\Omega_\pi^\vee) + \deg\lambda_\pi\mathcal{O}_X.$$

Definition. Let X be a curve over $\overline{\mathbf{Q}}$. Let K be a number field such that X has a semi-stable model $\pi\colon X_{\mathbf{Z}_K}\to\operatorname{Spec}\mathbf{Z}_K$ over the ring of integers $\mathbf{Z}_K$ of K. The *Faltings height* of X is

$$h_{\mathrm{Faltings}}(X) = \frac{1}{[K:\mathbf{Q}]}\deg\lambda_\pi\mathcal{O}_{X_{\mathbf{Z}_K}};$$

this is independent of the choice of K and of the semi-stable model.

2.1. Heights

Let K be a number field. For every place v of K, let K_v denote the completion of K at v, and let

$$|\ \ |_v\colon K_v\to[0,\infty)$$

denote the absolute value corresponding to v, normalised so that multiplication by x scales the Haar measure on K_v by a factor $|x|_v$.

Let K be a number field, let x be an element of $\bar K$, and let $L\subset\bar K$ be any finite extension of K containing x. The *height* of x (relative to K) is the real number defined by

$$h_K(x) = \frac{1}{[L:K]}\sum_v\log\max\{1,|x|_v\},$$

where v runs over all (finite and infinite) places of L; this is independent of the choice of L. More generally, for any point $x=(x_0:x_1:\ldots:x_n)$ in some $\mathbf{P}^n(\bar K)$, we define

$$h_{\mathbf{P}^n/K}(x) = \frac{1}{[L:K]}\sum_v\log\max\{|x_0|_v,\ldots,|x_n|_v\},$$

where L is any finite extension of K containing all the x_i. This is well defined because of the product formula for the places of K.

2.2. The Néron–Tate pairing and points of small height

Let A be an Abelian variety over a number field K, and let $\mathcal{L}$ be a symmetric ample line bundle on A. The *Néron–Tate height* on A with respect to $\mathcal{L}$ is the real-valued quadratic form $h^{\mathcal{L}}_{A/K}$ on the Abelian group $A(\bar{K})$ defined as follows. We choose an integer m such that $\mathcal{L}^{\otimes m}$ is very ample and a K-basis $(b_0, \ldots, b_r)$ of $\mathrm{H}^0(A, \mathcal{L}^{\otimes m})$. These choices define a projective embedding $i\colon A(\bar{K}) \to \mathbf{P}^r(\bar{K})$. Then the sequence $\{n^{-2} h_{\mathbf{P}^r/K}(i(nx))\}_{n \geq 1}$ converges as $n \to \infty$, and the limit

$$h^{\mathcal{L}}_{A/K}(x) = m^{-1} \lim_{n \to \infty} n^{-2} h_{\mathbf{P}^r/K}(i(nx))$$

does not depend on the choice of m and $(b_0, \ldots, b_r)$. The map

$$h^{\mathcal{L}}_{A/K}\colon A(\bar{K}) \to \mathbf{R}$$

is a quadratic form. The associated symmetric bilinear form is called the *Néron–Tate pairing* on A and is denoted by $\langle\ ,\ \rangle^{\mathcal{L}}_{A/K}$.

Let K be a number field, let B denote the spectrum of its ring of integers, and let $\pi\colon X \to B$ be a regular and semi-stable arithmetic surface with fibres of genus $g \geq 2$ whose generic fibre is smooth and geometrically connected. Let J be the Jacobian of X over K. We write $\mathrm{h}^{\mathrm{NT}}_J$ and $\langle\ ,\ \rangle_J$ for the Néron–Tate height and the Néron–Tate pairing on J with respect to the ample line bundle $\mathcal{O}_J(\Theta)$, where Θ is a symmetric theta divisor. The basic relation between the Néron–Tate pairing and Arakelov theory is the *Faltings–Hriljac formula*; see Faltings [37, Theorem 4(c)] or Moret-Bailly [106, exposé II, théorème 6.15]. It says that if $\mathcal{L}$ and $\mathcal{M}$ are two admissible line bundles of degree 0 and at least one of them has intersection number 0 with every irreducible component of every fibre, then

$$(\mathcal{L} \,.\, \mathcal{M}) = -[K : \mathbf{Q}] \langle [\mathcal{L}_K], [\mathcal{M}_K] \rangle_J,$$

where the square brackets denote the point of the Jacobian corresponding to a line bundle of degree 0.

We will later need the following fact due to Zhang [116, Theorem 5.6]: if D is a divisor of degree 1 on X_K such that $[(2g-2)D - \Omega_{\pi_K}]$ is a torsion point of the Jacobian, then for every $\epsilon > 0$ there are infinitely many points $x \in X(\bar{K})$ such that the Néron–Tate height (relative to the base field K) of the point $[x - D]$ in the Jacobian of X is less than $\Omega^2_{X/K,\mathrm{a}}/(2g-2) + \epsilon$, where $\Omega^2_{X/K,\mathrm{a}}$ is the self-intersection of the relative dualising sheaf of X_K in the sense of Zhang [116]. A consequence of this is the following generalisation.

Lemma 2.1. *Let X_K be a proper, smooth and geometrically connected curve of genus $g \geq 2$ over a number field K, let J be the Jacobian of X_K, and let D be a divisor of degree 1 on X_K such that $[(2g-2)D - \Omega_{X_K/K}]$ is a torsion point of J. Then for every positive integer d and every $\epsilon > 0$ there exist infinitely many effective divisors R of degree d on $X_{\bar{K}}$ such that*

$$\mathrm{h}^{\mathrm{NT}}_{J/K}([R - dD]) < d^2 \frac{\Omega^2_{X/K,\mathrm{a}}}{2g-2} + \epsilon.$$

Moreover, if $\mathcal{L}$ is a given line bundle of degree at most $g-d-1$ without non-zero global sections, there are infinitely many such R for which $\mathcal{L}(R)$ still does not have a non-zero global section.

Proof. Consider an effective divisor R of degree d on $X_{\bar{K}}$, and write

$$R = P_1 + \cdots + P_d \quad (P_i \in X(\bar{K})).$$

The fact that $\mathrm{h}^{\mathrm{NT}}_{J/K}$ is a quadratic form taking non-negative values implies that

$$\begin{aligned}
\mathrm{h}^{\mathrm{NT}}_{J/K}([R-dD]) &= \mathrm{h}^{\mathrm{NT}}_{J/K}([P_1-D]+\cdots+[P_d-D]) \\
&= \sum_{i=1}^{d} \mathrm{h}^{\mathrm{NT}}_{J/K}([P_i-D]) + \sum_{i\neq j}\langle [P_i-D],[P_j-D]\rangle \\
&\le \sum_{i=1}^{d} \mathrm{h}^{\mathrm{NT}}_{J/K}([P_i-D]) + \frac{1}{2}\sum_{i\neq j}\bigl(\mathrm{h}^{\mathrm{NT}}_{J/K}([P_i-D]) + \mathrm{h}^{\mathrm{NT}}_{J/K}([P_j-D])\bigr) \\
&= d\sum_{i=1}^{d} \mathrm{h}^{\mathrm{NT}}_{J/K}([P_i-D]).
\end{aligned}$$

By Zhang's theorem cited above, there are infinitely many ways to choose points P_i on $X_{\bar{K}}$ such that

$$\mathrm{h}^{\mathrm{NT}}_{J/K}([P_i-D]) < \frac{\Omega^2_{X/K,\mathrm{a}}}{2g-2} + \epsilon/d^2.$$

This implies the first claim of the lemma. Now if $\mathcal{L}$ is a line bundle of degree at most $g-d-1$ without non-zero global sections, then

$$\dim \mathrm{H}^1(X_{\bar{k}},\mathcal{L}) = g-1-\deg\mathcal{L}$$

by the Riemann–Roch formula. Via Serre duality, we see that there are infinitely many ways to choose the P_i such that

$$\dim_{\bar{k}} \mathrm{H}^1(X_{\bar{k}},\mathcal{L}(P_1,\ldots,P_i)) = g-1-\deg\mathcal{L}-i \quad \text{for } i=0,\,1,\,\ldots,\,d.$$

For every such choice of the P_i, applying the Riemann–Roch formula again shows that $\mathrm{H}^0(X_{\bar{k}},\mathcal{L}(P_1,\ldots,P_d))=0$, which proves the second claim. □

If X_K is the generic fibre of a semi-stable arithmetic surface $\pi\colon X\to B$, the real number $\Omega^2_{X/K,\mathrm{a}}$ is related to the self-intersection (in the sense of Arakelov's intersection theory) of the dualising sheaf Ω_π of X over B via the formula

$$\Omega^2_{X/K,\mathrm{a}} = (\Omega_\pi\,.\,\Omega_\pi) - \sum_{v\in K_{\mathrm{fin}}} r_v \log\#k(v)$$

(see [116, Theorem 5.5]), where the r_v (defined in [116, § 4]) are certain non-negative real numbers that vanish for all finite places of K such that the corresponding fibre of X is smooth.

3. Bounds on analytic data for modular curves

In this section we derive bounds on various analytic data associated to Riemann surfaces that are compactifications of quotients of the hyperbolic plane by Fuchsian groups Γ that are of finite index in a fixed cofinite Fuchsian group Γ_0. Most importantly, in §§ 3.2–3.5 we derive bounds on canonical Green functions of such Riemann surfaces. These bounds are stated in terms of those obtained in Sections II.4 and II.5. Another result that is proved in this section and is used later is a bound on the function H_Γ defined by (1.9), which relates the difference between the admissible and Petersson metrics on the line bundle of differentials by Lemma 1.1. Finally, in § 3.7 we find an upper bound on a certain integral that is, roughly speaking, the average of the logarithm of the norm of a given differential with respect to the canonical admissible metric.

Remark. A different approach to the problem of bounding canonical Green functions was taken by Jorgenson and Kramer in [51], who found an interesting expression for the canonical Green function purely in terms of data associated with the hyperbolic metric; see [51, Theorem 3.8] (we note that a minus sign is missing in the cited theorem). All things considered, however, their methods appear to be more involved than ours.

3.1. Notation

Let Γ_0 be a cofinite Fuchsian group. For every cusp $\mathfrak{c}$ of Γ_0 we fix a real number $\epsilon_\mathfrak{c} > 0$ such that the discs $B_\mathfrak{c}(\epsilon_\mathfrak{c})$ of area $\epsilon_\mathfrak{c}$ around $\mathfrak{c}$, as defined in § II.1.2, are well-defined and pairwise disjoint. We define a compact subset Y_0 of $\Gamma_0\backslash\mathbf{H}$ by

$$Y_0 = (\Gamma_0\backslash\mathbf{H}) \setminus \bigsqcup_{\mathfrak{c}} B_\mathfrak{c}(\epsilon_\mathfrak{c}),$$

where $\mathfrak{c}$ runs over the cusps of Γ_0. Furthermore, we choose a real number $\delta > 1$ such that for all $z, w \in \mathbf{H}$ whose images in $\Gamma_0\backslash\mathbf{H}$ lie in Y_0, the set

$$\{\gamma \in \Gamma_0 \mid \gamma \text{ is not elliptic and } u(z, \gamma w) \le \delta\}$$

contains at most one element. Finally, we fix a positive real number λ.

Let Γ be a subgroup of finite index in Γ_0, and let X be the compactification of $\Gamma\backslash\mathbf{H}$. We assume that X has genus $g_X \ge 1$, that Γ does not contain any elliptic elements, and (as in § II.5.4) that the non-zero eigenvalues of the Laplace operator on $\Gamma\backslash\mathbf{H}$ are bounded from below by λ. For every cusp $\mathfrak{c}$ of Γ, we denote by $m_\mathfrak{c}$ the ramification index at $\mathfrak{c}$ of the map from the compactification of $\Gamma\backslash\mathbf{H}$ to that of $\Gamma_0\backslash\mathbf{H}$. We abbreviate

$$\epsilon_\mathfrak{c} = m_\mathfrak{c}\epsilon_{\mathfrak{c}_0},$$

where $\mathfrak{c}_0$ is the cusp of Γ_0 over which $\mathfrak{c}$ lies. We write Y for the inverse image of Y_0 in $\Gamma\backslash\mathbf{H}$; this is the complement of the discs $B_\mathfrak{c}(\epsilon_\mathfrak{c})$, where $\mathfrak{c}$ runs over the cusps of Γ.

3.2. Comparison between hyperbolic and canonical Green functions

There are two interesting Green functions associated to the Riemann surface X. First, we have the Green function gr_Γ outside the diagonal on $\Gamma\backslash\mathbf{H}\times\Gamma\backslash\mathbf{H}$ given by the structure of $\Gamma\backslash\mathbf{H}$ as a quotient of the upper half-plane by a Fuchsian group. This Green function has doubly logarithmic singularities near the cusps; this was made precise in § II.3.6. Second, we have the canonical Green function $\mathrm{gr}_X^{\mathrm{can}}$ outside the diagonal on $X\times X$, given by the structure of X as a compact Riemann surface of genus at least 1. There is a standard way to relate these two Green functions, which we will use to find explicit bounds on canonical Green functions.

Let μ_X^{can} be the canonical $(1,1)$-form on X as in § 1.1. We define a real-valued function h_Γ on $\Gamma\backslash\mathbf{H}$ by

$$\begin{aligned} h_\Gamma(z) &= \int_{w\in\Gamma\backslash\mathbf{H}} \mathrm{gr}_\Gamma(z,w)\mu_X^{\mathrm{can}}(w) \\ &= \frac{1}{g_X}\int_{w\in\Gamma\backslash\mathbf{H}} \mathrm{gr}_\Gamma(z,w)F_\Gamma(w)\mu_{\mathbf{H}}(w), \end{aligned}$$

where F_Γ is the function defined by (1.7). This integral converges since F_Γ is smooth and bounded on $\Gamma\backslash\mathbf{H}$.

By the definition of the Laplace operator Δ and the Green function gr_Γ in §§ II.1.1 and II.3.6, respectively, the function h_Γ satisfies

$$\begin{aligned} -\Delta h_\Gamma &= \frac{1}{g_X}F_\Gamma - \frac{1}{g_X\,\mathrm{vol}_\Gamma}\int_X F_\Gamma\mu_{\mathbf{H}} \\ &= \frac{1}{g_X}F_\Gamma - \frac{1}{\mathrm{vol}_\Gamma}, \end{aligned}$$

or equivalently

$$2i\partial\bar\partial h_\Gamma = \mu_X^{\mathrm{can}} - \frac{1}{\mathrm{vol}_\Gamma}\mu_{\mathbf{H}},$$

on $\Gamma\backslash\mathbf{H}$. Furthermore, if $\mathfrak{c}$ is a cusp of Γ and $q_{\mathfrak{c}}\colon\mathbf{H}\to(0,\infty)$ is the function defined in § II.1.2, then both $h_\Gamma(z)$ and $\mathrm{gr}_\Gamma(z,w)$ for fixed w have a singularity of the form $\mathrm{vol}_\Gamma^{-1}\log y_{\mathfrak{c}}(z)$ as $y_{\mathfrak{c}}(z)\to\infty$. This implies that the canonical Green function of X can be expressed as

$$\mathrm{gr}_X^{\mathrm{can}}(z,w) = \mathrm{gr}_\Gamma(z,w) - h_\Gamma(z) - h_\Gamma(w) + \int_{\Gamma\backslash\mathbf{H}} h_\Gamma\mu_X^{\mathrm{can}}. \tag{3.1}$$

We will use this expression to find bounds on $\mathrm{gr}_X^{\mathrm{can}}$.

3.3. Bounds on the function h_Γ

We are going to bound the function h_Γ on Y, uniformly in Γ, using the results of §§ II.4.2, II.4.3, II.5.4 and II.5.5. For $z\in Y$, we decompose the integral defining $h_\Gamma(z)$ as

$$h_\Gamma(z) = \frac{1}{g_X}\int\limits_{w\in Y}\mathrm{gr}_\Gamma(z,w)F_\Gamma(w)\mu_{\mathbf{H}}(w) + \frac{1}{g_X}\sum_{\mathfrak{c}}\int\limits_{w\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\mathrm{gr}_\Gamma(z,w)F_\Gamma(w)\mu_{\mathbf{H}}(w),$$

where $\mathfrak{c}$ runs over the cusps of Γ. We have seen in Theorem II.5.4 that there is a real number B not depending on Γ such that

$$\mathrm{gr}_\Gamma(z,w) \le B \quad \text{for all } z, w \in Y.$$

From this we get

$$\int_{w\in Y} \mathrm{gr}_\Gamma(z,w) F_\Gamma(w) \mu_{\mathbf{H}}(w) \le B \sup_Y F_\Gamma \int_Y \mu_{\mathbf{H}}.$$

We note that we obtained upper bounds on $\sup_Y F_\Gamma$ in § II.4.2.

For every cusp $\mathfrak{c}$, the results of §§ II.5.5 and II.4.3 show that furthermore

$$\mathrm{gr}_\Gamma(z,w) \le \frac{1}{\mathrm{vol}_\Gamma} \log(\epsilon_\mathfrak{c} y_\mathfrak{c}(w)) + B \quad \text{for all } z \in Y \text{ and } w \in B_\mathfrak{c}(\epsilon_\mathfrak{c})$$

and

$$F_\Gamma(w) \le (\epsilon_\mathfrak{c} y_\mathfrak{c}(w))^2 \exp(4\pi/\epsilon_\mathfrak{c} - 4\pi y_\mathfrak{c}(w)) \sup_Y F_\Gamma \quad \text{for all } w \in B_\mathfrak{c}(\epsilon_\mathfrak{c}).$$

We recall that $B_\mathfrak{c}(\epsilon_\mathfrak{c})$ is the image of the strip $\{x + iy \mid 0 \le x < 1 \text{ and } y > 1/\epsilon\}$ under the map $\mathbf{H} \to \Gamma\backslash\mathbf{H}$ sending z to $\Gamma\sigma_\mathfrak{c} z$. Using the above bounds and integrating over $B_\mathfrak{c}(w)$, we therefore get

$$\begin{aligned}\int_{w\in B_\mathfrak{c}(\epsilon_\mathfrak{c})} \mathrm{gr}_\Gamma(z,w) F_\Gamma(w)\mu_{\mathbf{H}}(w) &\le \epsilon_\mathfrak{c}^2 \sup_Y F_\Gamma \\ &\qquad \cdot \int_{1/\epsilon_\mathfrak{c}}^\infty \Bigl(\frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_\mathfrak{c} y) + B\Bigr) \exp(4\pi/\epsilon_\mathfrak{c} - 4\pi y) dy \\ &= \frac{\epsilon_\mathfrak{c}^2}{4\pi} \frac{\sup_Y F_\Gamma}{\mathrm{vol}_\Gamma} \int_0^\infty \log\Bigl(1 + \frac{\epsilon_\mathfrak{c}}{4\pi} u\Bigr) \exp(-u) du \\ &\qquad + \frac{\epsilon_\mathfrak{c}^2}{4\pi} B \sup_Y F_\Gamma,\end{aligned}$$

using the substitution

$$u = 4\pi y - 4\pi/\epsilon_\mathfrak{c}.$$

The integral on the right-hand side can be bounded using Jensen's inequality on convex functions:

$$\begin{aligned}\int_0^\infty \log(1 + au)\exp(-u) du &\le \log \int_0^\infty (1 + au)\exp(-u) du \\ &= \log(\Gamma(1) + a\Gamma(2)) \\ &= \log(1 + a).\end{aligned} \tag{3.2}$$

This gives

$$\int_{w\in B_\mathfrak{c}(\epsilon_\mathfrak{c})} \mathrm{gr}_\Gamma(z,w) F_\Gamma(w) \mu_{\mathbf{H}}(w) \le \frac{\epsilon_\mathfrak{c}^2}{4\pi} \sup_Y F_\Gamma \Bigl(\frac{1}{\mathrm{vol}_\Gamma} \log\Bigl(1 + \frac{\epsilon_\mathfrak{c}}{4\pi}\Bigr) + B\Bigr).$$

Summing the contributions from Y and from the discs around the cusps, we get the upper bound

$$\sup_Y h_\Gamma \leq \frac{\sup_Y F_\Gamma}{g_X}\Bigl(B\int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\Bigl(\frac{1}{\mathrm{vol}_\Gamma}\log\Bigl(1+\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)+B\Bigr)\Bigr). \tag{3.3}$$

For the lower bound, we do a computation that is identical except that by Theorem II.5.4 we get an extra (negative) term

$$S(z) = \frac{1}{g_X}\int\limits_{\substack{w\in\Gamma\backslash\mathbf{H}\\ d(z,w)\leq\delta}} (k_1(\delta) - k_1(d(z,w)))F_\Gamma(w)\mu_{\mathbf{H}}(w),$$

where

$$k_1(u) = \frac{1}{4\pi}\log\frac{u+1}{u-1}.$$

This term can be bounded as

$$S(z) \geq \frac{\sup_X F_\Gamma}{g_X}\int\limits_{d(z,w)\leq\delta} (k_1(\delta) - k_1(d(z,w)))\mu_{\mathbf{H}}(w);$$

we note that an upper bound for $\sup_X F_\Gamma$ is given by Lemma II.4.1. The integral can be evaluated as follows:

$$\begin{aligned}\int\limits_{\substack{w\in\Gamma\backslash\mathbf{H}\\ d(z,w)\leq\delta}} (k_1(\delta) - k_1(d(z,w)))\mu_{\mathbf{H}}(w) &= 2\pi(\delta-1)k_1(\delta) - 2\pi\int_1^\delta k_1(u)du\\ &= -\log\frac{\delta+1}{2}.\end{aligned}$$

This gives the lower bound

$$\begin{aligned}\inf_Y h_\Gamma \geq \frac{\sup_Y F_\Gamma}{g_X}&\Bigl(A\int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\Bigl(\frac{1}{\mathrm{vol}_\Gamma}\log\Bigl(1+\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)+A\Bigr)\Bigr)\\ &- \frac{\sup_X F_\Gamma}{g_X}\log\frac{\delta+1}{2}.\end{aligned} \tag{3.4}$$

We now extend our bounds on h_Γ to the discs $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$. By construction, h_Γ satisfies the differential equation

$$2i\partial\bar\partial h_\Gamma = \mu_X^{\mathrm{can}} - \frac{1}{\mathrm{vol}_\Gamma}\mu_{\mathbf{H}}.$$

This implies that h_Γ can be written on $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ as

$$h_\Gamma(z) = \int\limits_{w\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \mathrm{gr}_{\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(z,w)\mu_X^{\mathrm{can}}(w) + \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_{\mathfrak{c}}y_{\mathfrak{c}}(z)) + H(z),$$

where $\mathrm{gr}_{\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}$ is the Green function on the closed disc $\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ as defined at the end of § II.5.5, and where H is the unique harmonic function on the compactification $\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ that is equal to h_Γ on the boundary. It follows from the non-positivity of $\mathrm{gr}_{\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}$ and the maximum principle for harmonic functions that

$$h_\Gamma(z) \le \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z)) + \sup_{\partial\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} h_\Gamma \quad \text{for all } z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}). \tag{3.5}$$

Similarly, it follows from the bound for F_Γ given in § II.4.3 and the differential equation satisfied by $\mathrm{gr}_{\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}$ in § II.5.5 that

$$\begin{aligned}\int\limits_{w\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \mathrm{gr}_{\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(z,w)\mu_X^{\mathrm{can}}(w) &\ge \frac{\sup_Y F_\Gamma}{g_X}\epsilon_{\mathfrak{c}}^2\exp(4\pi/\epsilon_{\mathfrak{c}}) \\ &\quad\cdot \int\limits_{w\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \mathrm{gr}_{\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(z,w)y_{\mathfrak{c}}(w)^2\exp(-4\pi y_{\mathfrak{c}}(w))\mu_{\mathbf{H}}(w) \\ &= \frac{\sup_Y F_\Gamma}{g_X}\epsilon_{\mathfrak{c}}^2\exp(4\pi/\epsilon_{\mathfrak{c}}) \\ &\quad\cdot\frac{1}{(4\pi)^2}\bigl(\exp(-4\pi y_{\mathfrak{c}}(z)) - \exp(-4\pi/\epsilon_{\mathfrak{c}})\bigr) \\ &= -\frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2\bigl(1-\exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}}(z))\bigr) \\ &\ge -\frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2.\end{aligned} \tag{3.6}$$

Therefore a lower bound is given by

$$h_\Gamma(z) \ge \inf_{\partial\bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} h_\Gamma - \frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2 + \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z)) \quad \text{for all } z \in \bar{B}_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}). \tag{3.7}$$

3.4. Bounds on the integral $\int_{\Gamma\backslash\mathbf{H}} h_\Gamma\mu_X^{\mathrm{can}}$

We are now going to bound the constant term in (3.1), which we split up as

$$\int_{\Gamma\backslash\mathbf{H}} h_\Gamma\mu_X^{\mathrm{can}} = \int_Y h_\Gamma\mu_X^{\mathrm{can}} + \sum_{\mathfrak{c}}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} h_\Gamma\mu_X^{\mathrm{can}}.$$

Plugging in the upper bound (3.5), we obtain

$$\int_{\Gamma\backslash\mathbf{H}} h_\Gamma\mu_X^{\mathrm{can}} \le \sup_Y h_\Gamma\int_Y\mu_X^{\mathrm{can}} + \sum_{\mathfrak{c}}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl(\sup_{\partial\bar{B}_{\mathfrak{c}}(m_{\mathfrak{c}}\epsilon)} h_\Gamma + \frac{1}{\mathrm{vol}_\Gamma}\log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}})\Bigr)\mu_X^{\mathrm{can}}.$$

Next we use the fact that $\int_X \mu_X^{\mathrm{can}} = 1$, the equation (1.8), which relates μ^{can} to $\mu_{\mathbf{H}}$ via the function F_Γ, and the bounds

$$F_\Gamma(z) \le \begin{cases}\sup_Y F_\Gamma & \text{if } z\in Y;\\ (\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z))^2\exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}}(z))\sup_Y F_\Gamma & \text{if } z\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})\end{cases}$$

proved in §§ II.4.2 and II.4.3. This gives

$$\int_{\Gamma\backslash\mathbf{H}} h_\Gamma \mu_X^{\mathrm{can}} \le \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma} \sum_{\mathfrak{c}} \int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}})(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}})^2 \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}})\mu_{\mathbf{H}} + \sup_Y h_\Gamma.$$

The integral over $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ can be bounded as follows:

$$\begin{aligned}
\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}})(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}})^2 \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}})\mu_{\mathbf{H}} &= \epsilon_{\mathfrak{c}}^2 \int_{1/\epsilon_{\mathfrak{c}}}^\infty \log(\epsilon_{\mathfrak{c}} y) \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y) dy \\
&= \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \int_0^\infty \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi} u\Bigr) \exp(-u) du \\
&\le \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr),
\end{aligned}$$

where the last inequality follows from (3.2). We therefore get

$$\int_{\Gamma\backslash\mathbf{H}} h_\Gamma \mu_X^{\mathrm{can}} \le \sup_Y h_\Gamma + \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma} \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr).$$

Finally, plugging in the upper bound (3.3) for h_Γ, we conclude that

$$\begin{aligned}
\int_{\Gamma\backslash\mathbf{H}} h_\Gamma \mu_X^{\mathrm{can}} \le{}& \frac{\sup_Y F_\Gamma}{g_X} \Bigl(B \int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \Bigl(\frac{1}{\operatorname{vol}_\Gamma} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) + B\Bigr)\Bigr) \\
&+ \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma} \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr).
\end{aligned} \tag{3.8}$$

An entirely analogous computation using (3.7) and (3.4) leads to the lower bound

$$\begin{aligned}
\int_{\Gamma\backslash\mathbf{H}} h_\Gamma \mu_X^{\mathrm{can}} \ge{}& \frac{\sup_Y F_\Gamma}{g_X} \Bigl(A \int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \Bigl(\frac{1}{\operatorname{vol}_\Gamma} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) + A\Bigr)\Bigr) \\
&- \Bigl(\frac{\sup_Y F_\Gamma}{g_X}\Bigr)^2 \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^4}{(4\pi)^3} - \frac{\sup_X F_\Gamma}{g_X} \log\frac{\delta + 1}{2}.
\end{aligned} \tag{3.9}$$

3.5. Bounds on canonical Green functions

From (3.1), the upper bound for gr_Γ from Theorem II.5.4, the lower bound for h_Γ given by (3.4) and the bound (3.8), we can now conclude that

$$\begin{aligned}
\sup_{Y\times Y} \mathrm{gr}_X^{\mathrm{can}} \le{}& B + \min\{0, k_1(\delta) - k_1(d(z,w))\} \\
&- \frac{2\sup_Y F_\Gamma}{g_X} \Bigl(A \int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \Bigl(\frac{1}{\operatorname{vol}_\Gamma} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) + A\Bigr)\Bigr) \\
&+ \frac{\sup_Y F_\Gamma}{g_X} \Bigl(B \int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \Bigl(\frac{1}{\operatorname{vol}_\Gamma} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) + B\Bigr)\Bigr) \\
&+ \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma} \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) + 2\frac{\sup_X F_\Gamma}{g_X} \log\frac{\delta + 1}{2}.
\end{aligned}$$

Simplifying this, we obtain

$$\sup_{Y\times Y} \mathrm{gr}_X^{\mathrm{can}} \le B + \min\{0, k_1(\delta) - k_1(d(z,w))\} + \frac{(B-2A)\sup_Y F_\Gamma}{g_X}\Bigl(\int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\Bigr) + 2\frac{\sup_X F_\Gamma}{g_X}\log\frac{\delta+1}{2}. \tag{3.10}$$

Similarly, combining the lower bound for gr_Γ from Theorem II.5.4 with the inequalities (3.3) and (3.9) leads to the following bound for all $z, w \in Y$:

$$\begin{aligned}\mathrm{gr}_X^{\mathrm{can}}(z,w) \ge A &+ \min\{0, k_1(\delta) - k_1(d(z,w))\} \\ &+ \frac{(A-2B)\sup_Y F_\Gamma}{g_X}\Bigl(\int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\Bigr) \\ &- \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma}\sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) \\ &- \Bigl(\frac{\sup_Y F_\Gamma}{g_X}\Bigr)^2 \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^4}{(4\pi)^3} - \frac{\sup_X F_\Gamma}{g_X}\log\frac{\delta+1}{2}\end{aligned} \tag{3.11}$$

We now imitate § II.5.5 to extend the above bounds on the canonical Green function $\mathrm{gr}_X^{\mathrm{can}}(x,y)$ to the case where one or both of x and y lies in a neighbourhood of a cusp $\mathfrak{c}$ of Γ. For any y not in the disc $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$, we consider $\mathrm{gr}_X^{\mathrm{can}}(x,y)$ as a function of $x \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$. This function satisfies

$$2i\partial\bar\partial\,\mathrm{gr}_X^{\mathrm{can}}(x,y) = -\mu_X^{\mathrm{can}}(x),$$

so we can write

$$\mathrm{gr}_X^{\mathrm{can}}(x,y) = -\int\limits_{z\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \mathrm{gr}_{\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(x,z)\mu_X^{\mathrm{can}}(x) + h_y(x) \quad \text{for all } x \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}),$$

where $\mathrm{gr}_{\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}$ is the Green function on the disc $\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ as defined in § II.5.5. By construction, the function $h_y(x)$ coincides with $\mathrm{gr}_X^{\mathrm{can}}(x,y)$ for x on the boundary of $\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$. The inequality (3.6) now implies that

$$\mathrm{gr}_X^{\mathrm{can}}(x,y) \le \frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2 + \sup_{z\in Y}\mathrm{gr}_X^{\mathrm{can}}(z,y) \quad \text{for all } x \in \bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}), y \notin \bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}).$$

Finally, considering the case where x and y are both in a disc $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$, we get

$$\mathrm{gr}_X^{\mathrm{can}}(x,y) \le \mathrm{gr}_{\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(x,y) + \frac{2\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2 + \sup_{Y\times Y}\mathrm{gr}_X^{\mathrm{can}} \tag{3.12}$$

for all $x, y \in \bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$.

3.6. A lower bound for the function H_Γ

Now that we have a lower bound for $\mathrm{gr}_X^{\mathrm{can}}$, we can deduce a lower bound for the function H_Γ defined by (1.9). Namely, it follows immediately from (1.9) and (3.11) that

$$\begin{aligned}\inf_Y H_\Gamma \geq A + k_1(\delta) &+ \frac{(A-2B)\sup_Y F_\Gamma}{g_X}\Bigl(\int_Y \mu_{\mathbf{H}} + \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\Bigr) \\ &- \frac{\sup_Y F_\Gamma}{g_X \operatorname{vol}_\Gamma} \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^2}{4\pi} \log\Bigl(1 + \frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr) - \Bigl(\frac{\sup_Y F_\Gamma}{g_X}\Bigr)^2 \sum_{\mathfrak{c}} \frac{\epsilon_{\mathfrak{c}}^4}{(4\pi)^3} \\ &- \frac{\sup_X F_\Gamma}{g_X} \log \frac{\delta+1}{2}.\end{aligned} \tag{3.13}$$

We extend this to the cusps using the differential equation

$$2i\partial\bar\partial H_\Gamma = (2g_X - 2)\mu_X^{\mathrm{can}} - \frac{1}{2\pi}\mu_{\mathbf{H}} + \sum_{\mathfrak{c}\text{ cusp}} \delta_{\mathfrak{c}}.$$

proved in Lemma 1.1. This differential equation implies that

$$H_\Gamma(z) = (2g_X - 2)\int_{w\in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \mathrm{gr}_{\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}(z,w)\mu_X^{\mathrm{can}}(w) + \frac{1}{2\pi}\log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z)) + \frac{1}{\epsilon_{\mathfrak{c}}} - y_{\mathfrak{c}}(z) + h(z)$$

for $z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$, where h is the unique harmonic function on $\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$ that coincides with H_Γ on the boundary of $\bar B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$. By (3.6) and the minimum principle for harmonic functions, we get

$$H_\Gamma(z) \geq -(2g_X - 2)\frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2 + \frac{1}{2\pi}\log(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z)) + \frac{1}{\epsilon_{\mathfrak{c}}} - y_{\mathfrak{c}}(z) + \inf_Y H_\Gamma \tag{3.14}$$

for all $z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$.

3.7. An upper bound for the integral $\int_X \log|\alpha|_{\Omega^1_{X/\mathbf{C}}} \mu_X^{\mathrm{can}}$

Let α be a non-zero element of $\mathrm{H}^0(X, \Omega^1_{X/\mathbf{C}})$. We are interested in an upper bound for the integral

$$I(\alpha) = \int_X \log|\alpha|_{\Omega^1_{X/\mathbf{C}}} \mu_X^{\mathrm{can}}$$

in terms of the norm

$$\langle\alpha,\alpha\rangle = \frac{i}{2}\int_X \alpha \wedge \bar\alpha.$$

Let f be the element of $\mathrm{S}_2(\Gamma)$ corresponding to α via the isomorphism (1.6). We rewrite $I(\alpha)$ using Lemma 1.1 as

$$I(\alpha) = \int_X \log|f|_{2,\mathrm{Pet}}\mu_X^{\mathrm{can}} + \log 2 - 2\pi\int_X H_\Gamma \mu_X^{\mathrm{can}}. \tag{3.15}$$

Jensen's inequality implies that

$$\int_{\Gamma\backslash\mathbf{H}} \log\bigl(|f|^2_{2,\mathrm{Pet}}\bigr)\mu_X^{\mathrm{can}} \le \log\int_{\Gamma\backslash\mathbf{H}} |f|^2_{2,\mathrm{Pet}}\mu_X^{\mathrm{can}}.$$

From (1.8) we now get

$$\begin{aligned}\int_{\Gamma\backslash\mathbf{H}} |f|^2_{2,\mathrm{Pet}}\mu_X^{\mathrm{can}} &\le \frac{\sup_X F_\Gamma}{g_X}\int_{\Gamma\backslash\mathbf{H}} |f|^2_{2,\mathrm{Pet}}\mu_{\mathbf{H}} \\ &= \frac{\sup_X F_\Gamma}{g_X}\langle f,f\rangle_\Gamma \qquad (3.16)\\ &= \frac{\sup_X F_\Gamma}{g_X}\langle \alpha,\alpha\rangle.\end{aligned}$$

Furthermore, the bound (3.14) implies

$$\begin{aligned}\int_X H_\Gamma\mu_X^{\mathrm{can}} &\ge \int_Y (\inf_Y H_\Gamma)\mu_X^{\mathrm{can}} + \sum_{\mathfrak{c}}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl(\inf_Y H_\Gamma - (2g_X-2)\frac{\sup_Y F_\Gamma}{g_X}\Bigl(\frac{\epsilon_{\mathfrak{c}}}{4\pi}\Bigr)^2 \\ &\qquad + \frac{1}{2\pi}\log(\epsilon_{\mathfrak{c}}y_{\mathfrak{c}}(z)) + \frac{1}{\epsilon_{\mathfrak{c}}} - y_{\mathfrak{c}}(z)\Bigr)\mu_X^{\mathrm{can}} \\ &\ge \inf_Y H_\Gamma - \sum_{\mathfrak{c}}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl((2g_X-2)\frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2} \\ &\qquad - \frac{1}{2\pi}\log(\epsilon_{\mathfrak{c}}y_{\mathfrak{c}}) - \frac{1}{\epsilon_{\mathfrak{c}}} + y_{\mathfrak{c}}\Bigr)\mu_X^{\mathrm{can}} \\ &\ge \inf_Y H_\Gamma - \sum_{\mathfrak{c}}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl((2g_X-2)\frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2} - \frac{1}{\epsilon_{\mathfrak{c}}} + y_{\mathfrak{c}}\Bigr)\mu_X^{\mathrm{can}} \\ &= \inf_Y H_\Gamma - \sum_{\mathfrak{c}}\Bigl((2g_X-2)\frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\mu_X^{\mathrm{can}} \\ &\qquad + \int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl(y_{\mathfrak{c}} - \frac{1}{\epsilon_{\mathfrak{c}}}\Bigr)\mu_X^{\mathrm{can}}\Bigr).\end{aligned}$$

The integrals can be bounded as follows:

$$\begin{aligned}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\mu_X^{\mathrm{can}} &= \frac{1}{g_X}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} F_\Gamma\mu_{\mathbf{H}} \\ &\le \frac{\sup_Y F_\Gamma}{g_X}\epsilon_{\mathfrak{c}}^2\int_{1/\epsilon_{\mathfrak{c}}}^\infty \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y)dy \\ &= \frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{4\pi}\end{aligned}$$

and similarly

$$\begin{aligned}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl(y_{\mathfrak{c}} - \frac{1}{\epsilon_{\mathfrak{c}}}\Bigr)\mu_X^{\mathrm{can}} &= \frac{1}{g_X}\int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})}\Bigl(y_{\mathfrak{c}} - \frac{1}{\epsilon_{\mathfrak{c}}}\Bigr)F_\Gamma\mu_{\mathbf{H}} \\ &\le \frac{\sup_Y F_\Gamma}{g_X}\epsilon_{\mathfrak{c}}^2\int_{1/\epsilon_{\mathfrak{c}}}^\infty \Bigl(y - \frac{1}{\epsilon_{\mathfrak{c}}}\Bigr)\exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y)dy \\ &= \frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2}.\end{aligned}$$

This gives

$$\begin{aligned}\int_X H_\Gamma\mu^{\mathrm{can}} &\ge \inf_Y H_\Gamma - \sum_{\mathfrak{c}}\biggl((2g_X-2)\Bigl(\frac{\sup_Y F_\Gamma}{g_X}\Bigr)^2\frac{\epsilon_{\mathfrak{c}}^4}{(4\pi)^3} + \frac{\sup_Y F_\Gamma}{g_X}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2}\biggr)\\ &= \inf_Y H_\Gamma - (2g_X-2)\Bigl(\frac{\sup_Y F_\Gamma}{g_X}\Bigr)^2\sum_{\mathfrak{c}}\frac{\epsilon_{\mathfrak{c}}^4}{(4\pi)^3}\\ &\qquad - \frac{\sup_Y F_\Gamma}{g_X}\sum_{\mathfrak{c}}\frac{\epsilon_{\mathfrak{c}}^2}{(4\pi)^2}.\end{aligned}\tag{3.17}$$

From (3.15), (3.16), (3.17) and (3.13), we now get the desired upper bound for $I(\alpha)$. Because the resulting formula is not very enlightening, we do not write it down.

4. Intersection theory at the finite places

Let R be a complete discrete valuation ring with field of fractions K and algebraically closed residue field k. Let X be a proper, smooth and geometrically connected curve over K. To X there is attached a graph G_X describing the system of all regular and semi-stable models of X over finite extensions of R. This G_X will be a *metrised graph*, as will be defined below. In the spirit of Zhang [116], we will describe the relevance of such graphs for arithmetic intersection theory.

4.1. Metrised graphs

We first make precise what we mean by piecewise smooth functions. We then define metrised graphs in essentially the same way as Zhang [116, Appendix], and we define the Laplace operator and the corresponding Green function on a metrised graph.

By an *interval* we mean a subset of $\mathbf{R}$ of the form $\{x \in \mathbf{R} \mid a \le x \le b\}$, where $a < b$ are real numbers. Let $I \subset \mathbf{R}$ be an interval, and let $f\colon I \to \mathbf{R}$ be a continuous function. Then f is called *piecewise linear* (resp. *piecewise smooth*) if I can be written as a finite union of intervals $I_1, \ldots, I_k$ such that the restriction of f to each I_k is linear (resp. infinitely continuously differentiable). Here “differentiable” means “left (resp. right) differentiable” at the endpoints.

Definition. A *metrised graph* is a topological space G with a measure μ such that G is isomorphic to a quotient of a finite disjoint union of intervals $I_1, \ldots, I_n$ by some equivalence relation on the set of endpoints, and such that μ is induced from the Lebesgue measure on the I_k.

If G is a metrised graph given as a quotient

$$q\colon I_1 \sqcup \ldots \sqcup I_n \to G$$

by an equivalence relation on the endpoints, the continuous functions $G \to \mathbf{R}$ are (by definition of the quotient) the continuous functions on $I_1 \sqcup \ldots \sqcup I_n$ that respect the equivalence relation. The $\mathbf{R}$-vector space of *piecewise linear functions* on G, denoted by $\mathrm{PL}(G)$, is the space of continuous functions $f\colon G \to \mathbf{R}$ such that the restriction of f to each I_k is piecewise linear. The $\mathbf{R}$-vector space $\mathrm{PS}(G)$ of *piecewise smooth functions* is defined analogously.

By construction, a metrised graph G has a natural measure μ. We write vol_G for the volume of G with respect to μ. Furthermore, we have the *Laplace operator*

$$\Delta\colon \mathrm{PS}(G) \to \mathrm{PS}(G)^\vee.$$

This is an $\mathbf{R}$-linear map which is positive semi-definite in the sense that

$$(\Delta f)(f) \geq 0 \quad \text{for all } f \in \mathrm{PS}(G);$$

the kernel of Δ consists of the locally constant functions. For the definition we refer to Zhang [116, Appendix].

From now on we assume for simplicity that G is connected. The *Green function* for the Laplace operator on G is the unique continuous function

$$\mathrm{gr}_G\colon G \times G \to \mathbf{R}$$

that is symmetric, piecewise smooth in both variables, and satisfies the differential equation

$$-\Delta\,\mathrm{gr}_G(p,q) = \delta_q - \frac{1}{\mathrm{vol}_{G_X}}\mu \quad \text{and} \quad \int_{p\in G} \mathrm{gr}_G(p,q)\mu(p) = 0 \qquad \text{for all } q \in G.$$

(In keeping with our convention for the other Green functions employed in this thesis, our Green function is minus that of Zhang.) We also define

$$\begin{aligned} g_{q,r}\colon G &\to \mathbf{R} \\ p &\mapsto \mathrm{gr}_G(p,q) - \mathrm{gr}_G(p,r). \end{aligned}$$

Then $g_{q,r}$ is the unique function satisfying

$$-\Delta g_{q,r} = \delta_q - \delta_r \quad \text{and} \quad \int_{p\in G} g_{q,r}(p)\mu(p) = 0 \qquad \text{for all } q, r \in G.$$

By viewing G as a one-dimensional object made of electrically conducting material and $g_{q,r}$ as the potential function corresponding to point charges $+1$ and -1 at q and r, we see that

$$\begin{aligned} \sup_G g_{q,r} - \inf_G g_{q,r} &= g_{q,r}(r) - g_{q,r}(q) \\ &\leq d(q,r), \end{aligned}$$

where d is the distance between q and r. Since $\inf_G g_{q,r} \leq 0$, we get

$$\begin{aligned} \sup_{p,q,r\in G} \bigl(\mathrm{gr}_G(p,q) - \mathrm{gr}_G(p,r)\bigr) &= \sup_{p,q,r\in G} g_{q,r}(p) \\ &\leq \mathrm{diam}(G), \end{aligned}$$

where $\mathrm{diam}(G)$ denotes the diameter of G.

4.2. Reduction graphs

Let R be a complete discrete valuation ring with field of fractions K and algebraically closed residue field k. Let X be a proper, smooth and geometrically connected curve over K. We associate to X a metrised graph G_X in the following way. By the semi-stable reduction theorem [22], there exists a finite extension K' of K such that $X\times_{\operatorname{Spec} K}\operatorname{Spec} K'$ has a regular and semi-stable model $X_{R'}$ over the integral closure R' of R in K'. We can identify the residue field of R' with k, and we write $e(K'/K)$ for the ramification index of K' over K. Furthermore, we write

$$\tilde X_{K'} = X_{R'}\times_{\operatorname{Spec} R'}\operatorname{Spec} k,$$

and we let $V(\tilde X_{K'})$ denote the set of irreducible components of $\tilde X_{K'}$. We take a set of intervals of length $1/e(K'/K)$ in $\mathbf{R}$ indexed by the set of singular points of $\tilde X_{K'}(k)$, and we label the endpoints of the interval corresponding to a singular point x by the two irreducible components on which x lies; these are possibly equal. For each $C\in V(\tilde X_{K'})$ we identify the set of endpoints labelled C. The result is by definition a metrised graph $G(\tilde X_{K'})$. We may identify $V(\tilde X_{K'})$ with a finite subset of $G(\tilde X_{K'})$. If $K\subseteq K'\subseteq K''$ are finite extensions such that X has semi-stable reduction over K', and if $\tilde X_{K'}$ and $\tilde X_{K''}$ are the corresponding regular and semi-stable models, then there is a canonical isomorphism

$$G(\tilde X_{K'})\cong G(\tilde X_{K''})$$

of metrised graphs. We may therefore denote the graph by G_X, the choice of an extension K' of K being understood. We call G_X the *reduction graph* of X.

We define a non-negative real number $\gamma(X)$ as

$$\gamma(X)=\sup_{x,y,z\in G_X}\bigl(\mathrm{gr}_{G_X}(x,y)-\mathrm{gr}_{G_X}(x,z)\bigr),$$

where gr_{G_X} is the Green function of the metrised graph G_X. It follows from the results of § 4.1 that

$$\gamma(X)\le\operatorname{diam}(G_X).$$

For any finite extension K' of K over which X has semi-stable reduction, we define the finite dimensional $\mathbf{R}$-vector space

$$D(\tilde X_{K'})=\mathbf{R}^{V(\tilde X_{K'})}$$

of formal $\mathbf{R}$-linear combinations of the irreducible components of the special fibre. The intersection pairing between irreducible components gives rise to an $\mathbf{R}$-linear map

$$\begin{aligned}M\colon D(\tilde X_{K'})&\longrightarrow D(\tilde X_{K'})\\ \sum_C a_C C&\longmapsto\sum_C\Bigl(\sum_{C'}(C\,.\,C')a_{C'}\Bigr)C.\end{aligned}$$

Lemma 4.1. *Consider an element $w\in D(\tilde X_{K'})$ of the form*

$$w=\sum_{C\in V(\tilde X_{K'})}b_C C\quad\textit{with}\quad\sum_{C\in V(\tilde X_{K'})}b_C=0.$$

Then there is an element

$$v = \sum_{C \in V(\tilde{X}_{K'})} a_C C \in D(\tilde{X}_{K'})$$

such that

$$Mv = w.$$

This v is unique up to addition of a multiple of $\sum_C C$. For any v as above the inequality

$$\max_C a_C - \min_C a_C \leq 2e(K'/K)\gamma(X) \sum_{C' : b_{C'} > 0} b_{C'}$$

holds for all $C \in V(\tilde{X}_{K'})$.

Proof. The existence of v and its uniqueness up to addition of multiples of $\sum_C C$ follow from the symmetry of the matrix M and the fact that the kernel of M is spanned by $\sum_C C$. For the bound on the a_C, we use the inclusion

$$i \colon D(\tilde{X}_{K'}) \rightarrowtail \mathrm{PS}(G_X)$$

sending an element $\sum_{C \in V(\tilde{X}_{K'})} a_C C \in D(\tilde{X}_{K'})$ to the unique continuous function that takes the value a_C at C for every $C \in V(\tilde{X}_{K'})$ and is linear outside $V(\tilde{X}_{K'})$. There is a second inclusion

$$\begin{aligned} j \colon D(\tilde{X}_{K'}) &\rightarrowtail \mathrm{PS}(G_X)^\vee \\ \sum_C a_C C &\mapsto \sum_C a_C \delta_C, \end{aligned}$$

where δ_C is the Dirac δ-distribution at C, defined by

$$\delta_C(f) = f(C).$$

The above maps fit in a commutative diagram

$$\begin{array}{ccc} D(\tilde{X}_{K'}) & \xrightarrow{i} & \mathrm{PS}(G_X) \\ {\scriptstyle -e(K'/K)M}\big\downarrow & & \big\downarrow{\scriptstyle \Delta} \\ D(\tilde{X}_{K'}) & \xrightarrow{j} & \mathrm{PS}(G_X)^\vee. \end{array}$$

This can be seen by means of a straightforward calculation going along the same lines as Zhang [116, (a.5)]. The assumption that $\sum_{C'} b_{C'} = 0$ implies that one solution v of $Mv = w$ is given by

$$v = \sum_{C \in V(X_{K'})} a_C C,$$

where

$$a_C = e(K'/K) \sum_{C' \in V(X_{K'})} b_{C'} \operatorname{gr}_{G_X}(C, C').$$

In particular, this implies

$$\begin{aligned}
|a_C| &\le e(K'/K) \sum_{C' : b_{C'} > 0} b_{C'} \sup_{G_X} \mathrm{gr}_{G_X}(C, \ ) + \sum_{C' : b_{C'} < 0} b_{C'} \inf_{G_X} \mathrm{gr}_{G_X}(C, \ ) \\
&= e(K'/K) \sum_{C' : b_{C'} > 0} b_{C'} \Bigl(\sup_{G_X} \mathrm{gr}_{G_X}(C, \ ) - \inf_{G_X} \mathrm{gr}_{G_X}(C, \ ) \Bigr) \\
&\le e(K'/K)\gamma(X) \sum_{C' : b_{C'} > 0} b_{C'}.
\end{aligned}$$

The proposition follows since $\max_C a_C - \min_C a_C$ is independent of the choice of v. $\square$

5. Bounds on some Arakelov-theoretic invariants of modular curves

For every positive integer n, let $\mathrm{X}_1(n)$ denote the coarse moduli space for the modular stack $\mathcal{M}_{\Gamma_1(n)}$ over $\operatorname{Spec} \mathbf{Z}$ defined in § I.1.1. We only consider n such that that the fibres of $\mathrm{X}_1(n)$ are of genus at least 1. In this section we will find bounds on certain Arakelov invariants of the arithmetic surface $\mathrm{X}_1(n)$.

5.1. Self-intersection of the relative dualising sheaf

We consider the map

$$T \colon \operatorname{Spec} \mathbf{Z}[[q]] \to \mathrm{X}_1(n)$$

corresponding to the Tate curve $\underline{\mathrm{Tate}}(q^n)$ over $\operatorname{Spec} \mathbf{Z}[[q]]$ together with the n-torsion point q modulo q^n, as defined in § I.2.4. The zero locus of q gives a section

$$O \colon \operatorname{Spec} \mathbf{Z} \to \mathrm{X}_1(n).$$

Although $\mathrm{X}_1(n)$ is not semi-stable, the image of O lies in the open subset where the morphism $\mathrm{X}_1(n) \to \operatorname{Spec} \mathbf{Z}$ is smooth and has reduced fibres, so in this open subset the relative dualising sheaf exists and coincides with the line bundle of differentials. This means that $O^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}}$ is a metrised line bundle on $\operatorname{Spec} \mathbf{Z}$. To find an upper bound for its degree, we use the fact that T is unramified, so that

$$\begin{aligned}
T^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}} &\cong \Omega^1_{\mathbf{Z}[[q]]/\mathbf{Z}} \\
&= \mathbf{Z}[[q]]dq.
\end{aligned}$$

This implies that $O^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}}$ is a free $\mathbf{Z}$-module of rank 1 generated by dq. We deduce from (2.2), (2.1) and (1.4) that

$$\begin{aligned}
\deg O^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}} &= -\log |dq_O|_{\Omega^1_{\mathrm{X}_1(n)/\mathbf{C}}}(O) \\
&= \lim_{z \to O} \bigl(2\pi \, \mathrm{gr}^{\mathrm{can}}_{\mathrm{X}_1(n)(\mathbf{C})}(z, O) - \log |q_O(z)| \bigr),
\end{aligned}$$

where q_O is the standard coordinate around the cusp O.

We choose a real number $\epsilon \in (0,1)$, and we write $B_\infty(\epsilon)$ for the standard disc of area ϵ around the unique cusp ∞ of $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$ as in § II.1.2, and we define Y_0 as the complement of $B_\infty(\epsilon)$ in $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$. Furthermore, we define a compact subset Y of $\mathrm{X}_1(n)(\mathbf{C})$ as the inverse image of Y_0 under the map $\Gamma_1(n)\backslash\mathbf{H} \to \mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$. Then the complement of Y is the disjoint union of the discs $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$, where $\mathfrak{c}$ runs over the cusps of $\Gamma_1(n)$ and $\epsilon_{\mathfrak{c}}$ is ϵ times the the ramification index at $\mathfrak{c}$. In particular, the fact that the ramification index at O equals n implies that $\epsilon_O = n\epsilon$. By (3.12) and the explicit formula for $\mathrm{gr}_{\bar{B}_O(\epsilon_O)}$ given in § II.5.5, we therefore have

$$\begin{aligned}\deg O^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}} &\le \lim_{z\to O}\bigl(2\pi\,\mathrm{gr}_{\bar{B}_O(n\epsilon)}(z,O) - \log|q_O(z)|\bigr) + \frac{\sup_Y F_\Gamma}{g_{\mathrm{X}_1(n)}}\frac{n^2\epsilon^2}{4\pi} \\ &\quad + 2\pi\sup_{Y\times Y}\mathrm{gr}^{\mathrm{can}}_{\mathrm{X}_1(n)(\mathbf{C})} \\ &= \frac{2\pi}{n\epsilon} + \frac{\sup_Y F_\Gamma}{g_{\mathrm{X}_1(n)}}\frac{n^2\epsilon^2}{4\pi} + 2\pi\sup_{Y\times Y}\mathrm{gr}^{\mathrm{can}}_{\mathrm{X}_1(n)(\mathbf{C})}\,.\end{aligned}$$

We now consider a number field K such that $\mathrm{X}_1(n)$ has a semi-stable model $\mathcal{X}$ over $\operatorname{Spec}\mathbf{Z}_K$, where $\mathbf{Z}_K$ is the ring of integers of K. We abbreviate

$$\Omega^2_{\mathrm{X}_1(n)/\mathbf{Z}} = \frac{1}{[K:\mathbf{Q}]}(\Omega_{\mathcal{X}/\mathbf{Z}_K}\,.\,\Omega_{\mathcal{X}/\mathbf{Z}_K})_{\mathcal{X}};$$

this does not depend on the choice of K. It follows from the Hodge index theorem for Arakelov's intersection pairing that

$$\Omega^2_{\mathrm{X}_1(n)/\mathbf{Z}} \le 4g_{\mathrm{X}_1(n)}(g_{\mathrm{X}_1(n)}-1)\deg O^*\Omega_{\mathrm{X}_1(n)/\mathbf{Z}};$$

see Faltings [37, Theorem 5]. This also gives us an upper bound for $\Omega^2_{\mathrm{X}_1(n)/\mathbf{Q},\mathrm{a}}$, the self-intersection of the relative dualising sheaf in the sense of Zhang, via the inequality

$$\Omega^2_{\mathrm{X}_1(n)/\mathbf{Q},\mathrm{a}} \le \Omega^2_{\mathrm{X}_1(n)/\mathbf{Z}}$$

from § 2.2.

5.2. Bounds on Green functions on reduction graphs of modular curves

We fix a positive integer a. We consider integers n of the form ab, where b is a squarefree positive integer coprime to a.

Let p be a prime number, let $\mathrm{W}(\overline{\mathbf{F}}_p)$ be the ring of Witt vectors of $\overline{\mathbf{F}}_p$, and let $W_p = \mathrm{W}(\overline{\mathbf{F}}_p)[1/p]$ be its field of fractions. We will study how $\gamma(\mathrm{X}_1(n)_{W_p})$ varies as a function of n. We distinguish several cases, depending on how often p divides n.

First we assume $p \nmid n$. Then $\mathrm{X}_1(n)$ has a smooth model over $\mathrm{W}(\overline{\mathbf{F}}_p)$, so we get

$$\gamma(\mathrm{X}_1(n)_{W_p}) = 0.$$

Next we assume p divides n exactly once and that $n/p \ge 5$. Then $\mathrm{X}_1(n)$ has a regular and semi-stable model over the tame extension $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$ of degree $p-1$ of $\mathrm{W}(\overline{\mathbf{F}}_p)$, and the special fibre is the union of two smooth curves intersecting transversally in m points for some positive integer m; see Katz and Mazur [53, Theorem 13.11.4]. This implies that the reduction graph of $\mathrm{X}_1(n)_{W_p}$ consists of two vertices connected by m edges of length $1/(p-1)$, so the observation in § 4.2 that $\gamma(\mathrm{X}_1(n)_{W_p})$ is bounded above by the diameter of the reduction graph implies that

$$\gamma(\mathrm{X}_1(n)_{W_p}) \le \frac{1}{p-1}.$$

Remark. In fact, one can explicitly compute the Green function of the reduction graph and thereby show that

$$\gamma(\mathrm{X}_1(n)_{W_p}) = \frac{1}{4(p-1)}.$$

We continue with the case where p divides n exactly once and $n/p \le 4$. The assumption that $\mathrm{X}_1(n)$ has genus at least 1 implies that we have the following possibilities:

(1) $n = p \ge 11$;

(2) $n = 2p$ and $p \ge 7$;

(3) $n = 3p$ and $p \ge 5$;

(4) $n = 4p$ and $p \ge 5$.

As before, $\mathrm{X}_1(n)$ has a semi-stable model over $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$ consisting of two smooth curves intersecting transversally in a finite number of points. This model is, however, not necessarily regular, since certain supersingular points in the special fibre of $\mathrm{X}_1(n)$ over $\mathrm{W}(\overline{\mathbf{F}}_p)$ correspond to objects with extra automorphisms. These are supersingular elliptic curves over $\overline{\mathbf{F}}_p$ with j-invariant 0 (in which case $p \equiv 2 \bmod 3$) or 1728 (in which case $p \equiv 3 \bmod 4$), together with a torsion point of order n/p. Let x be such a non-regular supersingular point, and let G be the automorphism group of the corresponding object. Then G is cyclic of order g, where the possibilities for g are given by the table below.

	$n = p$	$n = 2p$	$n = 3p$	$n = 4p$
$j \equiv 0 \bmod p$ and $p \equiv 2 \bmod 3$	6	2	1 or 3	1
$j \equiv 1728 \bmod p$ and $p \equiv 3 \bmod 4$	4	2 or 4	1	1

We choose a moduli problem $\mathcal{P}$ on elliptic curves over $\mathrm{W}(\overline{\mathbf{F}}_p)$ that is representable, finite étale and Galois with group G. Then we have a finite surjective morphism

$$\mathrm{X}(\mathcal{P};\Gamma_1(n)) \to \mathrm{X}_1(n),$$

where $\mathrm{X}(\mathcal{P};\Gamma_1(n))$ is the fine moduli scheme classifying elliptic curves together with a $\mathcal{P}$-structure and $\Gamma_1(n)$-structure. This is a regular two-dimensional $\mathrm{W}(\mathbf{F}_p)$-scheme. For any point x as above, we choose a point $\tilde{x}$ mapping to x. Let $G_{\tilde{x}}$ denote the stabiliser of $\tilde{x}$ in G, and let $\widehat{\mathcal{O}}_x$ and $\widehat{\mathcal{O}}_{\tilde{x}}$ denote the complete local rings of $\mathrm{X}_1(n)$ and $\mathrm{X}(\mathcal{P};\Gamma_1(n))$ at x and $\tilde{x}$, respectively. Then $\widehat{\mathcal{O}}_x$ can be identified with the ring of $G_{\tilde{x}}$-invariants in $\widehat{\mathcal{O}}_{\tilde{x}}$. We apply the following algebraic result to this situation.

Lemma 5.1 (Edixhoven and A. J. de Jong; see de Jong [19, Lemma 4.3]). *Let R be a complete local Noetherian domain, and let A be an R-algebra. Let H be a finite subgroup of $\operatorname{Aut}_R A$, and let A^H denote the R-algebra of A-invariants.*

(1) If $A \cong R[[u]]$, then $A^H \cong R[[x]]$.

(2) If $A \cong R[[u,v]]/(uv - f)$ for some f in the maximal ideal of R, then A^H is isomorphic to $R[[x]]$ or to $R[[x,y]]/(xy - f^{\#H})$. □

The regularity of the complete local $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$-algebra $\widehat{\mathcal{O}}_{\tilde{x}}$ implies that it is isomorphic to $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p][[u,v]]/(uv-\pi)$ for some uniformiser π of $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$. Taking $G_{\tilde{x}}$-invariants, we see that

$$\widehat{\mathcal{O}}_x \cong \mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p][[x,y]]/(xy - \pi^{\#G_x}).$$

This implies that in passing from the semi-stable model of $\mathrm{X}_1(n)$ over $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$ to its minimal regular model, the point x is replaced by a chain of $e-1$ projective lines, where $e \le 6$. From this it follows that the diameter of the reduction graph is at most $6/(p-1)$, so we conclude

$$\gamma(\mathrm{X}_1(n)_{W_p}) \le \frac{6}{p-1}.$$

Next we treat the general situation where $n = p^a m$ with $a \ge 2$ and $m \ge 5$ not divisible by p. In this case $\mathrm{X}_1(n)$ still has a model over the discrete valuation ring $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$ whose special fibre consists of $a+1$ smooth and irreducible components; see Katz and Mazur [53, Theorem 13.11.4]. We denote this model by $\mathrm{X}_1(n)_{\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]}$. However, the special fibre $\mathrm{X}_1(n)_{\overline{\mathbf{F}}_p}$ of $\mathrm{X}_1(n)_{\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]}$ is not semi-stable. We choose a finite extension R of $\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]$ over which $\mathrm{X}_1(n)$ acquires semi-stable reduction. We consider the minimal resolution

$$\pi\colon \tilde{\mathrm{X}}_1(n)_R \to \mathrm{X}_1(n)_{\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]} \otimes_{\operatorname{Spec}\mathrm{W}(\overline{\mathbf{F}}_p)[\zeta_p]} \operatorname{Spec} R.$$

Then $\tilde{\mathrm{X}}_1(n)_R$ is a regular model of $\mathrm{X}_1(n)$ whose special fibre $\tilde{\mathrm{X}}_1(n)_{\overline{\mathbf{F}}_p}$ is semi-stable, and π induces a morphism

$$\pi_{\overline{\mathbf{F}}_p}\colon \tilde{\mathrm{X}}_1(n)_{\overline{\mathbf{F}}_p} \to \mathrm{X}_1(n)_{\overline{\mathbf{F}}_p}.$$

We write G_n for the reduction graph of $\mathrm{X}_1(n)$ over $\mathrm{W}(\overline{\mathbf{F}}_p)$ as defined in § 4.2. For every singular point x of $\mathrm{X}_1(n)_{\overline{\mathbf{F}}_p}$, we write H_x for the union of the edges in G_n corresponding to singular points $\tilde{x} \in \tilde{\mathrm{X}}_1(n)_{\overline{\mathbf{F}}_p}$ with $\pi_{\overline{\mathbf{F}}_p}\tilde{x} = x$. For every irreducible component I of X, we write $\tilde{I}$ for the unique irreducible component of $\tilde{X}$ that maps isomorphically to I under $\pi_{\overline{\mathbf{F}}_p}$. We let T denote the finite subset of G_n consisting of the points that correspond to one of the $\tilde{I}$. Then G_n is the union of the H_x, with x running over the singular points of $\mathrm{X}_1(n)_{\overline{\mathbf{F}}_p}$, and the intersection of any two distinct H_x equals T. For every singular point x of $\mathrm{X}_1(n)_{\overline{\mathbf{F}}_p}$, we define

$$d_x = \max_{I \in T} \max_{g \in H_x} (\text{distance between } I \text{ and } g).$$

If g and h are two points of G_n lying on H_x and H_y, respectively, there is a path of length at most d_x from g to any $I \in T$, and there is a path of length at most d_y from I to h. This implies that

$$\operatorname{diam}(G_n) \le 2 \max_x d_x.$$

Now let m and m' be positive integers with $m \mid m'$, $m \ge 5$ and $p \nmid m'$, and write $n = p^a m$ and $n' = p^a m'$. Let x be a singular point of $\mathrm{X}_1(n)_{\bar{\mathbf{F}}_p}$, and let x' be a point of $\mathrm{X}_1(n')_{\bar{\mathbf{F}}_p}$ mapping to x under the map

$$b_1^{n',n}\colon \mathrm{X}_1(n') \to \mathrm{X}_1(n).$$

The map $b_1^{n',n}$ is étale at x', so the subgraphs $H_{x'}$ of G'_n and H_x of G_n are isometric. This implies that

$$\max_x d_x = \max_{x'} d_{x'},$$

where x and x' run over the singular points of $\mathrm{X}_1(n)_{\bar{\mathbf{F}}_p}$ and $\mathrm{X}_1(n')_{\bar{\mathbf{F}}_p}$, respectively. We conclude that the diameter of $G_{\mathrm{X}_1(p^a m)}$ is bounded for all m such that $m \ge 5$ and $p \nmid n$ by a real number $c(p^a)$ that does not depend on m.

Finally, for $m \le 4$, the same reasoning as that used above for the case $a = 1$ implies that the diameter of $G_{\mathrm{X}_1(n)}$ is bounded by $6c(p^a)$.

Chapter IV

Computational tools

In this chapter we describe the computational techniques that will be used in the next chapter to compute modular Galois representations.

Most of the chapter is taken up by a "toolbox" for computing with divisors on curves over (finite) fields. Roughly speaking, we describe and extend the methods of Khuri-Makdisi for computing with projective curves, and we show that certain results of Couveignes [16] and Diem [27] can be transferred to this setting. The remainder of the chapter is devoted to some computational questions related to finite $\mathbf{F}$-vector space schemes over $\mathbf{Q}$ and to finite-dimensional $\mathbf{F}$-linear representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, where $\mathbf{F}$ is a finite field.

Many of the algorithms we describe are probabilistic. All of these are of the *Las Vegas* type. This means that the running time depends on random data generated during the execution of the algorithm, but that the outcome is guaranteed to be correct. The epithet *Las Vegas* distinguishes such algorithms from those of the *Monte Carlo* type, in which the randomness influences the correctness of the outcome instead of the running time.

1. Algorithms for computing with finite algebras

In this section, we describe some techniques for solving two computational problems about finite algebras over a field. The first is how to find the primary decomposition of such an algebra; the second is how to reconstruct such an algebra from a certain kind of bilinear map between modules over it.

The algebras to which we are going to apply these techniques in the next section are of the form $\Gamma(E, \mathcal{O}_E)$, where E is an effective divisor on a smooth curve over a field k. In this section, however, we place ourselves in the more general setting of arbitrary finite commutative k-algebras.

1.1. Primary decomposition and radicals

Let k be a perfect field. We assume that we have a way to represent elements of k, to perform field operations in k and to test whether an element in our representation is zero. We assume furthermore that have a (probabilistic) algorithm to factor polynomials $f \in k[x]$ in an (expected) number of operations in k that is bounded by a polynomial in the degree of f.

In this situation, there are (probabilistic) algorithms to find the primary decomposition of a finite commutative k-algebra A that finish in an (expected) number of operations in k that is bounded by a polynomial in $[A : k]$. Such algorithms have been known for some time, but do not seem to be easily available in published form; see Khuri-Makdisi's preprint [57, draft version 2, § 7]. For an algorithm to find the primary decomposition of arbitrary (not necessarily commutative) finite algebras over *finite* fields, see Eberly and Giesbrecht [30].

1.2. Reconstructing an algebra from a perfect bilinear map

Let A be a commutative ring. If M, N and O are free A-modules of rank one and

$$\mu\colon M \times N \to O$$

is an A-bilinear map, we say that μ is *perfect* if it induces an isomorphism

$$M \otimes_A N \xrightarrow{\sim} O$$

of free A-modules of rank 1.

Now let k be a field, and let a finite commutative k-algebra A be specified implicitly in the following way. We are given k-vector spaces M, N and O of the same finite dimension, together with a k-bilinear map

$$\mu\colon M \times N \to O$$

We assume there exists a commutative k-algebra A such that M, N and O are free A-modules of rank 1 and μ is a perfect A-bilinear map. The following observation implies that A is the *unique* k-algebra with this property, and also shows how to compute A as a subalgebra of $\operatorname{End}_k M$, provided we are able to find a generator of N as an A-module. We note that the roles of M and N can also be interchanged.

Lemma 1.1. *In the above situation, let g be a generator of the A-module N. The ring homomorphism $A \to \operatorname{End}_k M$ sending a to multiplication by a is, as an A-linear map, the composition of*

$$\begin{aligned} A &\xrightarrow{\sim} N \\ a &\longmapsto ag \end{aligned}$$

and

$$\begin{aligned} N &\longrightarrow \operatorname{End}_k M \\ n &\longmapsto \mu(\ \ ,g)^{-1} \circ \mu(\ \ ,n). \end{aligned}$$

In particular, the image of A in $\operatorname{End}_k M$ equals the image of the second map.

Proof. This is a straightforward verification. ∎

In the case where k is a finite field, a way to find a generator for N as an A-module is simply to pick random elements $g \in N$ until we find one that generates N. Since μ is perfect, checking whether g generates N comes down to checking whether $\mu(\ \ , g)\colon M \to O$ is an isomorphism. In particular, we can do this without knowing A.

To get a reasonable expected running time for this approach, we need to ensure that N contains sufficiently many elements n such that $N = An$. Since N is free of rank 1, the number of generators equals the number of units in A. Let us therefore estimate under what conditions a random element of A is a unit with probability at least $1/2$. Write d for the degree of A over k. Decomposing A into a product of finite local k-algebras, and noting that the proportion of units in a finite local k-algebra is equal to the proportion of units in its residue field, we see that

$$\frac{\#A^\times}{\#A} \ge \frac{(\#k^\times)^d}{\#k^d} = \left(1 - \frac{1}{\#k}\right)^d;$$

equality occurs if and only if A is a product of d copies of k. Now it is not hard to show that

$$\#k \ge 2d \implies \left(1 - \frac{1}{\#k}\right)^d \ge \frac{1}{2}.$$

Taking a finite extension k' of k of cardinality at least $2d$, we therefore see that a random element of $A_{k'}$ is a unit with probability at least $1/2$. There are well-known algorithms to generate such an extension, such as that of Rabin [83], which runs in probabilistic polynomial time and simply tries random polynomials until it finds one that is irreducible, and the deterministic algorithm of Adleman and Lenstra [1].

Algorithm 1.2 *(Reconstruct an algebra from a bilinear map).* Let k be a finite field, let A be a finite k-algebra, and let

$$\mu\colon M \times N \to O$$

be a perfect A-bilinear map between free A-modules of rank 1. Given the coefficients of μ with respect to some k-bases of M, N and O, this algorithm outputs a k-basis for the image of A in $\operatorname{End}_k M$, consisting of matrices with respect to the given basis of M.

1. Choose an extension k' of k of degree $\left\lceil \frac{\log \max\{2[A:k], q\}}{\log q} \right\rceil$. Let M', N', O' and μ' denote the base extensions of M, N, O and μ to k'.
2. Choose a uniformly random element $g \in N'$.
3. Check whether $\mu'(\ \ , g)\colon M' \to O'$ is an isomorphism; if not, go to step 2.
4. For n ranging over a k'-basis of N', compute the endomorphism
$$a_n = \mu'(\ \ , g)^{-1} \circ \mu'(\ \ , n) \in \operatorname{End}_{k'} M'.$$
Let $A' \subseteq \operatorname{End}_{k'} M'$ denote the k'-span of the a_n.
5. Output a basis for the k-vector space $\operatorname{End}_k M \cap A'$.

Analysis. It follows from Lemma 1.1 that A' equals the image of $k' \otimes_k A$ in $\mathrm{End}_{k'} M$. This implies that the basis returned by the algorithm is indeed a k-basis for the image of A in $\mathrm{End}_k M$. Because of the choice of k', steps 2 and 3 are executed at most twice on average. It is therefore clear that the expected running time of the algorithm is polynomial in $[A : k]$ and $\log \#k$. ◇

If k is infinite (or finite and sufficiently large), we have the following variant. Let Σ be a finite subset of k, and let V be a k-vector space of dimension d with a given basis $v_1, \ldots, v_d$. Consider the set

$$V_\Sigma = \{\sum_{i=1}^{d} \sigma_i v_i \mid \sigma_1, \ldots, \sigma_d \in \Sigma\}$$

of Σ-*linear combinations* of $v_1, \ldots, v_n$. Choosing the σ_i uniformly randomly in Σ, we get the uniform distribution on V_Σ. If $H_1, \ldots, H_l$ are proper linear subspaces of V, then a uniformly random element of V_Σ lies in at least one of the H_i with probability at most $l/\#\Sigma$. Now if A is a finite commutative k-algebra, it contains at most $[A : k]$ maximal ideals. This implies that if Σ is a finite subset of k with $\#\Sigma \geq 2[A : k]$, then a Σ-linear combination of any k-basis of A is a unit with probability at least $1/2$. This leads to the following variant of Algorithm 1.2.

Algorithm 1.3 *(Reconstruct an algebra from a bilinear map).* Let k be a field, let A be a finite k-algebra, and let

$$\mu : M \times N \to O$$

be a perfect A-bilinear map between free A-modules of rank 1. Suppose that we can pick uniformly random elements of some subset Σ of k with $\#\Sigma \geq 2[A : k]$. Given the coefficients of μ with respect to some k-bases of M, N and O, this algorithm outputs a k-basis for the image of A in $\mathrm{End}_k M$, consisting of matrices with respect to the given basis of M.

1. Choose a uniformly random Σ-linear combination g of the given basis of N.
2. Check whether $\mu(\ \ , g) : M \to O$ is an isomorphism; if not, go to step 2.
3. For n ranging over a k-basis of N, compute the endomorphism

$$a_n = \mu(\ \ , g)^{-1} \circ \mu(\ \ , n) \in \mathrm{End}_k M,$$

and output the a_n.

Analysis. This works for the same reason as Algorithm 1.2. ◇

Let us sketch how to solve the problem if k is an arbitrary field. Let p be the characteristic of k. If $p = 0$ or $p \geq 2[A : d]$, we can apply Algorithm 1.3 with $\Sigma = \{0, 1, \ldots, 2[A : d] - 1\}$. Otherwise, we consider the subfield k_0 of k generated by the coefficients of the multiplication table of A over k. Then A is obtained by base extension to k of the finite k_0-algebra A_0 defined by the same multiplication table. We can check whether k_0 is a finite field with $\#k_0 < 2d$ by checking whether each coefficient of the multiplication table satisfies a polynomial of small degree. If this is the case, then we compute an $\mathbf{F}_p$-basis and multiplication table for k_0 and apply Algorithm 1.2 to A_0 over k_0. Otherwise we obtain at some point a finite subset Σ of k, with $\#\Sigma \geq 2d$, consisting of polynomials in the coefficients of the multiplication table. We then apply Algorithm 1.3 to A over k with this Σ.

2. Computing with divisors on a curve

In this section and the next we describe a collection of algorithms, developed by Khuri-Makdisi in [56] and [57], that allow us to compute efficiently with divisors on a curve over a field. In particular, we will describe algorithms for computing in the Picard group of a curve. Many of the results of this section can be found in [56] and [57]. In contrast, §§ 2.6, 2.9 and 2.11 seem to be new.

The curves we consider are complete, smooth and geometrically connected curves over a field k. In this section, the base field is arbitrary, although for some of the algorithms we assume that given a finite k-algebra we can find its primary decomposition. In Section 3, we will study a few computational problems particular to curves over finite fields.

The basic idea is to describe such a curve using a projective embedding via a very ample line bundle $\mathcal{L}$, and to represent divisors as subspaces of the k-vector space $\Gamma(X, \mathcal{L})$ of global sections of $\mathcal{L}$. Using this representation of the curve and of divisors on it, Khuri-Makdisi [56] has given algorithms for computing with divisors and elements of the Picard group. Taking advantage of some improvements to this basic idea, described in [57], his algorithms are at the time of writing the asymptotically fastest known algorithms (measured in operations in the field k) for general curves.

Remark. When the field k is finite (as it is in the applications that we will describe in Chapter V), the fact that the complexity is measured in field operations is no problem. However, if k is a number field, one cannot avoid numerical explosion of the data describing the divisors during computations, even when lattice reduction algorithms are used to reduce the size of the data between operations; see Khuri-Makdisi [57, page 2214].

2.1. Representing the curve

Let X be a complete, smooth, geometrically connected curve over a field k. We fix a line bundle $\mathcal{L}$ on X such that

$$\deg \mathcal{L} \geq 2g + 1.$$

Then $\mathcal{L}$ is very ample (see for example Hartshorne [43, IV, Corollary 3.2(b)]), so it gives rise to a closed immersion

$$i_{\mathcal{L}}: X \to \mathbf{P}\Gamma(X, \mathcal{L})$$

into a projective space of dimension $\deg \mathcal{L} - g$. (We write $\mathbf{P}V$ for the projective space of hyperplanes in a k-vector space V.) The assumption that $\deg \mathcal{L} \geq 2g + 1$ implies moreover that the multiplication maps

$$\mu_{i,j}: \Gamma(X, \mathcal{L}^{\otimes i}) \otimes_k \Gamma(X, \mathcal{L}^{\otimes j}) \longrightarrow \Gamma(X, \mathcal{L}^{\otimes (i+j)}).$$

are surjective for all $i, j \geq 0$, or equivalently that the embedding $i_{\mathcal{L}}$ is projectively normal. This is a classical theorem of Castelnuovo [13], Mattuck [75, page 194] and Mumford [81, page 55]. Below we will state a more general result due to Khuri-Makdisi [56, Lemma 2.2].

Remark. In the context of projective embeddings, the line bundle $\mathcal{L}$ is usually denoted by $\mathcal{O}_X(1)$. However, we often need to deal with line bundles of the form $\mathcal{L}(D)$ for a divisor D, and the author does not like the notation $\mathcal{O}_X(1)(D)$.

We write S_X for the homogeneous coordinate ring of X with respect to the embedding $i_{\mathcal{L}}$. By the fact that $i_{\mathcal{L}}$ is projectively normal, we have a canonical isomorphism

$$S_X \xrightarrow{\sim} \bigoplus_{i \geq 0} \Gamma(X, \mathcal{L}^{\otimes i})$$

of graded k-algebras; see Hartshorne [43, Chapter II, Exercise 5.14]. It turns out that to be able to compute with divisors on X we do not need to know the complete structure of this graded algebra. For all $h \geq 0$ we define the finite graded k-algebra $S_X^{(h)}$ as S_X modulo the ideal generated by homogeneous elements of degree greater than h. The above isomorphism shows that specifying $S_X^{(h)}$ is equivalent to giving the k-vector spaces $\Gamma(X, \mathcal{L}^{\otimes i})$ for $1 \leq i \leq h$ together with the multiplication maps $\mu_{i,j}$ for $i + j \leq h$.

When we speak of a *projective curve* X in the remainder of this section, we will assume without further mention that X is a complete, smooth and geometrically connected curve of genus $g \geq 0$, and that a line bundle $\mathcal{L}$ of degree at least $2g + 1$ has been chosen. We will often write $\mathcal{L}_X$ for this line bundle and g_X for the genus of X to emphasise that they are part of the data.

In the algorithms in this section, the curve X is part of the input in the guise of the graded k-algebra $S_X^{(h)}$ for some sufficiently large h. A lower bound for h is specified in each case. One way to specify the multiplication in $S_X^{(h)}$ is to fix a basis for each of the spaces $\Gamma(X, \mathcal{L}^{\otimes i})$, and to give the matrices for multiplication with each basis element. However, as Khuri-Makdisi explains in [57], a more efficient representation is to choose a trivialisation of $\mathcal{L}$ (and hence of its powers) over an effective divisor of sufficiently large degree or, even better, at sufficiently many distinct rational points of X, so that the multiplication maps can be computed pointwise.

Remarks. (1) The integers g and $\deg \mathcal{L}$ can of course be stored as part of the data describing X. However, they can also be extracted from the dimensions of the k-vector spaces $\Gamma(X, \mathcal{L})$ and $\Gamma(X, \mathcal{L}^{\otimes 2})$; this follows easily from the Riemann–Roch formula.

(2) If the degree of $\mathcal{L}$ is at least $2g + 2$, then the homogeneous ideal defining the embedding $i_{\mathcal{L}}$ is generated by homogeneous elements of degree 2, according to a theorem of Fujita and Saint-Donat; see Lazarsfeld [64, § 1.1]. This makes it possible to deduce equations for X from the k-algebra $S_X^{(2)}$. However, we will not need to do this.

(3) The way of representing curves and divisors described by Khuri-Makdisi in [56] and [57] is especially suited for modular curves. Namely, we can represent a modular

curve X using the projective embedding given by a line bundle of modular forms, and computing the k-algebra $S_X^{(h)}$ for a given h comes down to computing q-expansions of modular forms of a suitable weight to a sufficiently large order. This can be done using modular symbols; see Stein [104] and Section 4 below. If the modular curve has at least 3 cusps (which is the case, for example, for $\mathrm{X}_1(n)$ for all $n \geq 5$), then we can restrict ourselves to modular forms of weight 2, for which the formalism of modular symbols is particularly simple [104, Chapter 3].

2.2. Representing divisors

Let X be a projective curve of genus g in the sense of § 2.1, and let $\mathcal{L}$ be the line bundle of degree at least $2g + 1$ giving the projective embedding of X. To represent divisors on X, it is enough to consider effective divisors, since an arbitrary divisor can be represented by a formal difference of two effective divisors.

Consider an effective divisor D on X such that $\mathcal{L}(-D)$ is generated by global sections. (In terms of the projective embedding, this means that D is the intersection of X and a linear subvariety of $\mathbf{P}\Gamma(X, \mathcal{L})$, or equivalently that D is defined by a system of linear equations.) Such a divisor can be represented as the subspace $\Gamma(X, \mathcal{L}(-D))$ of $\Gamma(X, \mathcal{L})$ consisting of sections vanishing on D. The codimension of $\Gamma(X, \mathcal{L}(-D))$ in $\Gamma(X, \mathcal{L})$ is equal to the degree of D.

A sufficient condition for the line bundle $\mathcal{L}(-D)$ to be generated by global sections is

$$\deg D \leq \deg \mathcal{L} - 2g; \tag{2.1}$$

see for example Hartshorne [43, IV, Corollary 3.2(a)]. However, we note that in general not every subspace of codimension at most $\deg \mathcal{L} - 2g$ is of the form $\Gamma(X, \mathcal{L}(-D))$ for an effective divisor D of the same degree.

Remark. This way of representing divisors comes down (at least for divisors of degree $d \leq \deg \mathcal{L} - 2g$) to embedding the d-th symmetric power of X into the Grassmannian variety parametrising subspaces of codimension d in $\Gamma(X, \mathcal{L})$ and viewing divisors of degree d as points on this Grassmannian variety.

It will often be necessary to consider divisors D of degree larger than the bound $\deg \mathcal{L} - 2g$ of (2.1). In such cases we can represent D as a subspace of $\Gamma(X, \mathcal{L}^{\otimes i})$ for i sufficiently large such that

$$\deg D \leq i \deg \mathcal{L} - 2g, \tag{2.2}$$

provided of course that we know $S_X^{(h)}$ for some $h \geq i$.

Khuri-Makdisi's algorithms rest on the following two results. The first is a generalisation of the theorem of Castelnuovo, Mattuck and Mumford mentioned above. It says in effect that to compute the space of global sections of the tensor product of two line bundles of sufficiently large degree, it is enough to multiply global sections of those line bundles.

Lemma 2.1 (Khuri-Makdisi [56, Lemma 2.2]). *Let X be a complete, smooth, geometrically connected curve of genus g over a field k, and let $\mathcal{M}$ and $\mathcal{N}$ be line bundles* *on X whose degrees are at least $2g+1$. Then the canonical k-linear map*

$$\Gamma(X,\mathcal{M})\otimes_k\Gamma(X,\mathcal{N})\longrightarrow\Gamma(X,\mathcal{M}\otimes_{\mathcal{O}_X}\mathcal{N})$$

is surjective.

The second result shows how to find the space of global sections of a line bundle that vanish on a given effective divisor, where this divisor is represented as a subspace of global sections of a second line bundle.

Lemma 2.2 (Khuri-Makdisi [56, Lemma 2.3]). *Let X be a complete, smooth, geometrically connected curve of genus g over a field k, let $\mathcal{M}$ and $\mathcal{N}$ be line bundles on X such that $\mathcal{N}$ is generated by global sections, and let D be any effective divisor on X. Then the inclusion*

$$\Gamma(X,\mathcal{M}(-D))\subseteq\bigl\{s\in\Gamma(X,\mathcal{M})\bigm| s\Gamma(X,\mathcal{N})\subseteq\Gamma(X,\mathcal{M}\otimes\mathcal{N}(-D))\bigr\}\tag{2.3}$$

is an equality.

Thanks to these two lemmata, one can give algorithms to do basic operations on divisors; see Khuri-Makdisi [56, § 3]. For example, we can add, subtract and intersect divisors of sufficiently small degree, and we can test whether a given subspace of $\Gamma(X,\mathcal{L}^{\otimes i})$ is of the form $\Gamma(X,\mathcal{L}^{\otimes i}(-D))$ for some effective divisor D. See also Algorithm 2.11 below for an example where Lemmata 2.1 and 2.2 are used.

2.3. Deflation and inflation

An ingredient that Khuri-Makdisi uses in [57] to speed up the algorithms is *deflation* of subspaces. Suppose we want to compute the space $\Gamma(X,\mathcal{M}(-D))$ using (2.3) in the case where $\mathcal{M}=\mathcal{L}^{\otimes i}$ and $\mathcal{N}=\mathcal{L}^{\otimes j}(-E)$ with i and j positive integers and where D and E are effective divisors satisfying (2.2). On the right-hand side of (2.3), we may replace $\Gamma(X,\mathcal{N})$ by any basepoint-free subspace; this is clear from the proof of [56, Lemma 2.3]. It turns out that there always exists such a subspace of dimension $O(\log(\deg\mathcal{N}))$, and a subspace of dimension 2 exists if the base field is either infinite or finite of sufficiently large cardinality. Moreover, one can efficiently find such a subspace by random trial; see Khuri-Makdisi [57, Proposition/Algorithm 3.7].

Remark. This random search for small basepoint-free subspaces is the reason why Khuri-Makdisi's algorithms in [57] are probabilistic, as opposed to those in [56].

Suppose we are given a basepoint-free subspace W of $\Gamma(X,\mathcal{L}^{\otimes i}(-D))$ for some i and D such that $\Gamma(X,\mathcal{L}^{\otimes i}(-D))$ is basepoint-free. Then we can reconstruct the complete space $\Gamma(X,\mathcal{L}^{\otimes i}(-D))$ from W. This procedure is called *inflation*. To describe how this can be done, we first state the following slight generalisation of a result of Khuri-Makdisi [57, Theorem 3.5(2)].

Lemma 2.3. *Let X be a complete, smooth, geometrically connected curve of genus g over a field k, and let $\mathcal{M}$ and $\mathcal{N}$ be line bundles on X. Let V be a non-zero subspace of $\Gamma(X,\mathcal{M})$, and let D be the common divisor of the elements of V. If the inequality*

$$-\deg\mathcal{M}+\deg\mathcal{N}+\deg D\ge 2g-1$$

is satisfied, the canonical k-linear map

$$V \otimes_k \Gamma(X, \mathcal{N}) \longrightarrow \Gamma(X, \mathcal{M} \otimes_{\mathcal{O}_X} \mathcal{N}(-D)) \tag{2.4}$$

is surjective.

Proof. We note that $\mathcal{M}(-D)$ is generated by global sections, since we can view V as a subspace of $\Gamma(X, \mathcal{M}(-D))$ and the elements of V have common divisor 0 as sections of $\mathcal{M}(-D)$. We also note that $\deg \mathcal{M} \geq \deg D$. Therefore the assumption on the degrees of $\mathcal{M}$, $\mathcal{N}$ and D implies the inequalities

$$\deg \mathcal{N} \geq 2g - 1$$

and

$$\deg(\mathcal{M} \otimes \mathcal{N}(-D)) \geq 2g - 1.$$

After extending the field k, we may assume it is infinite. Then there exist elements $s, t \in V$ with common divisor D; see Khuri-Makdisi [57, Lemma 4.1]. The space

$$s\Gamma(X, \mathcal{N}) + t\Gamma(X, \mathcal{N})$$

lies in the image of (2.4), so it suffices to show that

$$\dim_k(s\Gamma(X, \mathcal{N}) + t\Gamma(X, \mathcal{N})) = \dim_k \Gamma(X, \mathcal{M} \otimes \mathcal{N}(-D)).$$

Write

$$\operatorname{div} s = D + E \quad \text{and} \quad \operatorname{div} s = D + F$$

where E and F are disjoint effective divisors. Then we have

$$\begin{aligned} \dim_k(s\Gamma(X, \mathcal{N}) + t\Gamma(X, \mathcal{N})) &= 2 \dim_k \Gamma(X, \mathcal{N}) - \dim_k(s\Gamma(X, \mathcal{N}) \cap t\Gamma(X, \mathcal{N})) \\ &= 2 \dim_k \Gamma(X, \mathcal{N}) \\ &\qquad - \dim_k \Gamma(X, \mathcal{M} \otimes \mathcal{N}(-D - E - F)) \\ &= 2 \dim_k \Gamma(X, \mathcal{N}) - \dim_k \Gamma(X, \mathcal{M}^\vee \otimes \mathcal{N}(D)). \end{aligned}$$

The last equality follows from the fact that multiplication by st induces an isomorphism

$$\mathcal{M}^\vee(D) \xrightarrow{\sim} \mathcal{M}(-D - E - F).$$

Using the fact that the various line bundles have degrees at least $2g - 1$, we see that

$$\begin{aligned} \dim_k(s\Gamma(X, \mathcal{N}) + t\Gamma(X, \mathcal{N})) &= 2(1 - g + \deg \mathcal{N}) - (1 - g + \deg \mathcal{M}^\vee \otimes \mathcal{N}(D)) \\ &= 1 - g + \deg \mathcal{M} + \deg \mathcal{N} - \deg D \\ &= \dim_k \Gamma(X, \mathcal{M} \otimes \mathcal{N}(-D)). \end{aligned}$$

This finishes the proof. □

To find the inflation of a basepoint-free subspace W of $\Gamma(X, \mathcal{L}^{\otimes i}(-D))$, we choose a positive integer j such that

$$(j-i)\deg \mathcal{L} + \deg D \geq 2g-1.$$

By Lemma 2.3 we can then compute $\Gamma(X, \mathcal{L}^{\otimes(i+j)}(-D))$ as the image of the bilinear map

$$W \otimes_k \Gamma(X, \mathcal{L}^{\otimes j}) \longrightarrow \Gamma(X, \mathcal{L}^{\otimes(i+j)}).$$

Then we compute

$$\Gamma(X, \mathcal{L}^{\otimes i}(-D)) = \bigl\{ s \in \Gamma(X, \mathcal{L}^{\otimes i}) \bigm| s\Gamma(X, \mathcal{L}^{\otimes j}) \subseteq \Gamma(X, \mathcal{L}^{\otimes(i+j)}(-D)) \bigr\}$$

using Lemma 2.2. We note that for this last step we can use a small basepoint-free subspace of $\Gamma(X, \mathcal{L}^{\otimes j})$ computed in advance.

2.4. Decomposing divisors into prime divisors

Let X be a complete, smooth, geometrically connected curve of genus g over a field k, with a projective embedding via a line bundle $\mathcal{L}$ as in § 2.1. The problem we are now going to study is how to find the decomposition of a given divisor on X as a linear combination of prime divisors. We will see below that this can be done if we are given the algebra $S_X^{(h)}$ for sufficiently large h and if we are able to compute the primary decomposition of a finite commutative k-algebra. We have seen in § 1.1 that this is possible in the case where k is perfect and we have an algorithm for factoring polynomials in one variable over k.

Let i be a positive integer, and let D be an effective divisor such that

$$\deg D \leq i \deg \mathcal{L} - 2g + 1.$$

We view D as a closed subscheme of X via the canonical closed immersion

$$j_D \colon D \to X.$$

For every line bundle $\mathcal{M}$ on X, the k-vector space $\Gamma(D, j_D^*\mathcal{M})$ is in a natural way a free module of rank one over $\Gamma(D, \mathcal{O}_D)$. The multiplication map

$$\mu_{i,i} \colon \Gamma(X, \mathcal{L}^{\otimes i}) \times \Gamma(X, \mathcal{L}^{\otimes i}) \longrightarrow \Gamma(X, \mathcal{L}^{\otimes 2i})$$

descends to a bilinear map

$$\mu_{i,i}^D \colon \Gamma(D, j_D^*\mathcal{L}^{\otimes i}) \times \Gamma(D, j_D^*\mathcal{L}^{\otimes i}) \longrightarrow \Gamma(D, j_D^*\mathcal{L}^{\otimes 2i})$$

of free modules of rank 1 over $\Gamma(D, \mathcal{O}_D)$. This map is perfect in the sense of § 1.2.

We now assume that the graded k-algebra $S_X^{(h)}$ as in § 2.1 is given for some $h \geq 2$. From the subspace $\Gamma(X, \mathcal{L}^{\otimes i}(-D))$ of $\Gamma(X, \mathcal{L}^{\otimes i})$ we can then determine $\Gamma(D, j_D^*\mathcal{L}^{\otimes i})$ as a k-vector space by means of the short exact sequence

$$0 \longrightarrow \Gamma(X, \mathcal{L}^{\otimes i}(-D)) \longrightarrow \Gamma(X, \mathcal{L}^{\otimes i}) \longrightarrow \Gamma(D, j_D^*\mathcal{L}^{\otimes i}) \longrightarrow 0. \tag{2.5}$$

(Note that exactness on the right follows from the assumption that $\deg \mathcal{L}^{\otimes i}(-D) \geq 2g-1$.) Similarly, we can compute $\Gamma(D, j_D^*\mathcal{L}^{\otimes 2i})$ from $\Gamma(X, \mathcal{L}^{\otimes 2i}(-D))$ using the same sequence with i replaced by $2i$. We can then determine the bilinear map $\mu_{i,i}^D$ induced by $\mu_{i,i}$ by standard methods from linear algebra.

We then use the method described in § 1.2 to compute the k-algebra $\Gamma(D, \mathcal{O}_D)$ together with its action on $\Gamma(D, j_D^*\mathcal{L}^{\otimes i})$. Next we find the primary decomposition of $\Gamma(D, \mathcal{O}_D)$, say

$$\Gamma(D, \mathcal{O}_D) \cong A_1 \times A_2 \times \cdots \times A_r,$$

where each factor A_i is a finite local k-algebra with maximal ideal P_i; we assume the field k is such that we can do this (see § 1.1). Such a prime ideal P_i corresponds to a prime divisor in the support of D, and the corresponding multiplicity equals

$$m_i = \frac{[A_i : k]}{[A_i/P_i : k]}.$$

Algorithm 2.4 *(Decomposition of a divisor).* Let X be a projective curve over a field k. Let i be a positive integer, and let D be an effective divisor such that

$$\deg D \leq i \deg \mathcal{L}_X - 2g_X + 1.$$

Suppose that we have a (probabilistic) algorithm to compute the primary decomposition of a finite commutative k-algebra A with (expected) running time polynomial in $[A : k]$, measured in operations in k. Given the k-algebra $S_X^{(2i)}$ and the subspaces $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes i})$ and $\Gamma(X, \mathcal{L}_X^{\otimes 2i}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes 2i})$, this algorithm outputs the decomposition of D as a linear combination of prime divisors as a list of pairs (P, m_P), where P is a prime divisor and m_P is the multiplicity of P in D.

1. Compute the spaces $\Gamma(D, j_D^*\mathcal{L}_X^{\otimes i})$ and $\Gamma(D, j_D^*\mathcal{L}_X^{\otimes 2i})$ using (2.5) and the analogous short exact sequence with $2i$ in place of i.

2. Compute the k-bilinear map $\mu_{i,i}^D$ from $\mu_{i,i}$.

3. Using the method of § 1.2, compute a k-basis for $\Gamma(D, \mathcal{O}_D)$ as a linear subspace of $\operatorname{End}_k \Gamma(D, j_D^*\mathcal{L}_X^{\otimes i})$, where elements of the latter k-algebra are expressed as matrices with respect to some fixed basis of $\Gamma(D, j_D^*\mathcal{L}_X^{\otimes i})$.

4. Compute the multiplication table of $\Gamma(D, \mathcal{O}_D)$ on the k-basis of $\Gamma(D, \mathcal{O}_D)$ found in the previous step.

5. Find the primary decomposition of $\Gamma(D, \mathcal{O}_D)$.

6. For each local factor A computed in the previous step, let P_A denote the maximal ideal of A, output the inverse image of $P_A \cdot \Gamma(D, j_D^*\mathcal{L}_X^{\otimes i})$ in $\Gamma(X, \mathcal{L}_X^{\otimes i})$ and the integer $[A : k]/[A/P_A : k]$.

Analysis. It follows from the above discussion that the algorithm returns the correct result. It is straightforward to check that the running time is polynomial in i and $\deg \mathcal{L}_X$, measured in operations in k. ◇

A special case of this algorithm is when D is the intersection of X with a hypersurface of degree $i-1$. Let s be a non-zero section of $\mathcal{L}_X^{\otimes(i-1)}$ defining this hypersurface. The subspaces that are used in this algorithm can then be computed as

$$\Gamma(X, \mathcal{L}_X^{\otimes i}(-D)) = s\Gamma(X, \mathcal{L}_X)$$

and

$$\Gamma(X, \mathcal{L}_X^{\otimes 2i}(-D)) = s\Gamma(X, \mathcal{L}_X^{\otimes(i+1)}).$$

2.5. Finite morphisms between curves

Let us now look at finite morphisms between curves. A finite morphism

$$f\colon X \to Y$$

of complete, smooth, geometrically connected curves induces two functors

$$f^*\colon \{\text{line bundles on } Y\} \to \{\text{line bundles on } X\}$$

and

$$\mathrm{N}_f\colon \{\text{line bundles on } X\} \to \{\text{line bundles on } Y\}.$$

Here $f^*\mathcal{N}$ denotes the usual inverse image of the line bundle $\mathcal{N}$ on Y, and $\mathrm{N}_f\mathcal{M}$ is the *norm* of the line bundle $\mathcal{M}$ on X under the morphism f.

Let us briefly explain the notion of the norm of a line bundle. The norm functor is a special case (that of $\mathbf{G}_\mathrm{m}$-torsors) of the *trace of a torsor* for a commutative group scheme under a finite locally free morphism; see Deligne [100, exposé XVII, n$^{\text{os}}$ 6.3.20–6.3.26]. We formulate the basic results for arbitrary finite locally free morphisms of schemes

$$f\colon X \to Y.$$

In this situation there exists a functor

$$\mathrm{N}_f\colon \{\text{line bundles on } X\} \to \{\text{line bundles on } Y\}$$

together with a collection of homomorphisms

$$\mathrm{N}_f^{\mathcal{L}}\colon f_*\mathcal{L} \to \mathrm{N}_f\mathcal{L}$$

of sheaves of sets, for all line bundles $\mathcal{L}$ on X, functorial under isomorphisms of line bundles on X, sending local generating sections on X to local generating sections on Y and such that the equality

$$\mathrm{N}_f^{\mathcal{L}}(xl) = \mathrm{N}_f(x) \cdot \mathrm{N}_f^{\mathcal{L}}(l)$$

holds for all local sections x of $f_*\mathcal{O}_X$ and l of $f_*\mathcal{L}$. Here $\mathrm{N}_f\colon f_*\mathcal{O}_X \to \mathcal{O}_Y$ denotes the usual norm map for a finite locally free morphism. Moreover, the functor N_f together with the collection of the $\mathrm{N}_f^{\mathcal{L}}$ is unique up to unique isomorphism. Instead of N_f we also write $\mathrm{N}_{X/Y}$ if the morphism f is clear from the context.

The basic properties of the norm functor are the following (see [100, exposé XVII, n° 6.3.26]):

(1) the functor N_f is compatible with any base change $Y' \to Y$;

(2) if $\mathcal{L}_1$ and $\mathcal{L}_2$ are two line bundles on X, there is a natural isomorphism

$$\mathrm{N}_f(\mathcal{L}_1 \otimes_{\mathcal{O}_X} \mathcal{L}_2) \cong \mathrm{N}_f\mathcal{L}_1 \otimes_{\mathcal{O}_Y} \mathrm{N}_f\mathcal{L}_2;$$

(3) if $X \xrightarrow{f} Y \xrightarrow{g} Z$ are finite locally free morphisms, there is a natural isomorphism

$$\mathrm{N}_{g\circ f} \xrightarrow{\sim} \mathrm{N}_g \circ \mathrm{N}_f.$$

Furthermore, there is a functorial isomorphism

$$\mathrm{N}_f\mathcal{L} \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{O}_Y}(\det_{\mathcal{O}_Y} f_*\mathcal{O}_X, \det_{\mathcal{O}_Y} f_*\mathcal{L}); \tag{2.6}$$

see Deligne [100, exposé XVIII, n° 1.3.17], and compare Hartshorne [43, IV, Exercise 2.6].

We now consider projective curves X and Y as defined in § 2.1. Suppose we have a finite morphism

$$f\colon X \to Y$$

with the property that f is induced by a graded homomorphism

$$f^\#\colon S_Y \to S_X$$

between the homogeneous coordinate rings of Y and X, or equivalently by a morphism of the corresponding affine cones over X and Y. Then $f^\#$ induces an isomorphism

$$f^*\mathcal{L}_Y \xrightarrow{\sim} \mathcal{L}_X$$

of line bundles on X; see Hartshorne [43, Chapter II, Proposition 5.12(c)]. In particular, this implies

$$\deg \mathcal{L}_X = \deg f \cdot \deg \mathcal{L}_Y.$$

We represent a finite morphism $f\colon X \to Y$ by the k-algebras $S_X^{(h)}$ and $S_Y^{(h)}$ for some $h \geq 2$, together with the k-algebra homomorphism

$$f^\#\colon S_Y^{(h)} \to S_X^{(h)}$$

induced by $f^\#\colon S_Y \to S_X$, given as a collection of linear maps $\Gamma(Y, \mathcal{L}_Y^{\otimes i}) \to \Gamma(X, \mathcal{L}_X^{\otimes i})$ compatible with the multiplication maps on both sides.

In the following, when we mention a *finite morphism* $f\colon X \to Y$ *between projective curves*, we assume that the k-algebras $S_X^{(h)}$ and $S_Y^{(h)}$ and the homomorphism $f^\#\colon S_Y^{(h)} \to S_X^{(h)}$ are given for some $h \geq 2$. In the algorithms described below, we will indicate where necessary how large h needs to be.

Remark. The homomorphism $f^\#$ gives rise to an injective k-linear map

$$\Gamma(Y, \mathcal{L}_Y) \to \Gamma(X, \mathcal{L}_X).$$

Given this map we can reconstruct $S(Y)$ as a subalgebra of $S(X)$ by noting that $S(Y)$ is generated as a k-algebra by $\Gamma(Y, \mathcal{L}_Y)$.

2.6. Images, pull-backs and push-forwards of divisors

Let us consider a finite morphism $f\colon X \to Y$ between complete, smooth, geometrically connected curves over a field k. Such a morphism f induces various maps between the groups of divisors on X and on Y.

First, for an *effective* divisor D on X, we write $f(D)$ for the schematic image of D under f. The definition implies that the ideal sheaf $\mathcal{O}_Y(-f(D))$ is the inverse image of $f_*\mathcal{O}_X(-D)$ under the natural map $\mathcal{O}_Y \to f_*\mathcal{O}_X$.

Second, for any divisor D on X, we have the "push-forward" f_*D of D by f; see Hartshorne [43, IV, Exercise 2.6]. If P is a prime divisor on X, then its image $f(P)$ under f is a prime divisor on Y, the residue field $k(P)$ is a finite extension of $k(f(P))$, and f_*P is given by the formula

$$f_*P = [k(P) : k(f(P))] \cdot f(P). \tag{2.7}$$

The residue field extension degree at P can simply be computed as

$$\begin{aligned} [k(P) : k(f(P))] &= \frac{[k(P) : k]}{[k(f(P)) : k]} \\ &= \frac{\deg P}{\deg f(P)}. \end{aligned}$$

Third, for any divisor E on Y, we have the "pull-back" f^*E of E by f; see for example Hartshorne [43, page 137]. If Q is a prime divisor on Y, then f^*Q is given by the formula

$$f^*Q = \sum_{P\colon f(P)=Q} e(P) \cdot P \tag{2.8}$$

where P runs over the prime divisors of X mapping to Q and $e(P)$ denotes the ramification index of f at P.

We extend both f_* and f^* to arbitrary divisors on X and Y by linearity. Note that (2.7) and (2.8) imply the well-known formula

$$f_*f^*E = (\deg f)E$$

for any divisor E on Y. Furthermore, if E is an *effective* divisor on Y, we have an equality

$$f^*E = E \times_Y X$$

of closed subschemes of X, and if $\mathcal{I}_E$ denotes the ideal sheaf of E, then its inverse image $f^{-1}\mathcal{I}_E$ is the ideal sheaf of f^*E.

Remark. The map $D \mapsto f(D)$ is not in general linear in D. We do not extend it to the divisor *group* on X, and in fact will only need schematic images of *prime* divisors on X in what follows. In contrast, the maps f_* and f^* are linear by definition.

Now assume f is a finite morphism between *projective* curves, in the sense of § 2.5. In particular, we have a homomorphism $f^\#\colon S_Y \to S_X$ of graded k-algebras. We will give algorithms to compute the image and the push-forward of a divisor on X as well as the pull-back of a divisor on Y.

The schematic image $f(D)$ of an effective divisor D on X can be computed using the following obvious algorithm.

Algorithm 2.5 *(Image of a divisor under a finite morphism).* Let $f\colon X \to Y$ be a finite morphism between projective curves, let i be a positive integer, and let D be an effective divisor on X. Given the k-algebras $S_X^{(i)}$ and $S_Y^{(i)}$, the homomorphism $f^\#\colon S_Y^{(i)} \to S_X^{(i)}$ and the subspace $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes i})$, this algorithm outputs the subspace $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-f(D)))$ of $\Gamma(Y, \mathcal{L}_Y^{\otimes i})$.

1. Output the inverse image of the subspace $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes i})$ under the linear map $\Gamma(Y, \mathcal{L}_Y^{\otimes i}) \to \Gamma(X, \mathcal{L}_X^{\otimes i})$.

Analysis. The definition of $f(D)$ implies that the line bundle $\mathcal{L}_Y^{\otimes i}(-f(D))$ equals the inverse image of $f_*\mathcal{L}_X^{\otimes i}(-D)$ under the natural map $\mathcal{L}_Y^{\otimes i} \to f_*\mathcal{L}_X^{\otimes i}$. Taking global sections, we see that $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-f(D)))$ is the inverse image of $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ under the natural map $\Gamma(Y, \mathcal{L}_Y^{\otimes i}) \to \Gamma(X, \mathcal{L}_X^{\otimes i})$. It is clear that the algorithm needs a number of operations in k that is polynomial in $\deg \mathcal{L}_X$ and i. $\diamond$

Remark. In the above algorithm, we have not placed any restrictions on the degrees of D and $f(D)$. However, $f(D)$ is not uniquely determined by $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-f(D)))$ if its degree is too large.

The algorithm to compute pull-backs that we will now give is based on the fact that the pull-back of an effective divisor E is simply the fibred product $E \times_Y X$, viewed as a closed subscheme of X. In particular, the algorithm does not have to compute the ramification indices, so instead we can *use* it to compute ramification indices. Namely, if P is a prime divisor on X, we see from (2.8) that the ramification index at P equals the multiplicity with which P occurs in the divisor $f^*(f(P))$.

Algorithm 2.6 *(Pull-back of a divisor under a finite morphism).* Let $f\colon X \to Y$ be a finite morphism between projective curves. Let i and j be positive integers, and let E be an effective divisor on Y such that

$$\deg f \cdot \deg E \le i \deg \mathcal{L}_X - 2g_X, \quad \deg E \le i \deg \mathcal{L}_Y - 2g_Y$$

and

$$(j - i) \deg \mathcal{L}_X + \deg f \cdot \deg E \ge 2g_X - 1.$$

(If we take $j \ge i+1$, the last equality does not pose an extra restriction on E.) Given the k-algebras $S_X^{(i+j)}$ and $S_Y^{(i+j)}$, the k-algebra homomorphism $f^\#\colon S_Y^{(i+j)} \to S_X^{(i+j)}$ and the subspace $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-E))$ of $\Gamma(Y, \mathcal{L}_Y^{\otimes i})$, this algorithm outputs the subspace $\Gamma(X, \mathcal{L}_X^{\otimes i}(-f^*E))$ of $\Gamma(X, \mathcal{L}_X^{\otimes i})$.

1. Compute the image W of $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-E))$ under the linear map

$$f^\#\colon \Gamma(Y, \mathcal{L}_Y^{\otimes i}) \to \Gamma(X, \mathcal{L}_X^{\otimes i}).$$

2. Compute the space $\Gamma(X, \mathcal{L}_X^{\otimes i+j}(-f^*E))$ as the product of W and $\Gamma(X, \mathcal{L}_X^{\otimes j})$ (see Lemma 2.3).
3. Compute $\Gamma(X, \mathcal{L}_X^{\otimes i}(-f^*E))$ using Lemma 2.2, and output the result.

Analysis. The ideal in S_Y defining E is generated by the linear forms vanishing on E, and the ideal of S_X defining f^*E is generated by the pull-backs of these forms. This shows that f^*E is defined by the forms in W. In the second and third step, we compute the space of all forms vanishing on f^*E, i.e. the inflation of W. That the method described is correct was proved in § 2.3. The running time is clearly polynomial in $\deg \mathcal{L}_X$, i and j. $\diamond$

Algorithm 2.7 *(Push-forward of a divisor under a finite morphism).* Let $f\colon X \to Y$ be a finite morphism between projective curves over a field k, let i be a positive integer, and let D be an effective divisor on X such that

$$\deg D \le i \deg \mathcal{L}_X - 2g_X - 1 \quad \text{and} \quad \deg D \le i \deg \mathcal{L}_Y - 2g_Y.$$

Suppose that we have a (probabilistic) algorithm to compute the primary decomposition of a finite commutative k-algebra A with (expected) running time polynomial in $[A : k]$, measured in operations in k. Given the k-algebras $S_X^{(2i)}$ and $S_Y^{(2i)}$, the homomorphism $f^\#\colon S_Y^{(2i)} \to S_X^{(2i)}$ and the subspace $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes i})$, this algorithm outputs the subspace $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-f_*D))$ of $\Gamma(Y, \mathcal{L}_Y^{\otimes i})$.

1. Compute $\Gamma(X, \mathcal{L}_X^{\otimes 2i}(-D))$ as the product of $\Gamma(X, \mathcal{L}_X^{\otimes i})$ and $\Gamma(X, \mathcal{L}_X^{\otimes i}(-D))$ (see Lemma 2.1).
2. Find the decomposition of D as a linear combination $\sum_P n_P P$ of prime divisors using Algorithm 2.4.
3. For each prime divisor P in the support of D, compute $\Gamma(Y, \mathcal{L}^{\otimes i}(-f(P)))$ using Algorithm 2.5, and compute $[k(P) : k(f(P))]$.
4. Compute the space $\Gamma(Y, \mathcal{L}_Y^{\otimes i}(-f_*D))$, where

$$f_*D = \sum_P n_P [k(P) : k(f(P))] f(P),$$

and output the result.

Analysis. The correctness of the algorithm follows from the definition of f_*. It runs in (probabilistic) polynomial time in $\deg \mathcal{L}_X$ and i, measured in field operations in k. $\diamond$

We include here another algorithm that computes the push-forward of an effective divisor under a non-constant rational function $X \to \mathbf{P}^1$ in a slightly different setting than before. We only assume X to be given as a projective curve in the sense of § 2.1, and we represent effective divisors on $\mathbf{P}^1$ as zero loci of homogeneous polynomials. For simplicity, we only consider divisors of degree at most $\deg \mathcal{L}_X$.

Algorithm 2.8 *(Push-forward of an effective divisor by a rational function).* Let X be a projective curve over a field k, let i be a positive integer, let ψ be a non-constant rational function on X given as the quotient of two sections $s, t \in \Gamma(X, \mathcal{L}_X^{\otimes i})$ without common zeroes, and let D be an effective divisor on X of degree $d \le \deg \mathcal{L}_X$. Suppose that we have a (probabilistic) algorithm to compute the primary decomposition of a finite commutative k-algebra A with (expected) running time polynomial in $[A : k]$, measured in operations in k. Given the k-algebra $S_X^{(\max\{4,i\})}$ and the subspace $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-D))$, this algorithm outputs the homogeneous polynomial of degree d defining the closed subscheme $\psi_* D$ of $\mathbf{P}^1_k$. (This polynomial is unique up to multiplication by elements of $k^\times$)

1. Compute the space $\Gamma(X, \mathcal{L}_X^{\otimes 4}(-D))$, and use Algorithm 2.4 to compute the decomposition of D as a linear combination $D = \sum_Q n_Q Q$ of prime divisors.

2. For each prime divisor Q occurring in the decomposition of D:

3. Compute the base change $X_{k(Q)}$, where $k(Q)$ is the residue field of Q. Compute the primary decomposition of $Q_{k(Q)}$ and pick a rational point Q' in it.

4. Compute $\Gamma(X_{k(Q)}, \mathcal{L}_X^{\otimes 2}(-Q'))$, then compute the (one-dimensional) intersection of this space with $k \cdot s + k \cdot t$, and express some generator of this intersection as $b_Q s - a_Q t$ with $a_Q, b_Q \in k(Q)$. The element $\psi(Q') \in \mathbf{P}^1(k(Q))$ now has homogeneous coordinates $(a_Q : b_Q)$.

5. Compute the homogeneous polynomial

$$f_{\psi_* Q} = \mathrm{N}_{k(Q)/k}(b_Q u - a_Q v) \in k[u, v]$$

defining $\psi_* Q$.

6. Output the homogeneous polynomial

$$f_{\psi_* D} = \prod_Q f_{\psi_* Q}^{n_Q} \in k[u, v]$$

of degree d defining $\psi_* D$.

Analysis. It is straightforward to check that the algorithm is correct and has expected running time polynomial in i and $\deg \mathcal{L}_X$, counted in operations in k. ◇

2.7. The norm functor for effective divisors

Let X be a proper, smooth, geometrically connected curve over a field k, and let E be an effective divisor on X. We view E as a closed subscheme of X, finite over k, and we write

$$j_E \colon E \to X$$

for the closed immersion of E into X. For the purposes of § 3.6 below, we will need an explicit description of the norm functor $\mathrm{N}_{E/k}$ (for the canonical morphism $E \to \operatorname{Spec} k$) that we saw in § 2.5. We view $\mathrm{N}_{E/k}$ as a functor from free $\mathcal{O}_E$-modules of rank 1 to k-vector spaces of dimension 1.

Let $\mathcal{M}$ be a line bundle on X. We abbreviate

$$\Gamma(E, \mathcal{M}) = \Gamma(E, j_E^* \mathcal{M})$$

and

$$\mathrm{N}_{E/k} \mathcal{M} = \mathrm{N}_{E/k}(j_E^* \mathcal{M}).$$

Suppose we have two line bundles $\mathcal{M}^+$ and $\mathcal{M}^-$, both of degree at least $\deg E+2g-1$, together with an isomorphism

$$\mathcal{M} \cong \mathcal{H}om_{\mathcal{O}_X}(\mathcal{M}^-, \mathcal{M}^+).$$

Then we can compute $\Gamma(E, \mathcal{M}^-)$ and $\Gamma(E, \mathcal{M}^+)$ using the short exact sequences

$$0 \longrightarrow \Gamma(X, \mathcal{M}^\pm(-E)) \longrightarrow \Gamma(X, \mathcal{M}^\pm) \longrightarrow \Gamma(E, \mathcal{M}^\pm) \longrightarrow 0,$$

and we can express $\mathrm{N}_{E/k}$ via the isomorphism

$$\mathrm{N}_{E/k}\mathcal{M} \cong \mathrm{Hom}_k\bigl(\det_k \Gamma(E, \mathcal{M}^-), \det_k \Gamma(E, \mathcal{M}^+)\bigr) \tag{2.9}$$

deduced from (2.6). We fix k-bases of $\Gamma(E, \mathcal{M}^-)$ and $\Gamma(E, \mathcal{M}^+)$. From the induced trivialisations of $\det_k \Gamma(E, \mathcal{M}^\pm)$ we then obtain a trivialisation of $\mathrm{N}_{E/k}\mathcal{M}$.

We now consider three line bundles $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{P}$ together with an isomorphism

$$\mu\colon \mathcal{M} \otimes_{\mathcal{O}_X} \mathcal{N} \xrightarrow{\sim} \mathcal{P}.$$

By the linearity of the norm functor, μ induces an isomorphism

$$\mathrm{N}_{E/k}\mathcal{M} \otimes_k \mathrm{N}_{E/k}\mathcal{N} \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{P}. \tag{2.10}$$

As above, we choose isomorphisms

$$\mathcal{M} \cong \mathcal{H}om_{\mathcal{O}_X}(\mathcal{M}^-, \mathcal{M}^+), \quad \mathcal{N} \cong \mathcal{H}om_{\mathcal{O}_X}(\mathcal{N}^-, \mathcal{N}^+), \quad \mathcal{P} \cong \mathcal{H}om_{\mathcal{O}_X}(\mathcal{P}^-, \mathcal{P}^+)$$

on X, where $\mathcal{M}^\pm$, $\mathcal{N}^\pm$ and $\mathcal{P}^\pm$ are line bundles of degree at least $\deg E + 2g + 1$. We fix bases of the six k-vector spaces

$$\Gamma(E, \mathcal{M}^\pm), \quad \Gamma(E, \mathcal{N}^\pm), \quad \Gamma(E, \mathcal{P}^\pm).$$

Then (2.9) gives trivialisations of $\mathrm{N}_{E/k}\mathcal{M}$, $\mathrm{N}_{E/k}\mathcal{N}$ and $\mathrm{N}_{E/k}\mathcal{P}$. Under these trivialisations, the isomorphism (2.10) equals multiplication by some element $\lambda \in k^\times$.

To find an expression for λ, we choose generators $\alpha_{\mathcal{M}}^\pm$ and $\alpha_{\mathcal{N}}^\pm$ of $\Gamma(E, \mathcal{M}^\pm)$ and $\Gamma(E, \mathcal{N}^\pm)$. To these we associate the isomorphisms

$$\alpha_{\mathcal{M}}\colon \Gamma(E, \mathcal{M}^-) \xrightarrow{\sim} \Gamma(E, \mathcal{M}^+)$$

and

$$\alpha_{\mathcal{N}}\colon \Gamma(E, \mathcal{N}^-) \xrightarrow{\sim} \Gamma(E, \mathcal{N}^+)$$

sending $\alpha_{\mathcal{M}}^-$ to $\alpha_{\mathcal{M}}^+$ and $\alpha_{\mathcal{N}}^-$ to $\alpha_{\mathcal{N}}^+$, respectively. Viewing $\alpha_{\mathcal{M}}$ and $\alpha_{\mathcal{N}}$ as generators of $\Gamma(E, \mathcal{M})$ and $\Gamma(E, \mathcal{N})$ and applying the isomorphism

$$\mu\colon \Gamma(E, \mathcal{M}) \otimes_{\Gamma(E, \mathcal{O}_E)} \Gamma(E, \mathcal{N}) \xrightarrow{\sim} \Gamma(E, \mathcal{P})$$

to $\alpha_{\mathcal{M}} \otimes \alpha_{\mathcal{N}}$ we obtain a generator of $\Gamma(E, \mathcal{P})$, which we can identify with an isomorphism

$$\alpha_{\mathcal{P}}\colon \Gamma(E, \mathcal{P}^-) \xrightarrow{\sim} \Gamma(E, \mathcal{P}^+).$$

We define $\delta_{\mathcal{M}}$ as the determinant of the matrix of $\alpha_{\mathcal{M}}$ with respect to the chosen bases. Under the given trivialisations of $\mathrm{N}_{E/k}\mathcal{M}$, the element $\mathrm{N}^{\mathcal{M}}_{E/k}\alpha_{\mathcal{M}}$ corresponds to $\delta_{\mathcal{M}}$. The same goes for $\mathcal{N}$ and $\mathcal{P}$. On the other hand, the isomorphism (2.10) maps $\mathrm{N}^{\mathcal{M}}_{E/k}\alpha_{\mathcal{M}} \otimes \mathrm{N}^{\mathcal{N}}_{E/k}\alpha_{\mathcal{N}}$ to $\mathrm{N}^{\mathcal{P}}_{E/k}\alpha_{\mathcal{P}}$. We conclude that we can express λ as

$$\lambda = \frac{\delta_{\mathcal{P}}}{\delta_{\mathcal{M}}\delta_{\mathcal{N}}}. \tag{2.11}$$

Let us turn the above discussion into an algorithm. Let X be a projective curve over k, embedded via a line bundle $\mathcal{L}$ as in § 2.1, and let E be an effective divisor on X. For simplicity, we restrict to the case where

$$\deg E = \deg \mathcal{L}.$$

We consider line bundles

$$\mathcal{M} = \mathcal{L}^{\otimes i}(-D_1) \quad \text{and} \quad \mathcal{N} = \mathcal{L}^{\otimes j}(-D_2),$$

where i and j are non-negative integers and D_1 and D_2 are effective divisors such that

$$\deg D_1 = i \deg \mathcal{L} \quad \text{and} \quad \deg D_2 = j \deg \mathcal{L}.$$

We take

$$\mathcal{M}^- = \mathcal{N}^- = \mathcal{P}^- = \mathcal{L}^{\otimes 2}$$

and

$$\mathcal{M}^+ = \mathcal{L}^{\otimes(i+2)}(-D_1), \quad \mathcal{N}^+ = \mathcal{L}^{\otimes(j+2)}(-D_2),$$
$$\mathcal{P}^+ = \mathcal{L}^{\otimes(i+j+2)}(-D_1 - D_2).$$

Algorithm 2.9 *(Linearity of the norm functor).* Let X be a projective curve over a field k, and let E, D_1 and D_2 be effective divisors on X such that

$$\deg E = \deg \mathcal{L}, \quad \deg D_1 = i \deg \mathcal{L}, \quad \deg D_2 = j \deg \mathcal{L}.$$

Fix bases of the four k-vector spaces

$$\Gamma(E, \mathcal{L}^{\otimes 2}), \quad \Gamma(E, \mathcal{L}^{\otimes(i+2)}(-D_1)),$$
$$\Gamma(E, \mathcal{L}^{\otimes(j+2)}(-D_2)), \quad \Gamma(E, \mathcal{L}^{\otimes(i+j+2)}(-D_1 - D_2)).$$

and consider the corresponding trivialisations

$$t_1\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}^{\otimes i}(-D_1), \quad t_2\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}^{\otimes j}(-D_2),$$
$$t_3\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}^{\otimes i+j}(-D_1 - D_2)$$

defined by (2.9). Given the k-algebra $S_X^{(i+j+4)}$, bases for the k-vector spaces

$$\Gamma(X,\mathcal{L}^{\otimes 2}),\quad \Gamma(X,\mathcal{L}^{\otimes(i+2)}),$$
$$\Gamma(X,\mathcal{L}^{\otimes(j+2)}(-D_2)),\quad \Gamma(X,\mathcal{L}^{\otimes(i+j+2)}(-D_1-D_2))$$

and the quotient maps

$$\begin{aligned}
\Gamma(X,\mathcal{L}^{\otimes 2}) &\longrightarrow \Gamma(E,\mathcal{L}^{\otimes 2}),\\
\Gamma(X,\mathcal{L}^{\otimes(i+2)}(-D_1)) &\longrightarrow \Gamma(E,\mathcal{L}^{\otimes i+2}(-D_1)),\\
\Gamma(X,\mathcal{L}^{\otimes(j+2)}(-D_2)) &\longrightarrow \Gamma(E,\mathcal{L}^{\otimes j+2}(-D_2)),\\
\Gamma(X,\mathcal{L}^{\otimes(i+j+2)}(-D_1-D_2)) &\longrightarrow \Gamma(E,\mathcal{L}^{\otimes i+2}(-D_1))
\end{aligned}$$

as matrices with respect to the given bases, this algorithm outputs the element $\lambda\in k^\times$ such that the diagram

$$\begin{array}{ccc}
k & \xrightarrow[\sim]{t_1\otimes t_2} & \mathrm{N}_{E/k}\mathcal{L}^{\otimes i}(-D_1)\otimes_k \mathrm{N}_{E/k}\mathcal{L}^{\otimes j}(-D_2)\\
\lambda\big\downarrow\wr & & \big\downarrow\wr\\
k & \xrightarrow[\sim]{t_3} & \mathrm{N}_{E/k}\mathcal{L}^{\otimes(i+j)}(-D_1-D_2)
\end{array}$$

is commutative.

1. Compute the spaces

$$\Gamma(E,\mathcal{L}^{\otimes(i+4)}(-D_1))\quad\text{and}\quad\Gamma(E,\mathcal{L}^{\otimes(i+j+4)}(-D_1-D_2))$$

and the multiplication maps

$$\begin{aligned}
\Gamma(E,\mathcal{L}^{\otimes 2})\times\Gamma(E,\mathcal{L}^{\otimes(i+2)}(-D_1))&\to\Gamma(E,\mathcal{L}^{\otimes(i+4)}(-D_1)),\\
\Gamma(E,\mathcal{L}^{\otimes(i+2)}(-D_1))\times\Gamma(E,\mathcal{L}^{\otimes(j+2)}(-D_2))&\to\Gamma(E,\mathcal{L}^{\otimes(i+j+4)}(-D_1-D_2)),\\
\Gamma(E,\mathcal{L}^{\otimes 2})\times\Gamma(E,\mathcal{L}^{\otimes(i+j+2)}(-D_1-D_2))&\to\Gamma(E,\mathcal{L}^{\otimes(i+j+4)}(-D_1-D_2)).
\end{aligned}$$

2. Apply the probabilistic method described in § 1.2 to the bilinear maps just computed to find generators β_0, β_1 and β_2 of the free $\Gamma(E,\mathcal{O}_E)$-modules $\Gamma(E,\mathcal{L}^{\otimes 2})$, $\Gamma(E,\mathcal{L}^{\otimes(i+2)}(-D_1))$ and $\Gamma(E,\mathcal{L}^{\otimes(j+2)}(-D_2))$ of rank 1.

(Note that we do not need the k-algebra structure on $\Gamma(E,\mathcal{L}^{\otimes 2})$. If k is small, we may have to extend the base field, but it is easy to see that this is not a problem.)

3. Compute the matrix (with respect to the given bases) of the isomorphism α_1 defined by the commutative diagram

$$\begin{array}{ccc}
\Gamma(E,\mathcal{L}^{\otimes 2}) & \xrightarrow[\sim]{\alpha_1} & \Gamma(E,\mathcal{L}^{\otimes(i+2)}(-D_1))\\
\Big\| & & \wr\big\downarrow\cdot\beta_0\\
\Gamma(E,\mathcal{L}^{\otimes 2}) & \xrightarrow[\sim]{\cdot\beta_1} & \Gamma(E,\mathcal{L}^{\otimes(i+4)}(-D_1)),
\end{array}$$

of the isomorphism α_2 defined by the similar diagram for $\mathcal{L}^{\otimes j}(-D_2)$ instead of $\mathcal{L}^{\otimes i}(-D_1)$ and of the isomorphism α_3 defined by the commutative diagram

$$\begin{array}{ccc} \Gamma(E, \mathcal{L}^{\otimes 2}) & \xrightarrow[\sim]{\alpha_3} & \Gamma(E, \mathcal{L}^{\otimes(i+j+2)}(-D_1 - D_2)) \\ \alpha_1 \downarrow \wr & & \wr \downarrow \cdot\beta_0 \\ \Gamma(E, \mathcal{L}^{\otimes(i+2)}(-D_1)) & \xrightarrow[\sim]{\cdot\beta_2} & \Gamma(E, \mathcal{L}^{\otimes(i+j+4)}(-D_1 - D_2)). \end{array}$$

4. Compute the elements δ_1, δ_2 and δ_3 of $k^\times$ as the determinants of the matrices of α_1, α_2 and α_3 computed in the previous step.

5. Output the element $\dfrac{\delta_3}{\delta_1\delta_2} \in k^\times$.

Analysis. We note that β_0 plays the role of $\alpha^-_{\mathcal{M}}$, $\alpha^-_{\mathcal{N}}$ and $\alpha^-_{\mathcal{P}}$ in the notation of the discussion preceding the algorithm, and that β_1, β_2 and $\beta_1\beta_2/\beta_0$ play the roles of $\alpha^+_{\mathcal{M}}$, $\alpha^+_{\mathcal{N}}$ and $\alpha^+_{\mathcal{P}}$. This means that α_1, α_2 and α_3 are equal to $\alpha_{\mathcal{M}}$, $\alpha_{\mathcal{N}}$ and $\alpha_{\mathcal{P}}$. It now follows from (2.11) that the output of the algorithm is indeed equal to λ. It is clear that the algorithm runs in (probabilistic) polynomial time in $\deg \mathcal{L}$, i and j (measured in field operations in k). ◇

2.8. Computing in the Picard group of a curve

We now turn to the question of computing with elements in the Picard group of a curve X, using the operations on divisors described in the first part of this section. We only consider the group $\operatorname{Pic}^0 X$ of isomorphism classes of line bundles of degree 0. This group can be identified in a canonical way with a subgroup of rational points of the Jacobian variety of X. If X has a rational point, then this subgroup consists of all the rational points of the Jacobian.

We will only describe Khuri-Makdisi's *medium model* of $\operatorname{Pic}^0 X$ relative to a fixed line bundle $\mathcal{L}$ of degree

$$\deg \mathcal{L} \geq 2g + 1,$$

but at the same time

$$\deg \mathcal{L} \leq c(g + 1)$$

for some constant $c \geq 1$, as described in Khuri-Makdisi [56, § 5].

Remark. Khuri-Makdisi starts with a divisor D_0 whose degree satisfies the above inequalities and takes $\mathcal{L} = \mathcal{O}_X(D_0)$. This is of course only a matter of language. Another difference in notation is that Khuri-Makdisi writes $\mathcal{L}_0$ for $\mathcal{L}$ and uses the notation $\mathcal{L}$ for $\mathcal{L}_0^{\otimes 2}$ (in the medium model) or $\mathcal{L}_0^{\otimes 3}$ (in the *large* and *small* models, which we do not describe here).

We represent elements of $\operatorname{Pic}^0 X$ by effective divisors of degree $\deg \mathcal{L}$ as follows: the isomorphism class of a line bundle $\mathcal{M}$ of degree 0 is represented by the divisor of some global section of the line bundle $\mathcal{H}om(\mathcal{M}, \mathcal{L})$ of degree $\deg \mathcal{L}$, i.e. by any effective divisor D such that

$$\mathcal{M} \cong \mathcal{L}(-D).$$

It follows from the inequality $\deg \mathcal{L} \geq 2g$ that we can represent any effective divisor D of degree $\deg \mathcal{L}$ by the subspace $\Gamma(X, \mathcal{L}^{\otimes 2}(-D))$ of codimension $\deg \mathcal{L}$ in $\Gamma(X, \mathcal{L}^{\otimes 2})$.

There are a few basic operations:

- *membership test*: given a subspace W of codimension $\deg \mathcal{L}$ in $\Gamma(X, \mathcal{L}^{\otimes 2})$, decide whether W represents an element of $\operatorname{Pic}^0 X$, i.e. whether W is of the form $\Gamma(X, \mathcal{L}^{\otimes 2}(-D))$ for an effective divisor D of degree $\deg \mathcal{L}$.
- *zero test*: given a subspace W of codimension $\deg \mathcal{L}$ in $\Gamma(X, \mathcal{L}^{\otimes 2})$, decide whether W represents the zero element of $\operatorname{Pic}^0 X$.
- *zero element*: output a subspace of codimension $\deg \mathcal{L}$ in $\Gamma(X, \mathcal{L}^{\otimes 2})$ representing the element $0 \in \operatorname{Pic}^0 X$.
- *addflip*: given two subspaces of $\Gamma(X, \mathcal{L}^{\otimes 2})$ representing elements $x, y \in \operatorname{Pic}^0 X$, compute a subspace of $\Gamma(X, \mathcal{L}^{\otimes 2})$ representing the element $-x - y$.

From the "addflip" operation, one immediately gets negation ($-x = -x - 0$), addition ($x + y = -(-x - y)$) and subtraction ($x - y = -(-x) - y$). Clearly, one can test whether two elements x and y are equal by computing $x - y$ and testing whether the result equals zero.

Remark. With regard to actual implementations of the above algorithms, we note that some of the operations can be implemented in a more efficient way than by composing the basic operations just described. We refer to [57] for details.

By Khuri-Makdisi's results in [57], the above operations can be implemented using randomised algorithms with expected running time of $O(g^{3+\epsilon})$ for any $\epsilon > 0$, measured in operations in the field k. This can be improved to $O(g^{2.376})$ by means of fast linear algebra algorithms. (The exponent 2.376 is an upper bound for the complexity of matrix multiplication.)

Multiplication by an integer n can be done efficiently by means of an *addition chain* for n. This is a sequence of positive integers $(a_1, a_2, \ldots, a_m)$ with $a_1 = 1$ and $a_m = n$ such that for each $l > 1$ there exist $i(l)$ and $j(l)$ in $\{1, 2, \ldots, l-1\}$ such that $a_l = a_{i(l)} + a_{j(l)}$. (We consider the indices $i(l)$ and $j(l)$ as given together with the addition chain.) The integer m is called the *length* of the addition chain. A more general and often slightly more efficient method of multiplying by n is to use an *addition-subtraction chain*, where a_l is allowed to be either $a_{i(l)} + a_{j(l)}$ or $a_{i(l)} - a_{j(l)}$. However, since the "addflip" operation in our set-up takes less time than the addition or subtraction algorithms, the most worthwhile option is to use an *anti-addition chain*, which is a sequence of (not necessarily positive) integers $(a_0, a_1, \ldots, a_m)$ such that

$$a_l = \begin{cases} 0 & \text{if } l = 0; \\ 1 & \text{if } l = 1; \\ -a_{i(l)} - a_{j(l)} & \text{if } 2 \leq l \leq m \end{cases}$$

and $a_m = n$; the $i(l)$ and $j(l)$ are given elements of $\{0, 1, \ldots, l-1\}$ for $2 \leq l \leq m$.

It is well known that for every positive integer n there exists an addition chain of length $O(\log n)$, and there are algorithms (such as the *binary method* used in repeated squaring) to find such an addition chain in time $O((\log n)^2)$. We leave it to the reader to write down a similar algorithm for finding an anti-addition chain.

For later use, we give versions of the "zero test" and "addflip" algorithms that are identical to those given by Khuri-Makdisi, except that some extra information computed in the course of the algorithm is part of the output.

Algorithm 2.10 *(Zero test).* Let X be a projective curve over a field k, and let x be an element of $\operatorname{Pic}^0 X$. Given the k-algebra $S_X^{(2)}$ and a subspace $\Gamma(\mathcal{L}_X^{\otimes 2}(-D))$ of $\Gamma(\mathcal{L}_X^{\otimes 2})$ representing x, this algorithm outputs `false` if $x \neq 0$ (i.e. if the line bundle $\mathcal{L}_X(-D)$ is non-trivial). If $\mathcal{L}_X(-D)$ is trivial, the algorithm outputs a pair $(\mathtt{true}, s)$, where s is a global section of $\mathcal{L}_X$ with divisor D.

1. Compute the space

$$\Gamma(\mathcal{L}_X(-D)) = \bigl\{s \in \Gamma(\mathcal{L}_X) \bigm| s\Gamma(\mathcal{L}_X) \subseteq \Gamma(\mathcal{L}_X^{\otimes 2}(-D))\bigr\}.$$

(The truth of this equality follows from Lemma 2.2.)

2. If $\Gamma(\mathcal{L}_X(-D)) = 0$, output `false`. Otherwise, output $(\mathtt{true}, s)$, where s is any non-zero element of the one-dimensional k-vector space $\Gamma(\mathcal{L}_X(-D))$.

Algorithm 2.11 *(Addflip).* Let X be a projective curve over a field k, and let x and y be elements of $\operatorname{Pic}^0 X$. Given the k-algebra $S_X^{(5)}$ and subspaces $\Gamma(\mathcal{L}_X^{\otimes 2}(-D))$ and $\Gamma(\mathcal{L}_X^{\otimes 2}(-E))$ of $\Gamma(\mathcal{L}_X^{\otimes 2})$ representing x and y, this algorithm outputs a subspace $\Gamma(\mathcal{L}_X^{\otimes 2}(-F))$ representing $-x-y$, as well as a global section s of $\mathcal{L}_X^{\otimes 3}$ such that

$$\operatorname{div} s = D + E + F.$$

1. Compute $\Gamma(\mathcal{L}_X^{\otimes 4}(-D-E))$ as the product of $\Gamma(\mathcal{L}_X^{\otimes 2}(-D))$ and $\Gamma(\mathcal{L}_X^{\otimes 2}(-E))$ (see Lemma 2.1).

2. Compute the space

$$\Gamma(\mathcal{L}_X^{\otimes 3}(-D-E)) = \bigl\{s \in \Gamma(\mathcal{L}_X^{\otimes 3}) \bigm| s\Gamma(\mathcal{L}_X) \subseteq \Gamma(\mathcal{L}_X^{\otimes 4}(-D-E))\bigr\}$$

(see Lemma 2.2).

3. Choose any non-zero $s \in \Gamma(\mathcal{L}_X^{\otimes 3}(-D-E))$. Let F denote the divisor of s as a global section of $\mathcal{L}_X^{\otimes 3}(-D-E)$.

4. Compute the space

$$\Gamma(\mathcal{L}_X^{\otimes 5}(-D-E-F)) = s\Gamma(\mathcal{L}_X^{\otimes 2}).$$

5. Compute the space

$$\Gamma(\mathcal{L}_X^{\otimes 2}(-F)) = \bigl\{t \in \Gamma(\mathcal{L}_X^{\otimes 2}) \bigm| t\Gamma(\mathcal{L}_X^{\otimes 3}(-D-E)) \subseteq \Gamma(\mathcal{L}_X^{\otimes 5}(-D-E-F))\bigr\}$$

(see again Lemma 2.2).

6. Output the space $\Gamma(\mathcal{L}_X^{\otimes 2}(-F))$ and the section $s \in \Gamma(\mathcal{L}_X^{\otimes 3})$.

2.9. Normalised representatives of elements of the Picard group

Let X be a projective curve over a field k in the sense of § 2.1, and let O be a k-rational point of X. Let x be an element of $\operatorname{Pic}^0 X$, and let $\mathcal{M}$ be a line bundle representing x. Let $r_x^{\mathcal{L}_X,O}$ be the greatest integer r such that

$$\Gamma(\mathcal{H}om(\mathcal{M}, \mathcal{L}_X(-rO))) \neq 0.$$

Then $\Gamma(\mathcal{H}om_{\mathcal{O}_X}(\mathcal{M}, \mathcal{L}_X(-r_x^{\mathcal{L}_X,O}O)))$ is one-dimensional, so there exists a unique effective divisor R such that

$$\mathcal{M} \cong \mathcal{L}_X(-R - r_x^{\mathcal{L}_X,O}O).$$

We define the $(\mathcal{L}_X, O)$-*normalised representative* of x as the effective divisor

$$R_x^{\mathcal{L}_X,\mathcal{O}} = R + r_x^{\mathcal{L}_X,O}O$$

of degree $\deg \mathcal{L}_X$; it is a canonically defined divisor (depending on O) with the property that x is represented by $\mathcal{L}_X(-R_x^{\mathcal{L}_X,O})$.

Remark. Since for any line bundle $\mathcal{N}$ we have

$$\deg \mathcal{N} \geq g \implies \Gamma(\mathcal{N}) \neq 0$$

and

$$\deg \mathcal{N} < 0 \implies \Gamma(\mathcal{N}) = 0,$$

the integer $r_x^{\mathcal{L}_X,O}$ satisfies

$$\deg \mathcal{L}_X - g_X \leq r_x^{\mathcal{L}_X,O} \leq \deg \mathcal{L}_X.$$

Algorithm 2.12 *(Normalised representative).* Let X be a projective curve over a field k, and let O be a k-rational point of X. Let x be an element of $\operatorname{Pic}^0 X$, and let $R_x^{\mathcal{L}_X,O}$ be the $(\mathcal{L}_X, O)$-normalised representative of x. Given the k-algebra $S_X^{(4)}$, the space $\Gamma(\mathcal{L}_X^{\otimes 2}(-O))$ and a subspace of $\Gamma(\mathcal{L}_X^{\otimes 2})$ representing x, this algorithm outputs the integer $r_x^{\mathcal{L}_X,O}$ and the subspace $\Gamma(\mathcal{L}_X^{\otimes 2}(-R_x^{\mathcal{L}_X,O}))$ of $\Gamma(\mathcal{L}_X^{\otimes 2})$.

1. Using the negation algorithm, find a subspace $\Gamma(\mathcal{L}_X^{\otimes 2}(-D))$ of $\Gamma(\mathcal{L}_X^{\otimes 2})$ representing $-x$. Put $r = \deg \mathcal{L}_X$.
2. Compute the space $\Gamma(\mathcal{L}_X^{\otimes 2}(-rO))$, then compute $\Gamma(\mathcal{L}_X^{\otimes 4}(-D - rO))$ as the product of $\Gamma(\mathcal{L}_X^{\otimes 2}(-D))$ and $\Gamma(\mathcal{L}_X^{\otimes 2}(-rO))$, and then compute
$$\Gamma(\mathcal{L}_X^{\otimes 2}(-D - rO)) = \{t \in \Gamma(\mathcal{L}_X^{\otimes 2}) \mid t\Gamma(\mathcal{L}_X^{\otimes 2}) \subseteq \Gamma(\mathcal{L}_X^{\otimes 4}(-D - rO))\}.$$
3. If $\Gamma(\mathcal{L}_X^{\otimes 2}(-D - rO)) = 0$, decrease r by 1 and go to step 2.
4. Let s be a non-zero element of $\Gamma(\mathcal{L}_X^{\otimes 2}(-D - rO))$. Compute
$$\Gamma(\mathcal{L}_X^{\otimes 4}(-D - R_x^{\mathcal{L}_X,O})) = s\Gamma(\mathcal{L}_X^{\otimes 2}),$$
and then compute
$$\Gamma(\mathcal{L}_X^{\otimes 2}(-R_x^{\mathcal{L}_X,O})) = \{t \in \Gamma(\mathcal{L}_X^{\otimes 2}) \mid t\Gamma(\mathcal{L}_X^{\otimes 2}(-D)) \subseteq \Gamma(\mathcal{L}_X^{\otimes 4}(-D - R_x^{\mathcal{L}_X,O}))\},$$
5. Output $r_x^{\mathcal{L}_X,O} = r$ and $\Gamma(\mathcal{L}_X^{\otimes 2}(-R_x^{\mathcal{L}_X,O}))$.

Analysis. It follows from the definition of $R_x^{\mathcal{L}_X,O}$ that this algorithm is correct. It is straightforward to check that its running time, measured in operations in k, is polynomial in $\deg \mathcal{L}_X$. ◇

Let us give a variant involving a different kind of representative, in which the line bundle $\mathcal{L}_X$ does not play a role. Again we fix a k-rational point O of X. Let x be an element of $\operatorname{Pic}^0 X$, and let $\mathcal{M}_x$ be a line bundle on X representing x, and let d_x^O denote the least non-negative integer d such that

$$\dim_k \mathrm{H}^0(X, \mathcal{M}_x(dO)) = 1.$$

Let s be a non-zero global section of $\mathcal{M}_x(d_x^O O)$, and define

$$D_x^O = \operatorname{div}(s) + (g_X - d_x^O)O.$$

This is an effective divisor of degree g_X, and is independent of the choice of s since $\mathrm{H}^0(X, \mathcal{M}_x(dO))$ is one-dimensional. We call D_x^O the *O-normalised representative* of x. It is straightforward to check that

$$d_x^O = \deg \mathcal{L}_X - r_{\hat{x}}^{\mathcal{L}_X,O}$$

and that

$$D_x^O = R_{\hat{x}}^{\mathcal{L}_X,O} - (\deg \mathcal{L}_X - g_X)O, \tag{2.12}$$

where $\hat{x} \in \operatorname{Pic}^0 X$ is defined by

$$\hat{x} = [\mathcal{L}_X(-(\deg \mathcal{L}_X)O)] - x.$$

This means that we can compute D_x^O by finding $\hat{x}$, computing its $(\mathcal{L}_X, O)$-normalised representative and then using (2.12).

2.10. Descent of elements of the Picard group

Now let k' be a finite extension of k, and write

$$X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k'.$$

Consider the natural group homomorphism

$$i\colon \operatorname{Pic}^0 X \to \operatorname{Pic}^0 X'.$$

It is injective since a line bundle $\mathcal{L}$ of degree 0 on X is trivial if and only if $\Gamma(X, \mathcal{L}) \neq 0$, and this is equivalent to the corresponding condition over k'.

Let x' be an element of $\operatorname{Pic}^0 X'$. We now explain how to use normalised representatives to decide whether x' lies in the image of i, and if so, to find the unique element $x \in \operatorname{Pic}^0 X$ such that $x' = i(x)$.

Algorithm 2.13 *(Descent).* Let X be a projective curve over a field k, and let O be a k-rational point of X. Let k' be a finite extension of k, write

$$X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k',$$

and let $\mathcal{L}_{X'}$ denote the pull-back of the line bundle $\mathcal{L}_X$ to X'. Let x' be an element of $\operatorname{Pic}^0 X'$. Given the k-algebra $S_X^{(4)}$, the spaces

$$\Gamma(X, \mathcal{L}_X^{\otimes 2}(-rO)) \quad \text{for} \quad \deg \mathcal{L}_X - g_X \leq r \leq \deg \mathcal{L}_X$$

and a subspace of $\Gamma(X', \mathcal{L}_{X'}^{\otimes 2})$ representing x', this algorithm outputs `false` if x' is not in the image of the canonical map

$$i\colon \operatorname{Pic}^0 X \to \operatorname{Pic}^0 X'.$$

Otherwise, the algorithm outputs $(\texttt{true}, \Gamma(X, \mathcal{L}_X^{\otimes 2}(-D)))$, where $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-D))$ represents the unique element $x \in \operatorname{Pic}^0 X$ such that $i(x) = x'$.

1. Compute the $(\mathcal{L}_{X'}, O)$-normalised representative $R_{x'}^{\mathcal{L}_{X'}, O}$ of x'.
2. Compute the k-vector space

$$V = \Gamma(X', \mathcal{L}_{X'}^{\otimes 2}(-R_x)) \cap \Gamma(X, \mathcal{L}_X^{\otimes 2}).$$

3. If the codimension of V in $\Gamma(X, \mathcal{L}_X^{\otimes 2})$ is less than $\deg \mathcal{L}_X$, output `false`; otherwise, output $(\texttt{true}, V)$.

Analysis. In step 3, we check whether $R_x^{\mathcal{L}_X, O}$ is defined over k or, equivalently, whether x is defined over k. If this is the case, the space V equals $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-R_x))$, where x is the unique element of $\operatorname{Pic}^0 X$ such that $i(x) = x'$. This shows that the algorithm is correct; its running time, measured in operations in k and k', is clearly polynomial in $\deg \mathcal{L}_X$. ◇

2.11. Computing Picard and Albanese maps

A finite morphism

$$f\colon X \to Y$$

between complete, smooth, geometrically connected curves over a field k induces two group homomorphisms

$$\operatorname{Pic} f\colon \operatorname{Pic}^0 Y \to \operatorname{Pic}^0 X$$

and

$$\operatorname{Alb} f\colon \operatorname{Pic}^0 X \to \operatorname{Pic}^0 Y,$$

called the *Picard* and *Albanese* maps, respectively. In terms of line bundles, they can be described as follows. The Picard map sends the class of a line bundle $\mathcal{N}$ on Y to the class of the line bundle $f^*\mathcal{N}$ on X, and the Albanese map sends the class of a line bundle $\mathcal{M}$ on X to the class of the line bundle $\mathrm{N}_f\mathcal{M}$ on Y.

Alternatively, these maps can be described in terms of divisor classes as follows. The group homomorphisms

$$f_*\colon \operatorname{Div}^0 X \to \operatorname{Div}^0 Y \quad \text{and} \quad f^*\colon \operatorname{Div}^0 Y \to \operatorname{Div}^0 X$$

between the groups of divisors of degree 0 on X and Y respect the relation of linear equivalence on both sides. The Picard map sends the class of a divisor E on Y to the class of the divisor f^*E on X, and the Albanese map sends the class of a divisor D on X to the class of the divisor f_*D on Y.

Let us now assume that $f\colon X \to Y$ is a finite morphism of *projective* curves in the sense of § 2.5; in particular, we are given an isomorphism $f^*\mathcal{L}_Y \xrightarrow{\sim} \mathcal{L}_X$. Using the following algorithms, we can compute the maps $\operatorname{Pic} f$ and $\operatorname{Alb} f$. The algorithm for the Albanese map actually only reduces the problem to a different one, namely that of computing *traces* in Picard groups with respect to finite extensions of the base field. If A is an Abelian variety over a field k and k' is a finite extension of k, then the trace of an element $y \in A(k')$ is defined by

$$\operatorname{tr}_{k'/k} y = [k' : k]_{\mathrm{i}} \sum_\sigma \sigma(y),$$

where σ runs over all k-embeddings of k' into an algebraic closure of k and $[k' : k]_{\mathrm{i}}$ is the inseparable degree of k' over k. Computing traces is a problem that can be solved at least for finite fields, as we will see in § 3.4.

Algorithm 2.14 *(Picard map).* Let $f\colon X \to Y$ be a finite morphism of projective curves, and let y be an element of $\operatorname{Pic}^0 Y$. Given the k-algebras $S_X^{(4)}$ and $S_Y^{(4)}$, the homomorphism $f^\#\colon S_Y^{(4)} \to S_X^{(4)}$ and a subspace $\Gamma(Y, \mathcal{L}_Y^{\otimes 2}(-E))$ of $\Gamma(Y, \mathcal{L}_Y^{\otimes 2})$ representing y, this algorithm outputs a subspace of $\Gamma(X, \mathcal{L}_X^{\otimes 2})$ representing $(\operatorname{Pic} f)(y) \in \operatorname{Pic}^0 X$.

1. Compute the subspace $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-D))$ for the divisor $D = f^*E$ using Algorithm 2.6 (taking $i = j = 2$ in the notation of that algorithm), and output the result.

Analysis. Since $(\operatorname{Pic} f)(y)$ is represented by the line bundle $\mathcal{L}_X(-f^*D)$, the correctness of this algorithm follows from that of Algorithm 2.6. Furthermore, the running time of Algorithm 2.6, measured in operations in k, is polynomial in $\deg \mathcal{L}_X$ for fixed i and j; therefore, the running time of this algorithm is also polynomial in $\deg \mathcal{L}_X$. $\diamond$

Algorithm 2.15 *(Albanese map).* Let $f\colon X \to Y$ be a finite morphism of projective curves over a field k. Let x be an element of $\operatorname{Pic}^0 X$, and let O be a k-rational point of Y. Suppose that we have a (probabilistic) algorithm to compute the primary decomposition of a finite commutative k-algebra A with (expected) running time polynomial in $[A : k]$, measured in operations in k. Suppose furthermore that for any finite separable extension k' of k and any element $y \in \operatorname{Pic}^0(Y_{k'})$, we can compute $\operatorname{tr}_{k'/k} y$ in time polynomial in $\deg \mathcal{L}_Y$ and $[k' : k]$, measured in operations in k. Given the k-algebras $S_X^{(6)}$ and $S_Y^{(6)}$, the homomorphism $f^\#\colon S_Y^{(6)} \to S_X^{(6)}$, the space $\Gamma(Y, \mathcal{L}_Y^{\otimes 2}(-O))$ and a subspace $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-D))$ of $\Gamma(X, \mathcal{L}_X^{\otimes 2})$ representing x, this algorithm outputs a subspace of $\Gamma(Y, \mathcal{L}_Y^{\otimes 2})$ representing $(\operatorname{Alb} f)(x) \in \operatorname{Pic}^0 Y$.

1. Compute $\Gamma(X, \mathcal{L}_X^{\otimes 4}(-D))$ as the product of $\Gamma(X, \mathcal{L}_X^{\otimes 2})$ and $\Gamma(X, \mathcal{L}_X^{\otimes 2}(-D))$.
2. Find the decomposition of D as a linear combination $\sum_P n_P P$ of prime divisors using Algorithm 2.4.
3. For each P occurring in the support of D:
4. Compute the base changes $X_{k(P)}$ and $Y_{k(P)}$.
5. Find the primary decomposition of the divisor $P_{k(P)}$ on $X_{k(P)}$, and pick a rational point P' in it.
6. Compute the space $\Gamma(Y_{k(P)}, \mathcal{L}_Y^{\otimes 2}(-f(P') - (\deg \mathcal{L}_Y - 1)O))$; this represents an element $y_{P'} \in \operatorname{Pic}^0(Y_{k(P)})$.
7. Compute the element $y_P = \operatorname{tr}_{k(P)/k} y_{P'}$ of $\operatorname{Pic}^0 Y_{k(P)}$. Apply Algorithm 2.13 to get a representation for y_P as an element of $\operatorname{Pic}^0 Y$.
8. Compute the element $y = \sum_P n_P y_P$ of $\operatorname{Pic}^0(Y)$.
9. Output the element $y - (\deg f)(\deg \mathcal{L}_Y - 1)y_0$ of $\operatorname{Pic}^0 Y$, where y_0 is the element of $\operatorname{Pic}^0 Y$ represented by $\Gamma(Y, \mathcal{L}_Y^{\otimes 2}(-(\deg \mathcal{L}_Y)O))$.

Analysis. The definition of $y_{P,i}$ implies that

$$y_{P'} = [\mathcal{L}_Y(-f(P') - (\deg \mathcal{L}_Y - 1)O)],$$

the definition of y_P implies that

$$y_P = [\mathcal{L}_Y^{\otimes [k(P):k]}(-f_*P - [k(P):k](\deg \mathcal{L}_Y - 1)O)]$$

and the definition of y, together with the fact that $\deg \mathcal{L}_X = (\deg f)(\deg \mathcal{L}_Y)$ implies that

$$\begin{aligned} y &= [\mathcal{L}_Y^{\otimes \deg \mathcal{L}_X}(-f_*D - (\deg \mathcal{L}_X)(\deg \mathcal{L}_Y - 1)O)] \\ &= [\mathcal{L}_Y^{\otimes \deg f}(-f_*D)] + (\deg f)(\deg \mathcal{L}_Y - 1)[\mathcal{L}_Y(-(\deg \mathcal{L}_Y)O)]. \end{aligned}$$

Together with the definition of y_0, this shows that

$$\begin{aligned} y - (\deg f)(\deg \mathcal{L}_Y - 1)y_0 &= [\mathcal{L}_Y^{\otimes \deg D}(-f_*D)] \\ &= \mathrm{N}_f \mathcal{L}_X(-D), \end{aligned}$$

and therefore that the output of the algorithm is indeed $(\operatorname{Alb} f)(x)$. Our computational assumptions imply that the running time is polynomial in $\deg \mathcal{L}_X$, measured in field operations in k. ◇

Finally we consider correspondences, i.e. diagrams of the form

$$\begin{array}{ccc} & X & \\ {\scriptstyle f}\swarrow & & \searrow{\scriptstyle g} \\ Y & & Z, \end{array}$$

where X, Y and Z are proper, smooth, geometrically connected curves over a field k. Such a correspondence induces group homomorphisms

$$\operatorname{Alb} g \circ \operatorname{Pic} f \colon \operatorname{Pic}^0 Y \to \operatorname{Pic}^0 Z$$

and

$$\operatorname{Alb} f \circ \operatorname{Pic} g \colon \operatorname{Pic}^0 Z \to \operatorname{Pic}^0 Y.$$

We suppose that X, Y and Z are given by projective embeddings using line bundles $\mathcal{L}_X$, $\mathcal{L}_Y$ and $\mathcal{L}_Z$ as in § 2.1, and that we are given isomorphisms

$$f^*\mathcal{L}_Y \cong \mathcal{L}_X \cong g^*\mathcal{L}_Z.$$

Then $\operatorname{Alb} g \circ \operatorname{Pic} f$ and $\operatorname{Alb} f \circ \operatorname{Pic} g$ can be computed by composing the two algorithms described above.

3. Curves over finite fields

In this section we give algorithms for computing with divisors on a curve over a finite field. After some preliminaries, we show how to compute the Frobenius map on divisors and how to choose uniformly random divisors of a given degree. Then we show how to do various operations in the Picard group of a curve over a finite field, such as choosing random elements, computing the Frey–Rück pairing and finding a basis of the l-torsion for a prime number l. Many of the results in this section, especially those in § 3.7, § 3.8 and § 3.9, are variants of work of Couveignes [16].

We switch from measuring the running time of algorithms in field operations to measuring it in bit operations. The usual field operations in a finite field k can be done in time polynomial in $\log \# k$.

Let k be a finite field of cardinality q, and let X be a complete, smooth, geometrically connected curve of genus g over k. The *zeta function* of X is the power series in $\mathbf{Z}[[t]]$ defined by

$$\begin{array}{ccc} \mathrm{Z}_X = \displaystyle\sum_{D \in \operatorname{Eff} X} t^{\deg D} & = & \displaystyle\sum_{n=0}^{\infty} (\# \operatorname{Eff}^n X) t^n \\ \| & & \| \\ \displaystyle\prod_{P \in \operatorname{PDiv} X} \frac{1}{1 - t^{\deg P}} & = & \displaystyle\prod_{d=1}^{\infty} (1 - t^d)^{-\# \operatorname{PDiv}^d X}. \end{array}$$

Here $\operatorname{Eff} X$ and $\operatorname{PDiv} X$ are the sets of effective divisors and prime divisors on X, respectively; a superscript denotes the subset of divisors of the indicated degree. The following properties of the zeta function are well known.

Theorem 3.1. *Let X be a complete, smooth, geometrically connected curve of genus g over a finite field of cardinality q.*

(1) The power series Z_X can be written as a rational function

$$\mathrm{Z}_X = \frac{L_X}{(1-t)(1-qt)}, \tag{3.1}$$

where $L_X \in \mathbf{Z}[t]$ is a polynomial of the form

$$L_X = 1 + a_1 t + \cdots + a_{2g-1} t^{2g-1} + q^g t^{2g}.$$

(2) *The factorisation of L_X over the complex numbers has the form*

$$L_X = \prod_{i=1}^{2g}(1-\alpha_i t), \tag{3.2}$$

where each α_i has absolute value $\sqrt{q}$.

(3) *The polynomial L_X satisfies the functional equation*

$$q^g t^{2g} L_X(1/qt) = L_X(t). \tag{3.3}$$

From the definition of Z_X and from (3.1) it is clear how one can compute the number of effective divisors of a given degree on X starting from the polynomial L_X. We now show how to extract the number of *prime* divisors of a given degree from L_X. Taking logarithmic derivatives in the definition of Z_X and the expression (3.1), we obtain

$$\frac{\mathrm{Z}'_X}{\mathrm{Z}_X} = \frac{1}{t}\sum_{n=1}^{\infty}\left(\sum_{d\mid n} d\cdot \#\operatorname{PDiv}^d X\right)t^n = \frac{L'_X}{L_X} + \frac{1}{1-t} + \frac{q}{1-qt}. \tag{3.4}$$

From L_X we can compute the coefficients of this power series. We can then compute $\#\operatorname{PDiv}^d X$ using the Möbius inversion formula. More explicitly, taking logarithmic derivatives in the factorisation (3.2), we obtain *Newton's identity*

$$L'_X/L_X = -\sum_{n=0}^{\infty} s_{n+1}t^n,$$

where the s_n are the power sums

$$s_n = \sum_{i=1}^{2g}\alpha_i^n \in \mathbf{Z} \quad (n \ge 0).$$

Expanding the right-hand side of (3.4) in a power series and comparing coefficients, we get

$$\sum_{d\mid n} d\,\#\operatorname{PDiv}^d X = 1 + q^n - s_n,$$

or equivalently, by the Möbius inversion formula,

$$n\,\#\operatorname{PDiv}^n X = \sum_{d\mid n}\mu(n/d)(1+q^d-s_d),$$

where μ is the usual Möbius function. We note that this simplifies to

$$\#\operatorname{PDiv}^n X = \begin{cases} 1+q-s_1 & \text{if } n=1;\\ \frac{1}{n}\sum_{d\mid n}\mu(n/d)(q^d-s_d) & \text{if } n\ge 2.\end{cases} \tag{3.5}$$

Let $J = \operatorname{Pic}^0_{X/k}$ denote the Jacobian variety of X. From the fact that the Brauer group of k vanishes it follows that the canonical inclusion

$$\operatorname{Pic}^0 X \to J(k)$$

is an equality. In other words, every rational point of J can be identified with a linear equivalence class of k-rational divisors of degree 0.

We note that from the functional equation (3.3) one can deduce that

$$\#\operatorname{Eff}^n X = \frac{q^{1-g+n}-1}{q-1} L_X(1) \quad \text{for } n \geq 2g,$$

which in turn is equivalent to the "class number formula"

$$\#J(k) = \#\operatorname{Pic}^0 X = L_X(1). \tag{3.6}$$

3.1. The Frobenius map

Let k be a finite field of cardinality q, and let X be a projective curve over k in the sense of § 2.1. We write $d = \deg \mathcal{L}_X$. Let $\operatorname{Sym}^d X$ denote the d-th symmetric power of X over k, and let $\operatorname{Gr}^d \Gamma(X, \mathcal{L}_X^{\otimes 2})$ denote the Grassmann variety of linear subspaces of codimension d in the k-vector space $\Gamma(X, \mathcal{L}_X^{\otimes 2})$. Then we have a commutative diagram

$$\begin{tikzcd} \operatorname{Gr}^d \Gamma(X, \mathcal{L}_X^{\otimes 2}) \arrow[d, "\mathrm{Frob}_q"'] & \operatorname{Sym}^d X \arrow[l] \arrow[d, "\mathrm{Frob}_q"] \\ \operatorname{Gr}^d \Gamma(X, \mathcal{L}_X^{\otimes 2}) & \operatorname{Sym}^d X \arrow[l] \end{tikzcd}$$

of varieties over k, where the vertical arrows are the q-power Frobenius morphisms. Now let k' be a finite extension of k, write

$$X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k',$$

and let D be an effective divisor on X'. The commutativity of the above diagram shows that the divisor $(\mathrm{Frob}_q)_* D$ on X' can be computed using the following algorithm.

Algorithm 3.2 *(Frobenius map on divisors).* Let X be a projective curve over a finite field k of q elements, and let Frob_q be the Frobenius map on the set of divisors on X. Let k' be a finite extension of k. Let $X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k'$, and let $\mathcal{L}_{X'}$ be the pull-back of the line bundle $\mathcal{L}_X$ to X'. Let i be a positive integer, and let D be an effective divisor on X'. Given the matrix M of the inclusion map

$$\Gamma(X', \mathcal{L}_{X'}^{\otimes i}(-D)) \longrightarrow \Gamma(X', \mathcal{L}_{X'}^{\otimes i})$$

with respect to any k'-basis of the left-hand side and the k'-basis induced from any k-basis of $\Gamma(X, \mathcal{L}_X^{\otimes i})$ on the right-hand side, this algorithm outputs the analogous matrix for the inclusion map

$$\Gamma(X', \mathcal{L}_{X'}^{\otimes i}(-(\mathrm{Frob}_q)_* D)) \longrightarrow \Gamma(X', \mathcal{L}_{X'}^{\otimes 2}).$$

1. Apply the Frobenius automorphism of k' over k to the coefficients of the matrix M, and output the result.

Analysis. It follows from the discussion preceding the algorithm that the output is indeed equal to $\Gamma(X', \mathcal{L}_{X'}^{\otimes i}(-(\mathrm{Frob}_q)_* D))$. The algorithm involves $O((\deg \mathcal{L}_X)^2)$ computations of a q-th power of an element in k', which can be done in time polynomial in $\deg \mathcal{L}_X$, i and $\log \# k'$. ◇

3.2. Choosing random prime divisors

Let X be a projective curve (in the sense of § 2.1) over a finite field. Our next goal is to generate random effective divisors of given degree on X. We start with an algorithm to generate random prime divisors. For this we do not yet need to know the zeta function of X, although we use its properties in the analysis of the running time of the algorithm.

Algorithm 3.3 *(Random prime divisor).* Let X be a projective curve over a finite field k. Let d and i be positive integers such that

$$d \le i \deg \mathcal{L}_X - 2g_X.$$

Given d, i and the k-algebra $S_X^{(2i+2)}$, this algorithm outputs a uniformly distributed prime divisor P of degree d on X, represented as the subspace $\Gamma(\mathcal{L}_X^{\otimes i}(-P))$ of $\Gamma(\mathcal{L}_X^{\otimes i})$, provided $\operatorname{PDiv}^d X$ is non-empty. (If $\operatorname{PDiv}^d X = \emptyset$, the algorithm does not terminate.)

1. Choose a non-zero element $s \in \Gamma(\mathcal{L}_X^{\otimes i})$ uniformly randomly, and let D denote the divisor of s. (In other words, choose a random hypersurface section of degree i of X.)
2. Compute the set $\operatorname{Irr}^d D$ of (reduced) irreducible components of D of degree d over k using Algorithm 2.4.
3. With probability $\frac{\#\operatorname{Irr}^d D}{\lfloor (i \deg \mathcal{L}_X)/d \rfloor}$, output a uniformly random element $P \in \operatorname{Irr}^d D$ and stop.
4. Go to step 1.

Analysis. Let q denote the cardinality of k, and let H denote the set of divisors D that are divisors of non-zero global sections of $\mathcal{L}_X^{\otimes i}$. By the Riemann–Roch formula, the cardinality of H is

$$\#H = \frac{q^{1-g+i\deg \mathcal{L}} - 1}{q-1}.$$

When the algorithm finishes, the probability $p(D, P)$ that a specific pair (D, P) has been chosen is

$$\begin{aligned} p(D,P) &= \frac{1}{\#H} \frac{\#\operatorname{Irr}^d D}{\lfloor (i \deg \mathcal{L})/d \rfloor} \frac{1}{\#\operatorname{Irr}^d D} \\ &= \frac{q-1}{q^{1-g+i\deg\mathcal{L}} - 1} \frac{1}{\lfloor (i\deg\mathcal{L})/d \rfloor}. \end{aligned}$$

For all prime divisors P of degree d, the number of $D \in H$ for which P is in the support of D is equal to

$$\#\{D \mid P \in \operatorname{supp} D\} = \frac{q^{1-g+i\deg\mathcal{L}-d} - 1}{q-1},$$

so the probability $p(P)$ that a given P is chosen equals

$$\begin{aligned} p(P) &= \#\{D \mid P \in \operatorname{supp} D\} \cdot p(D,P) \\ &= \frac{q^{1-g+i\deg\mathcal{L}-d} - 1}{q^{1-g+i\deg\mathcal{L}} - 1} \frac{1}{\lfloor (i\deg\mathcal{L})/d \rfloor}. \end{aligned}$$

This is independent of P and therefore shows that when the algorithm finishes, the chosen element $P \in \operatorname{PDiv}^d X$ is uniformly distributed. Furthermore, the probability p that the algorithm finishes in a given iteration is

$$\begin{aligned} p &= \#\operatorname{PDiv}^d X \cdot \frac{q^{1-g+i\deg\mathcal{L}-d}-1}{q^{1-g+i\deg\mathcal{L}}-1}\frac{1}{\lfloor (i\deg\mathcal{L})/d\rfloor} \\ &= \frac{\#\operatorname{PDiv}^d X}{q^d}\frac{q^{1-g+i\deg\mathcal{L}}-q^d}{q^{1-g+i\deg\mathcal{L}}-1}\frac{1}{\lfloor (i\deg\mathcal{L})/d\rfloor} \\ &\geq \frac{\#\operatorname{PDiv}^d X}{q^d}(1-q^{-1-g_X})\frac{d}{i\deg\mathcal{L}}. \end{aligned}$$

We claim that the expected running time is polynomial in $\deg\mathcal{L}$, i and $\log q$, under the assumption that $\#\operatorname{PDiv}^d X \neq \emptyset$. We distinguish two cases:

$$q^{d/2} < 2\sigma_0(d)(2g_X+1) \quad \text{and} \quad q^{d/2} \geq 2\sigma_0(d)(2g_X+1).$$

Here $\sigma_0(d)$ denotes the number of positive divisors of d. In the first case, we see that

$$p > (2\sigma_0(d)(2g_X+1))^2(1-q^{-1-g_X})\frac{d}{i\deg\mathcal{L}},$$

which shows that $1/p$ is bounded by a polynomial in $\deg\mathcal{L}$ and i. In the second case, we deduce from (3.5) the following estimate for $\#\operatorname{PDiv}^d X$:

$$\begin{aligned} |d\#\operatorname{PDiv}^d X - q^d| &\leq \sum_{\substack{e\mid d\\ e\neq d}} q^e + \sum_{e\mid d}|s_e| \\ &\leq (\sigma_0(d)-1)q^{d/2} + \sigma_0(d)\cdot 2g_X q^{d/2}\,. \\ &< \sigma_0(d)(2g_X+1)q^{d/2} \\ &\leq \frac{1}{2}q^d \end{aligned}$$

This implies that $\#\operatorname{PDiv}^d X > q^d/(2d)$, and hence

$$p > \frac{1-q^{-1-g_X}}{2i\deg\mathcal{L}}.$$

In both cases we conclude that the expected running time is bounded by a polynomial in $\deg\mathcal{L}$, i and $\log q$. ◇

3.3. Choosing random divisors

As before, let X be a projective curve over a finite field k. From now on we assume that we know the zeta function of X, or equivalently the polynomial L_X.

Below we will give an algorithm for generating uniformly random effective divisors of a given degree on the curve X. These divisors will be built up from prime divisors, so it will be useful to speak of the *decomposition type* of an effective divisor D. This is the sequence of integers $(l_1, l_2, \ldots)$, where l_d is the number of prime divisors of degree d (counted with multiplicities) occurring in D.

One of the ingredients is the concept of *m-smooth* divisors and decomposition types. An m-smooth divisor is a linear combination of prime divisors whose degrees are at most m, and an m-smooth decomposition type of degree n is an m-tuple $(l_1, \ldots, l_m)$ such that $\sum_{d=1}^{m} l_d d = n$. For every m-smooth effective divisor D of degree n, we may view the decomposition type of D as an m-smooth decomposition type, since only its first m coefficients are non-zero.

The algorithm that we will describe takes as input integers $n \geq 0$ and $m \geq 1$, and outputs a uniformly random m-smooth effective divisor of degree n. Clearly, all effective divisors of degree n are n-smooth, so that the algorithm can be used with $m = n$ to produce uniformly random effective divisors of degree n.

The first step is to generate the decomposition type of a uniformly random m-smooth effective divisor of degree n. The method we use for doing this is described by Diem in [27, page 150] and in [28]. The algorithm works by recursion on m. For every $m \geq 1$, we write $\mathrm{Eff}^n_{\leq m} X$ for the set of m-smooth effective divisors D of degree n. Furthermore, for $l \geq 0$ and $m \geq 1$ we write $\mathrm{Eff}^{lm}_{=m} X$ for the set of divisors of degree lm that are linear combinations of prime divisors of degree m. We note that the set $\mathrm{Eff}^n_{\leq m} X$ can be decomposed as

$$\mathrm{Eff}^n_{\leq m} X = \begin{cases} \mathrm{Eff}^n_{=1} X & \text{if } m = 1; \\ \displaystyle\bigsqcup_{l=0}^{\lfloor n/m \rfloor} \mathrm{Eff}^{lm}_{=m} X \times \mathrm{Eff}^{n-lm}_{\leq m-1} X & \text{if } m \geq 2. \end{cases} \tag{3.7}$$

The cardinality of $\mathrm{Eff}^{lm}_{=m} X$ equals the number of ways to choose l elements from the set $\mathrm{PDiv}^m X$ with repeats. For this we have the well-known formula

$$\#\, \mathrm{Eff}^{lm}_{=m} X = \binom{\#\, \mathrm{PDiv}^m X - 1 + l}{l}. \tag{3.8}$$

Furthermore, from the description (3.7) of $\mathrm{Eff}^n_{\leq m} X$ we see that

$$\#\, \mathrm{Eff}^n_{\leq m} X = \begin{cases} \#\, \mathrm{Eff}^n_{=1} X & \text{if } m = 1; \\ \displaystyle\sum_{l=0}^{\lfloor n/m \rfloor} \#\, \mathrm{Eff}^{lm}_{=m} X \cdot \#\, \mathrm{Eff}^{n-lm}_{\leq m-1} X & \text{if } m \geq 2. \end{cases} \tag{3.9}$$

From these relations we can compute $\#\, \mathrm{Eff}^n_{\leq m} X$ recursively, starting from the numbers $\#\, \mathrm{PDiv}^d X$ for $1 \leq d \leq m$. An alternative way to describe these recurrence relations is to use generating functions; see Diem [27, page 149] or [28, Lemma 3.14].

In order to generate decomposition types of uniformly random m-smooth divisors of degree n, we define a probability distribution μ^n_m on the set of m-smooth decomposition types of degree n by defining $\mu^n_m(l_1, \ldots, l_m)$ as the probability that a uniformly randomly chosen effective m-smooth divisor of degree n has decomposition type $(l_1, \ldots, l_m)$. The algorithm now works as follows. We first select an integer $l_m \in \{0, 1, \ldots, \lfloor n/m \rfloor\}$—the number of prime divisors of degree m (counted with multiplicities) occurring in the decomposition—according to the marginal distribution ν^n_m of the m-th coordinate. We then apply the algorithm recursively with $(n - l_m m, m - 1)$ in place of (n, m).

The marginal distribution ν^n_m of the coordinate l_m in an m-tuple $(l_1, \ldots, l_m)$ distributed according to μ^n_m is the following. If $m = 1$, then $l_1 = n$ with probability 1. When $m \geq 2$, the probability that l_m equals a given $l \in \{0, 1, \ldots, \lfloor n/m \rfloor\}$ is

$$\nu^n_m(l) = \frac{\#\operatorname{Eff}^{lm}_{=m} X \cdot \#\operatorname{Eff}^{n-lm}_{\leq m-1} X}{\#\operatorname{Eff}^n_{\leq m} X} \quad (0 \leq l \leq \lfloor n/m \rfloor). \tag{3.10}$$

We compute $\#\operatorname{Eff}^n_{\leq m} X$, as well as $\#\operatorname{Eff}^{lm}_{=m}$ and $\#\operatorname{Eff}^{n-lm}_{\leq m-1} X$ for $0 \leq l \leq \lfloor n/m \rfloor$, using (3.5), (3.8) and (3.9). We then generate a random $l_m \in \{0, 1, \ldots, \lfloor n/m \rfloor\}$, distributed according to ν^n_m, in the following way. We subdivide the interval

$$I = \{0, 1, \ldots, \#\operatorname{Eff}^n_{\leq m} X - 1\}$$

into $\lfloor n/m \rfloor + 1$ intervals I_l, with $0 \leq l \leq \lfloor n/m \rfloor$ and each I_l having length $\#\operatorname{Eff}^{lm}_{=m} X \cdot \#\operatorname{Eff}^{n-lm}_{\leq m-1} X$, we generate a uniformly random element $x \in I$, and we select the unique l such that $x \in I_l$.

Algorithm 3.4 *(Decomposition type of a random divisor).* Given the polynomial L_X for a curve X over a finite field and integers $n \geq 0$ and $m \geq 1$, this algorithm outputs a random m-smooth decomposition type $(l_1, \ldots, l_m)$ of degree n, distributed according to the distribution μ^n_m.

1. If $m = 1$, output the 1-tuple (n) and stop.
2. Choose a random element $l_m \in \{0, 1, \ldots, \lfloor n/m \rfloor\}$ according to the distribution ν^n_m from (3.10).
3. Call the algorithm recursively with $(n - l_m m, m - 1)$ in place of (n, m) to obtain an $(m - 1)$-smooth decomposition type $(l_1, \ldots, l_{m-1})$ of degree $n - l_m m$.
4. Output the m-tuple $(l_1, \ldots, l_m)$.

Analysis. The correctness of the algorithm follows from the above discussion. It is straightforward to check that it runs in time polynomial in g_X, $\log \#k$, n and m. ◇

The preceding algorithm reduces our problem to generating random linear combinations of l prime divisors of a given degree d. In other words, we have to pick a random *multiset* of cardinality l from $\operatorname{PDiv}^d X$. This can be done using the following algoritm.

Algorithm 3.5 *(Random multiset).* Let S be a finite non-empty set of known cardinality. Suppose we have algorithms to pick uniformly random elements of S and to decide whether two such elements are equal. Given a non-negative integer l, this algorithm outputs a uniformly random multiset of l elements from S.

1. Generate a uniformly random subset $\{x_1, \ldots, x_l\}$ of $\{1, 2, \ldots, l + \#S - 1\}$, with $x_1 < x_2 < \ldots < x_l$.

2. Define a multiset $(y_1, \ldots, y_l)$ of l elements from $\{0, 1, \ldots, \#S - 1\}$ by $y_i = x_i - i$; then $y_1 \leq y_2 \leq \ldots \leq y_l$.
3. For each i with $1 \leq i \leq l$, let a_i be the number of elements of $\{0, 1, \ldots, \#S - 1\}$ that occur with multiplicity i in $(y_1, \ldots, y_l)$.
4. Generate a uniformly random sequence

$$\begin{gathered} s_1^1, s_2^1, \ldots, s_{a_1}^1, \\ s_1^2, s_2^2, \ldots, s_{a_2}^2, \\ \vdots \\ s_1^l, s_2^l, \ldots, s_{a_l}^l \end{gathered}$$

   of $a_1 + a_2 + \cdots + a_l$ distinct elements of S.
5. Output the multiset consisting of the elements s_i^j of S, where s_i^j occurs with multiplicity j.

Analysis. By construction, $(y_1, \ldots, y_l)$ is a uniformly random multiset of l elements from $\{0, 1, \ldots, \#S - 1\}$, so the "multiplicity vector" $(a_1, \ldots, a_l)$ is the same as that of a uniformly random multiset of l elements from S. The multiset generated in the last step is uniformly random among the multisets with this "multiplicity vector". This implies that the result is a uniformly random multiset of l elements from S, as required. ◇

Combining Algorithms 3.3, 3.4 and 3.5, we obtain the following algorithm to generate a uniformly random effective divisor of a given degree.

Algorithm 3.6 *(Random divisor).* Let X be a projective curve over a finite field k. Given positive integers m and i, an integer n satisfying

$$0 \leq n \leq i \deg \mathcal{L}_X - 2g_X,$$

the graded k-algebra $S_X^{(2i+2)}$ and the polynomial L_X, this algorithm outputs a uniformly random m-smooth effective divisor D of degree n on X, represented as the subspace $\Gamma(\mathcal{L}_X^{\otimes i}(-D))$ of $\Gamma(\mathcal{L}_X^{\otimes i})$.

1. Generate a random m-smooth decomposition type $(l_1, \ldots, l_m)$ of degree n using Algorithm 3.4.
2. For $d = 1, \ldots, m$, generate a uniformly random linear combination D_d of l_d prime divisors of degree d on X using Algorithm 3.5 (with $S = \operatorname{PDiv}^d X$, and $l = l_d$), where we use Algorithm 3.3 to generate random elements of $\operatorname{PDiv}^d X$.
3. Compute the subspace $\Gamma(\mathcal{L}_X(-D))$ for the divisor $D = D_1 + \cdots + D_m$ using the addition algorithm described in § 2.2, and output $\Gamma(\mathcal{L}_X(-D))$.

Analysis. It follows from the above discussion that the algorithm outputs a uniformly random m-smooth divisor of degree n on X. The running time is clearly polynomial in m, n, i and $\deg \mathcal{L}_X$ (measured in field operations in k). ◇

Remark. In practice, the following method for picking a random effective divisor of degree n is faster, but does not give a uniformly distributed output. We first choose a uniformly random non-zero section s of $\Gamma(X, \mathcal{L}^{\otimes i})$, where i is a non-negative integer such that

$$i \deg \mathcal{L} - n \geq 2g + 1.$$

Then if the set of effective divisors D of degree n with $D \leq \operatorname{div} s$ is non-empty, we pick a uniformly random element from it; otherwise we keep going with a different section s.

3.4. The Frobenius endomorphism of the Jacobian

As before, let k be a finite field of cardinality q, and let X be a proper, smooth and geometrically connected curve over k. Let J be the Jacobian variety of X, and let Frob_q denote the Frobenius endomorphism of J; it is an isogeny of degree q^g. The Rosati dual of Frob_q is called the *Verschiebung* and denoted by Ver_q. The Albanese and Picard maps associated to the Frobenius morphism on X are the endomorphisms Frob_q and Ver_q of J, respectively.

Assume we have a point $O \in X(k)$. Then we have a commutative diagram

$$\begin{array}{ccc} \mathrm{Sym}^d X & \longrightarrow & J \\ {\scriptstyle \mathrm{Frob}_q}\downarrow & & \downarrow{\scriptstyle \mathrm{Frob}_q} \\ \mathrm{Sym}^d X & \longrightarrow & J \end{array}$$

of varieties over k, where the horizontal maps send a divisor D to the class of $D - dO$ and the vertical arrows are the q-power Frobenius morphisms. This shows that the Frobenius endomorphism of J is equal to the endomorphism $\mathrm{Alb}(\mathrm{Frob}_q)$ induced by the Frobenius map on X via Albanese functoriality.

We write $X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k'$. The results of § 3.1 imply that for any finite extension k' of k, the endomorphism Frob_q of $J(k') = \mathrm{Pic}^0(X')$ can be computed by applying Algorithm 3.2 to any subspace $\Gamma(X', \mathcal{L}_{X'}^{\otimes 2}(-D))$ of the k'-vector space

$$\Gamma(X', \mathcal{L}_{X'}^{\otimes 2}) \cong k' \otimes_k \Gamma(X, \mathcal{L}_X^{\otimes 2})$$

with D an effective divisor of degree $\deg \mathcal{L}_X$ on X' such that $\mathcal{L}_{X'}(-D)$ represents x.

If O is a k-rational point of X, then we can compute the trace map

$$\mathrm{tr}_{k'/k} \colon \mathrm{Pic}^0 X' \to \mathrm{Pic}^0 X$$

in the following way. For $x \in \mathrm{Pic}^0 X'$, we compute a subspace of $\Gamma(X', \mathcal{L}_{X'}^{\otimes 2})$ representing the element

$$y = \sum_{i=0}^{[k':k]-1} \mathrm{Frob}_q^i x \in \mathrm{Pic}^0 X'.$$

Now y is in fact the image of the element $\mathrm{tr}_{k'/k} x \in \mathrm{Pic}^0 X$ under the inclusion $\mathrm{Pic}^0 X \to \mathrm{Pic}^0 X'$, so we can apply Algorithm 2.13 to find a subspace of $\Gamma(X, \mathcal{L}_X^{\otimes 2})$ representing $\mathrm{tr}_{k'/k} x$.

In § 2.11, the problem of computing the Albanese map for a finite morphism of curves was reduced to the problem of computing trace maps. Since we can solve the latter problem, we can therefore compute Albanese maps for finite morphisms of curves over finite fields.

3.5. Picking random elements of the Picard group

The next problem we will study is that of picking uniformly random elements in the finite Abelian group $J(k) = \operatorname{Pic}^0 X$. We recall from § 2.8 that in the medium model of the Picard group, the class of a line bundle $\mathcal{M}$ of degree 0 is represented by an effective divisor D of degree $\deg \mathcal{L}$ such that $\mathcal{M} \cong \mathcal{L}(-D)$. Consider the map

$$\begin{aligned} \operatorname{Eff}^{\deg \mathcal{L}} X &\to \operatorname{Pic}^0 X \\ D &\mapsto [\mathcal{L}(-D)]. \end{aligned}$$

It follows from the Riemann–Roch theorem and the fact that $\deg \mathcal{L} \ge 2g_X - 1$ that all fibres of this map have cardinality $\frac{q^{1-g+\deg \mathcal{L}}-1}{q-1}$. This means that to pick a uniformly random element of $\operatorname{Pic}^0 X$ it suffices to pick a uniformly random divisor of degree $\deg \mathcal{L}$. A method for doing this is given by Algorithm 3.6, provided that we know $S_X^{(6)}$.

3.6. Computing Frey–Rück pairings

Let n be a positive integer. We assume k contains a primitive n-th root of unity; this is equivalent to

$$n \mid \#k^\times = q - 1$$

and implies that n is not divisible by the characteristic of k.

Let X be a complete, smooth, geometrically connected curve over k, and let J be its Jacobian variety. The *Frey–Rück pairing* of order n on $J(k) = \operatorname{Pic}^0 X$ is the bilinear map

$$[\ ,\ ]_n \colon J[n](k) \times J(k)/nJ(k) \to \mu_n(k)$$

defined as follows (see Frey and Rück [39] or Schaefer [93]). Let x and y be elements of $J(k)$ such that $nx = 0$. Choose divisors D and E such that x and y are represented by the line bundles $\mathcal{O}_X(D)$ and $O_X(E)$, respectively, and such that the supports of D and E are disjoint. By assumption, there exists a rational function f on X such that $nD = \operatorname{div}(f)$; now $[x, y]_n$ is defined as

$$[x, y]_n = f(E)^{\#k^\times/n}.$$

Here $f(E)$ is defined on $\bar{k}$-valued points (where $\bar{k}$ is an algebraic closure of k) by function evaluation, and then extended to the group of divisors on $X_{\bar{k}}$, by linearity in the sense that

$$f(E + E') = f(E) \cdot f(E').$$

It is known that the Frey–Rück pairing is *perfect* in the sense that it induces isomorphisms

$$J[n](k) \xrightarrow{\sim} \operatorname{Hom}(J(k)/nJ(k), \mu_n(k))$$

and

$$J(k)/nJ(k) \xrightarrow{\sim} \mathrm{Hom}(J[n](k), \mu_n(k))$$

of Abelian groups.

Let us now give a slightly different interpretation of $f(E)$ that brings us in the right situation to compute $[x,y]_n$. We consider an arbitrary non-zero rational function f and an arbitrary divisor E such that the divisors

$$D = \mathrm{div}(f)$$

and E have disjoint supports. Since $f(E)$ is by definition linear in E, it suffices to consider the case where E is an effective divisor. As in § 2.7, we write

$$j_E \colon E \to X$$

for the closed immersion of E into X, and if $\mathcal{M}$ is a line bundle on X we abbreviate

$$\mathrm{N}_{E/k}\mathcal{M} = \mathrm{N}_{E/k}(j_E^*\mathcal{M}).$$

Since D and E have disjoint supports, we have a canonical trivialisation

$$t_D \colon k \cong \mathrm{N}_{E/k}\mathcal{O}_X \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{O}_X(D).$$

On the other hand, multiplication by f induces an isomorphism

$$\mathrm{N}_{E/k}f \colon \mathrm{N}_{E/k}\mathcal{O}_X(D) \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{O}_X \cong k.$$

of one-dimensional k-vector spaces. We claim that the composed isomorphism

$$k \xrightarrow[\sim]{t_D} \mathrm{N}_{E/k}\mathcal{O}_X(D) \xrightarrow[\sim]{\mathrm{N}_{E/k}f} k \tag{3.11}$$

is multiplication by $f(E)$. This is true in the case where E is a single point, since then $\mathrm{N}_{E/k}$ is (canonically isomorphic to) the identity functor. We deduce the general case from this by extending the base field to an algebraic closure of k and using the fact that both $f(E)$ and the norm functor are linear in E. For the latter claim, we refer to Deligne [100, exposé XVII, n° 6.3.27].

Remark. The isomorphism (3.11) could be taken as a *definition* of $f(E)$ for effective divisors E.

Lemma 3.7. *Let x and y be elements of $J(k)$ with $nx = 0$, let $\mathcal{M}$ be a line bundle representing x, and let E^+ and E^- be effective divisors such that $\mathcal{O}_X(E^+ - E^-)$ represents y. (In particular, $\mathcal{M}$ has degree 0, and E^+ and E^- have the same degree.) For any pair of trivialisations*

$$t^\pm \colon k \xrightarrow{\sim} \mathrm{N}_{E^\pm/k}\mathcal{M}$$

of k-vector spaces and any trivialisation

$$s \colon \mathcal{O}_X \xrightarrow{\sim} \mathcal{M}^{\otimes n}$$

of line bundles on X, the isomorphism

$$k \xrightarrow[\sim]{(t^+)^n} \mathrm{N}_{E^+/k}\mathcal{M}^{\otimes n} \xrightarrow[\sim]{\mathrm{N}_{E^+/k}s^{-1}} k \xrightarrow[\sim]{\mathrm{N}_{E^-/k}s} \mathrm{N}_{E^-/k}\mathcal{M}^{\otimes n} \xrightarrow[\sim]{(t^-)^{-n}} k$$

is multiplication by an element of $k^\times$ whose $(\#k^\times/n)$-th power equals $[x,y]_n$.

(We have implicitly used the isomorphisms $\mathrm{N}_{E^\pm/k}(\mathcal{M}^{\otimes n}) \cong (\mathrm{N}_{E^\pm/k}\mathcal{M})^{\otimes n}$ expressing the linearity of $\mathrm{N}_{E/k}$, and denoted both sides of the isomorphism by $\mathrm{N}_{E^\pm/k}\mathcal{M}^{\otimes n}$.)

Proof. We fix a non-zero rational section h such that the divisor

$$D = \operatorname{div} h$$

is disjoint from $E^\pm$. Then we have canonical trivialisations

$$t_D^\pm\colon k \xrightarrow{\sim} \mathrm{N}_{E^\pm/k}\mathcal{O}_X(D)$$

as above. Composing these with the isomorphism

$$\mathrm{N}_{E^\pm/k}h\colon \mathrm{N}_{E^\pm/k}\mathcal{O}_X(D) \xrightarrow{\sim} \mathrm{N}_{E^\pm/k}\mathcal{M}$$

induced by multiplication by h gives trivialisations

$$t_h^\pm = \mathrm{N}_{E^\pm/k}h \circ t_D\colon k \xrightarrow{\sim} \mathrm{N}_{E^\pm/k}\mathcal{M}.$$

Now consider any isomorphism

$$s\colon \mathcal{O}_X \xrightarrow{\sim} \mathcal{M}^{\otimes n}$$

of line bundles on X, and define

$$f = s^{-1} \circ h^n\colon \mathcal{O}_X(nD) \xrightarrow{\sim} \mathcal{O}_X;$$

then f can be viewed as a rational function with divisor nD. We now have commutative diagrams

$$\begin{tikzcd}
k \arrow[r, "(t_D^\pm)^n", "\sim"'] \arrow[d, equal] & \mathrm{N}_{E^\pm/k}\mathcal{O}_X(nD) \arrow[r, "\mathrm{N}_{E^\pm/k}f", "\sim"'] \arrow[d, "\mathrm{N}_{E^\pm/k}h^n", "\sim"'] & k \arrow[d, equal] \\
k \arrow[r, "(t_h^\pm)^n", "\sim"'] & \mathrm{N}_{E^\pm/k}\mathcal{M}^{\otimes n} \arrow[r, "\mathrm{N}_{E^\pm/k}s^{-1}", "\sim"'] & k.
\end{tikzcd}$$

As we saw above, the top row is multiplication by $f(E^\pm)$; by the commutativity of the diagram, the same holds for the bottom row. Finally, we note that replacing $t_h^\pm$ by *any* pair of trivialisations

$$t^\pm\colon k \xrightarrow{\sim} \mathrm{N}_{E^\pm/k}\mathcal{M}$$

changes the isomorphism in the bottom row of the above diagram by some n-th power in $k^\times$. This implies that the isomorphism

$$k \xrightarrow[\sim]{(t^\pm)^n} \mathrm{N}_{E^\pm/k}\mathcal{M}^{\otimes n} \xrightarrow[\sim]{\mathrm{N}_{E^\pm/k}s^{-1}} k$$

equals multiplication by an element of $k^\times$ whose $(\#k/n)$-th power is $f(E^\pm)^{\#k^\times/n}$. The lemma follows from this by the definition of $[x,y]_n$. $\square$

Lemma 3.7 reduces the problem of computing the Frey–Rück pairing of order n to the following: given a line bundle $\mathcal{M}$ such that $\mathcal{M}^{\otimes n}$ is trivial, find an isomorphism

$$s\colon \mathcal{O}_X \xrightarrow{\sim} \mathcal{M}^n,$$

and, given moreover an effective divisor E and a trivialisation

$$t\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{M},$$

compute the isomorphism

$$I^E_{s,t}\colon k \underset{\sim}{\xrightarrow{t^n}} \mathrm{N}_{E/k}\mathcal{M}^{\otimes n} \underset{\sim}{\xrightarrow{\mathrm{N}_{E/k}s^{-1}}} k. \tag{3.12}$$

We assume that the curve X is specified by a projective embedding via a line bundle $\mathcal{L}$ as in § 2.1. We will describe an algorithm to compute isomorphisms of the type $I^E_{s,t}$, based on Khuri-Makdisi's algorithms for computing with divisors on X. Suppose we are given a line bundle $\mathcal{M}$ of degree 0 such that $\mathcal{M}^{\otimes n}$ is trivial and an effective divisor E. For simplicity, we assume that $\deg E = \deg \mathcal{L}$. As in § 2.2, we represent the class of $\mathcal{M}$ in $J(k)$ by the subspace $\Gamma(X, \mathcal{L}^{\otimes 2}(-D))$ of $\Gamma(X, \mathcal{L}^{\otimes 2})$, where D is any effective divisor of degree $\deg \mathcal{L}$ (not necessarily disjoint from E) such that

$$\mathcal{M} \cong \mathcal{L}(-D).$$

Likewise, we represent E as the subspace $\Gamma(X, \mathcal{L}^{\otimes 2}(-E))$ of $\Gamma(X, \mathcal{L}^{\otimes 2})$.

First, we will describe a construction of a trivialisation

$$s\colon \mathcal{O}_X \xrightarrow{\sim} \mathcal{L}(-D)^{\otimes n}.$$

For this we fix an anti-addition chain $(a_0, a_1, \ldots, a_m)$ for n, as described in § 2.8. In particular, for each l with $2 \le l \le m$ we are given $i(l)$ and $j(l)$ in $\{0, 1, \ldots, l-1\}$ such that

$$a_l = -a_{i(l)} - a_{j(l)}.$$

We fix any non-zero global section u of $\mathcal{L}$, and we put

$$D_0 = \operatorname{div}(u), \quad D_1 = D.$$

For $l = 2, 3, \ldots, m$, we iteratively apply Algorithm 2.11 to $D_{i(l)}$ and $D_{j(l)}$; this gives an effective divisor D_l of degree $\deg \mathcal{L}$ and a global section s_l of $\mathcal{L}^{\otimes 3}$ such that the line bundle $\mathcal{L}^{\otimes 3}(-D_l - D_{i(l)} - D_{j(l)})$ is trivial and

$$\operatorname{div}(s_l) = D_l + D_{i(l)} + D_{j(l)}.$$

We recursively define rational sections $h_1, h_2, \ldots, h_m$ of $\mathcal{L}^{\otimes(a_l - 1)}$ by

$$h_l = \begin{cases} u^{-1} & \text{for } l = 0; \\ 1 & \text{for } l = 1; \\ (h_{i(l)} h_{j(l)} s_l)^{-1} & \text{for } l = 2, 3, \ldots, m. \end{cases}$$

Then it follows immediately that each h_l has divisor $a_l D - D_l$. In particular, since $\mathcal{L}(-D)^{\otimes n}$ is trivial, so is $\mathcal{L}(-D_m)$ and Algorithm 2.10 provides us with a global section v of $\mathcal{L}$ such that

$$\operatorname{div}(v) = D_m.$$

The rational section

$$s = h_m v$$

of $\mathcal{L}^{\otimes n}$ has divisor nD and hence induces an isomorphism

$$s\colon \mathcal{O}_X \xrightarrow{\sim} \mathcal{L}(-D)^{\otimes n}.$$

Next, we assume that an effective divisor E has been given. We assume for simplicity that $\deg E = \deg \mathcal{L}$. We fix bases of the following k-vector spaces:

$$\begin{gathered}\Gamma(E, \mathcal{L}^{\otimes 2});\\ \Gamma(E, \mathcal{L}^{\otimes 3}(-D_l)) \text{ for } 1 \le l \le m;\\ \Gamma(E, \mathcal{L}^{\otimes 4}(-D_{i(l)} - D_{j(l)})) \text{ for } 2 \le l \le m.\end{gathered}$$

In addition, we fix a k-basis of $\Gamma(E, \mathcal{L}^{\otimes 3}(-D_0))$ by defining it as the image of the chosen basis of $\Gamma(E, \mathcal{L}^{\otimes 2})$ under the multiplication map

$$u\colon \Gamma(E, \mathcal{L}^{\otimes 2}) \xrightarrow{\sim} \Gamma(E, \mathcal{L}^{\otimes 3}(-D_0)).$$

For $0 \le l \le m$ we define a trivialisation

$$\begin{aligned} t_l\colon k &\xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}(-D_l)\\ &\xrightarrow{\sim} \operatorname{Hom}_k\bigl(\det_k \Gamma(E, \mathcal{L}^{\otimes 2}), \det_k \Gamma(E, \mathcal{L}^{\otimes 3}(-D_l))\bigr)\end{aligned}$$

using the given bases of $\Gamma(E, \mathcal{L}^{\otimes 2})$ and $\Gamma(E, \mathcal{L}^{\otimes 3}(-D_l))$, and we define an element γ_l of $k^\times$ by requiring that the diagram

$$\begin{array}{ccc} k & \overset{t_l}{\underset{\sim}{\longrightarrow}} & \mathrm{N}_{E/k}\mathcal{L}(-D_l)\\ {\scriptstyle\gamma_l}\big\downarrow{\scriptstyle\sim} & & {\scriptstyle\sim}\big\downarrow{\scriptstyle h_l}\\ k & \overset{t^{a_l}}{\underset{\sim}{\longrightarrow}} & \mathrm{N}_{E/k}\mathcal{L}(-D)^{\otimes a_l}\end{array}$$

be commutative. For $2 \le l \le m$, we define a trivialisation

$$t'_l\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}^{\otimes 2}(-D_{i(l)} - D_{j(l)})$$

by (2.9) using the given bases of $\Gamma(E, \mathcal{L}^{\otimes 4}(-D_{i(l)} - D_{j(l)}))$ and $\Gamma(E, \mathcal{L}^{\otimes 2})$. Furthermore, we endow $\Gamma(E, \mathcal{L}^{\otimes 5}(-D_l - D_{i(l)} - D_{j(l)}))$ with the basis obtained by transferring the given basis of $\Gamma(E, \mathcal{L}^{\otimes 2})$ via multiplication by s_l. This gives a trivialisation

$$t''_l\colon k \xrightarrow{\sim} \mathrm{N}_{E/k}\mathcal{L}^{\otimes 3}(-D_l - D_{i(l)} - D_{j(l)})$$

by (2.9) using the basis of $\Gamma(E, \mathcal{L}^{\otimes 5}(-D_l - D_{i(l)} - D_{j(l)}))$ just defined and the given basis of $\Gamma(E, \mathcal{L}^{\otimes 2})$.

Algorithm 3.8 *(Compute isomorphisms of the form $I^E_{s,t}$).* Let X be a projective curve over a field k, let D and E be effective divisors of degree $\deg\mathcal{L}$ on X, and let n be a positive integer such that $\mathcal{L}(-D)^{\otimes n}$ is trivial. Given the k-algebra $S_X^{(7)}$, an anti-addition chain $(a_0, a_1, \ldots, a_m)$ for n, a global section u of $\mathcal{L}$, effective divisors D_0, D_1, …, D_m, global sections s_2, …, s_m of $\mathcal{L}^3$ such that

$$D_0 = \operatorname{div}(u), D_1 = D \quad \text{and} \quad \operatorname{div}(s_l) = D_l + D_{i(l)} + D_{j(l)} \text{ for } 2 \le l \le m$$

and a global section v of the trivial line bundle $\mathcal{L}(-D_m)$, this algorithm outputs the isomorphism $I^E_{s,t}$ defined by (3.12), where s is defined using the given data, and where t is chosen by the algorithm. (Note that this means that the output of the algorithm is an element of $k^\times$ defined up to n-th powers in $k^\times$.)

1. Put $\gamma_0 = \gamma_1 = 1$.

2. For $l = 2, 3, \ldots, m$:

3. Using Algorithm 2.9, compute the elements $\lambda_l^{(1)}$ and $\lambda_l^{(2)}$ of $k^\times$ such that the diagrams

$$\begin{tikzcd} k \arrow[r, "t_{i(l)}\otimes t_{j(l)}", "\sim"'] \arrow[d, "\lambda_l^{(1)}"', "\sim"] & \mathrm{N}_{E/k}\mathcal{L}(-D_{i(l)}) \otimes \mathrm{N}_{E/k}\mathcal{L}(-D_{j(l)}) \arrow[d, "\sim"] \\ k \arrow[r, "t'_l", "\sim"'] & \mathrm{N}_{E/k}\mathcal{L}^{\otimes 2}(-D_{i(l)} - D_{j(l)}) \end{tikzcd}$$

and

$$\begin{tikzcd} k \arrow[r, "t_l\otimes t'_l", "\sim"'] \arrow[d, "\lambda_l^{(2)}"', "\sim"] & \mathrm{N}_{E/k}\mathcal{L}(-D_l) \otimes \mathrm{N}_{E/k}\mathcal{L}^{\otimes 2}(-D_{i(l)} - D_{j(l)}) \arrow[d, "\sim"] \\ k \arrow[r, "t''_l", "\sim"'] & \mathrm{N}_{E/k}\mathcal{L}^{\otimes 3}(-D_l - D_{i(l)} - D_{j(l)}) \end{tikzcd}$$

are commutative. Define $\lambda_l = \lambda_l^{(1)}\lambda_l^{(2)}$.

4. Put $\gamma_l = \dfrac{\lambda_l}{\gamma_{i(l)}\gamma_{j(l)}}$.

5. Compute $\delta \in k^\times$ as the determinant of the matrix of the isomorphism

$$v\colon \Gamma(E, \mathcal{L}^2) \xrightarrow{\sim} \Gamma(E, \mathcal{L}^3(-D_m))$$

with respect to the given bases.

6. Output the element $\dfrac{1}{\gamma_m\delta} \in k^\times$.

Analysis. The definition of λ_l given in the algorithm implies that the diagram

$$\begin{tikzcd} k \arrow[r, "t_l\otimes t_{i(l)}\otimes t_{j(l)}", "\sim"'] \arrow[d, "\lambda_l"', "\sim"] & \mathrm{N}_{E/k}\mathcal{L}(-D_l) \otimes \mathrm{N}_{E/k}\mathcal{L}(-D_{i(l)}) \otimes \mathrm{N}_{E/k}\mathcal{L}(-D_{j(l)}) \arrow[d, "\sim"] \\ k \arrow[r, "t''_l", "\sim"'] & \mathrm{N}_{E/k}\mathcal{L}^{\otimes 3}(-D_l - D_{i(l)} - D_{j(l)}) \end{tikzcd}$$

is commutative; furthermore, t''_l is induced by multiplication by s_l. The recursive definition of the h_l implies that the recurrence relation between the γ_l is as stated in the algorithm. Namely, it follows from the definition of D_0, from the special choice of basis of $\Gamma(E, \mathcal{L}^{\otimes 3}(-D_0))$ and from the fact that $t_1 = t$ that

$$\gamma_0 = \gamma_1 = 1.$$

Furthermore, the definitions of h_l, γ_l, $\gamma_{i(l)}$, $\gamma_{j(l)}$ and the property of λ_l that we have just proved imply that

$$\gamma_l = \frac{\lambda_l}{\gamma_{i(l)}\gamma_{j(l)}} \quad \text{for } l = 2, 3, \ldots, m.$$

Finally, it follows from the definitions of s, γ_m and the isomorphism $I^E_{s,t}$ from (3.12) that the relation between v, t_m, γ_m and $I^E_{s,t}$ is given by the commutativity of the diagram

$$\begin{tikzcd} k \arrow[r, "I^E_{s,t}", "\sim"'] & k \arrow[d, "\mathrm{N}_{E/k} v", "\sim"'] \\ k \arrow[u, "\gamma_m", "\sim"'] \arrow[r, "t_m", "\sim"'] & \mathrm{N}_{E/k}\mathcal{L}(-D_m). \end{tikzcd}$$

This proves that the element of $k^\times$ output by the last step is indeed $I^E_{s,t}$.

It is straightforward to check that the running time of the algorithm, measured in operations in k, is polynomial in $\deg \mathcal{L}$ and m. ◇

Algorithm 3.9 *(Frey–Rück pairing).* Let X be a projective curve over a finite field k, let n be an integer dividing $\#k^\times$, and let x and y be elements of $J(k)$ with $nx = 0$. Given the k-algebra $S^{(7)}_X$ and subspaces $\Gamma(\mathcal{L}^{\otimes 2}_X(-D))$ and $\Gamma(\mathcal{L}^{\otimes 2}_X(-E^-))$ of $\Gamma(\mathcal{L}^{\otimes 2}_X)$ representing x and y, this algorithm outputs the element $[x, y]_n \in \mu_n(k)$.

1. Find an anti-addition chain $(a_0, a_1, \ldots, a_m)$ for n.
2. Choose any non-zero global section u of $\mathcal{L}_X$, and let D_0 denote its divisor. Compute the space
$$\Gamma(\mathcal{L}^{\otimes 2}_X(-D_0)) = u\Gamma(\mathcal{L}_X).$$
Write $D_1 = D$.
3. Using Algorithm 2.11, find effective divisors D_2, D_3, …, D_m of degree $\deg \mathcal{L}_X$, where each D_l is represented as the space $\Gamma(\mathcal{L}^{\otimes 2}_X(-D_l))$, and non-zero global sections s_2, s_3, …, s_m of $\mathcal{L}^{\otimes 3}_X$ such that the line bundle $\mathcal{L}^{\otimes 3}_X(-D_{i(l)} - D_{j(l)} - D_l)$ is trivial and
$$\operatorname{div}(s_l) = D_{i(l)} + D_{j(l)} + D_l.$$
4. Using Algorithm 2.10, verify that $\mathcal{L}_X(-D_m)$ is trivial and find a non-zero global section v of $\mathcal{L}_X(-D_m)$.

5. Choose a non-zero global section w of $\mathcal{L}_X$, let E^+ denote its divisor, and compute

$$\Gamma(\mathcal{L}_X^{\otimes 2}(-E^+)) = w\Gamma(\mathcal{L}_X).$$

6. Compute $I_{s,t^+}^{E^+}$ and $I_{s,t^-}^{E^-}$, viewed as elements of $k^\times$, using Algorithm 3.8, where t^+ and t^- are certain trivialisations chosen by that algorithm.

7. Output $(I_{s,t^+}^{E^+}/I_{s,t^-}^{E^-})^{\#k^\times/n}$.

Analysis. The correctness of this algorithm follows from Lemma 3.7. The running time is polynomial in $\deg \mathcal{L}_X$, $\log \#k$ and $\log n$. ◇

3.7. Finding relations between torsion points

Let X be a projective curve over a finite field k, represented as in § 2.1, let J be its Jacobian, and let l be a prime number different from the characteristic of k. We will show how to find all the $\mathbf{F}_l$-linear relations between given elements of $J[l](k)$. In particular, given a basis $(b_1, \ldots, b_n)$ for a subspace V of $J[l](k)$ and another point $x \in J[l](k)$, this allows us to check whether $x \in V$, and if so, express x as a linear combination of $(b_1, \ldots, b_n)$.

Let k' be an extension of k containing a primitive l-th root of unity. It is well known that the problem just described can be reduced, via the Frey–Rück pairing, to the discrete logarithm problem in the group $\mu_l(k')$. Algorithm 3.11 below makes this precise. We begin with a bound on the number of elements needed to generate a finite-dimensional vector space over a finite field with high probability.

Lemma 3.10. *Let $\mathbf{F}$ be a finite field, and let V be an $\mathbf{F}$-vector space of finite dimension d. Let α be a real number with $0 < \alpha < 1$, and write*

$$m = \begin{cases} 0 & \text{if } d = 0; \\ d - 1 + \left\lceil \dfrac{\log \frac{1}{1-\alpha^{1/d}}}{\log \#\mathbf{F}} \right\rceil & \text{if } d > 0. \end{cases}$$

If $v_1, \ldots, v_m$ are uniformly random elements of V, the probability that V is generated by $v_1, \ldots, v_m$ is at least α.

Proof. Fix a basis of V. The matrix of the linear map

$$\begin{aligned} \mathbf{F}^m &\longrightarrow V \\ (c_1, \ldots, c_m) &\mapsto \sum_{i=1}^m c_i v_i \end{aligned}$$

is a uniformly random $d \times m$-matrix over $\mathbf{F}$. The probability that it has rank d is the probability that its rows (which are uniformly random elements of $\mathbf{F}^m$) are linearly independent. This occurs with probability

$$\begin{aligned} p &= \frac{(\#\mathbf{F}^m - 1)(\#\mathbf{F}^m - \#\mathbf{F}) \cdots (\#\mathbf{F}^m - \#\mathbf{F}^{d-1})}{\#\mathbf{F}^{dm}} \\ &\ge \frac{(\#\mathbf{F}^m - \#\mathbf{F}^{d-1})^d}{\#\mathbf{F}^{dm}} \\ &= \bigl(1 - (\#\mathbf{F})^{-(m-d+1)}\bigr)^d \end{aligned}$$

The choice of m implies that $p \ge \alpha$. □

Remark. The integer m defined in Lemma 3.10 is approximately $d-1+\frac{\log d}{\log \#F}$, in the sense that for any fixed α the difference is bounded for $d \ge 1$.

Algorithm 3.11 *(Relations between torsion points).* Let X be a projective curve over a finite field k, let J be its Jacobian, and let l be a prime number different from the characteristic of k. Let $x_1, \ldots, x_n$ be elements of $J[l](k)$. Given the k-algebra $S_X^{(7)}$ and subspaces $\Gamma(\mathcal{L}_X^{\otimes 2}(-D_i))$ of $\Gamma(\mathcal{L}_X^{\otimes 2})$ representing x_i for $1 \le i \le n$, this algorithm outputs an $\mathbf{F}_l$-basis for the kernel of the natural map

$$\begin{aligned}\Sigma\colon \mathbf{F}_l^n &\longrightarrow J[l](k)\\ (c_1, \ldots, c_n) &\longmapsto \sum_{i=1}^n c_i x_i.\end{aligned}$$

The algorithm depends on a parameter $\alpha \in (0,1)$.

1. Generate a minimal extension k' of k such that k' contains a primitive l-th root of unity ζ. Let
$$\lambda\colon \mu_l(k') \xrightarrow{\sim} \mathbf{F}_l$$
denote the corresponding discrete logarithm, i.e. the unique isomorphism of one-dimensional $\mathbf{F}_l$-vector spaces sending ζ to 1.
2. Define an integer $m \ge 0$ by
$$m = \begin{cases} 0 & \text{if } n = 0;\\ n-1+\left\lceil \dfrac{\log \frac{1}{1-\alpha^{1/n}}}{\log l} \right\rceil & \text{if } n > 0.\end{cases}$$
3. Choose m uniformly random elements $y_1, \ldots, y_m$ in $J(k')$ as described in § 3.5; their images in $J(k')/lJ(k')$ are again uniformly distributed.
4. Compute the $m \times n$-matrix
$$M = (\lambda([y_i, x_j]_l)) \quad (1 \le i \le m,\ 1 \le j \le n)$$
with coefficients in $\mu_l(k')$, where the pairing $[\ ,\ ]_l$ is evaluated using Algorithm 3.9 and the isomorphism λ is evaluated using some algorithm for computing discrete logarithms in $\mu_l(k)$.
5. Compute an $\mathbf{F}_l$-basis $(b_1, \ldots, b_r)$ for the kernel of M.
6. If $\Sigma(b_1) = \ldots = \Sigma(b_r) = 0$, output $(b_1, \ldots, b_r)$ and stop.
7. Go to step 3.

Analysis. We write V for the image of Σ and V' for the quotient of $J(k')/lJ(k')$ by the annihilator of V under the pairing $[\ ,\ ]_l$. Then we have an induced isomorphism

$$V \xrightarrow{\sim} \mathrm{Hom}_{\mathbf{F}_l}(V', \mu_l(k')).$$

Consider the map

$$\begin{aligned} \Sigma' \colon \mathbf{F}_l^m &\longrightarrow V' \\ (c_1, \dots, c_m) &\longmapsto \sum_{i=1}^m c_i y_i. \end{aligned}$$

Now we have a commutative diagram

$$\begin{array}{ccc} \mathbf{F}_l^n & \longrightarrow & \mathrm{Hom}_{\mathbf{F}_l}(\mathbf{F}_l^m, \mu_l(k')) \\ {\scriptstyle\Sigma}\big\downarrow & & \big\uparrow{\scriptstyle f\mapsto f\circ\Sigma'} \\ V & \xrightarrow{\sim} & \mathrm{Hom}_{\mathbf{F}_l}(V', \mu_l(k')) \end{array}$$

We identify $\mu_l(k')$ with $\mathbf{F}_l$ using the isomorphism λ and equip $\mathrm{Hom}_{\mathbf{F}_l}(\mathbf{F}_l^m, \mu_l(k'))$ with the dual basis of the standard basis of $\mathbf{F}_l^m$. Then the top arrow in the diagram is given by the matrix M defined in step 4. This means that we have an inclusion

$$\ker \Sigma \subseteq \ker M.$$

In step 6 we check whether this inclusion is an equality. The surjectivity of Σ implies that this is the case if and only if the rightmost map in the diagram is injective, i.e. if and only if Σ' is surjective. Since $\dim_{\mathbf{F}_l} V \le n$, this happens with probability at least α by Lemma 3.10. Therefore steps 3–7 are executed at most $1/\alpha$ times on average. This implies that (for fixed α) the algorithm runs in time polynomial in g_X, $\log \#k$, l and n. $\diamond$

Remarks. (1) If we know an upper bound for the dimension of the $\mathbf{F}_l$-vector space generated by the x_i, then we can use this upper bound instead of n in the expression for m in step 2.

(2) It does not matter much what algorithm we use for computing the discrete logarithm in $\mu_l(k')$, since the running time of Algorithm 3.11 is already polynomial in l. For example, we can simply tabulate the function λ.

3.8. The Kummer map on a divisible group

Let k be a finite field of cardinality q, and let l be a prime number. Let $\mathbf{G}$ be an étale l-divisible group over k. (The étaleness is automatic if l is different from the characteristic of k.) We denote by $\mathrm{Frob}_q \colon \mathbf{G} \to \mathbf{G}$ the (q-power) Frobenius endomorphism of $\mathbf{G}$; this is an automorphism because of the assumption that $\mathbf{G}$ is étale.

For any non-negative integer n such that all the points of $\mathbf{G}[l^n]$ are k-rational, the *Kummer map* of order l^n on $\mathbf{G}$ over k is the isomorphism

$$\begin{aligned} K_{l^n}^{\mathbf{G}/k} \colon \mathbf{G}(k)/l^n\mathbf{G}(k) &\xrightarrow{\sim} \mathbf{G}[l^n](k) \\ x &\longmapsto \mathrm{Frob}_q(y) - y, \end{aligned}$$

where y is any point of $\mathbf{G}$ over an algebraic closure of k such that the image of $l^n y$ in $\mathbf{G}(k)/l^n\mathbf{G}(k)$ equals x.

Let $\chi \in \mathbf{Z}_l[t]$ be the characteristic polynomial of the Frobenius automorphism of $\mathbf{G}$ on the Tate module of $\mathbf{G}$. Then the element $t \bmod \chi$ of $\mathbf{Z}_l[t]/(\chi)$ is invertible. Let n be any non-negative integer, and let a be a positive integer such that

$$t^a = 1 \quad \text{in } (\mathbf{Z}_l[t]/(l^n, \chi))^\times.$$

Then $t^a - 1$ is divisible by l^n in $\mathbf{Z}_l[t]/(\chi)$, and we let h_a be the unique element of $\mathbf{Z}_l[t]/(\chi)$ such that

$$t^a - 1 = l^n h_a \in \mathbf{Z}_l[t]/(\chi).$$

By the Cayley–Hamilton theorem, $\mathbf{Z}_l[t]/(\chi)$ acts on $\mathbf{G}$ with t acting as Frob_q. The above identity therefore implies that

$$\mathrm{Frob}_q^a - 1 = l^n h_a(\mathrm{Frob}_q) \quad \text{on } \mathbf{G}.$$

Let k_a be an extension of k with

$$[k_a : k] = a.$$

Then $\mathbf{G}[l^n]$ is defined over k_a, and we can express the Kummer map over k_a in terms of the Frobenius endomorphism over k as

$$\begin{aligned} K_{l^n}^{\mathbf{G}/k_a} \colon \mathbf{G}(k_a)/l^n\mathbf{G}(k_a) &\xrightarrow{\sim} \mathbf{G}[l^n](k_a) \\ x &\longmapsto h_a(\mathrm{Frob}_q)(x). \end{aligned}$$

In § 3.9 we are going to apply this to a certain l-divisible subgroup of the l-power torsion of the Jacobian of a projective curve over k.

3.9. Computing the l-torsion in the Picard group

Let X be a projective curve over k, represented as in § 2.1, and let J be its Jacobian. Let Frob_q denote the Frobenius endomorphism of J over k, and let $\chi \in \mathbf{Z}[t]$ be the characteristic polynomial of Frob_q.

Let l be a prime number different from the characteristic of k. We are going to apply the results of § 3.8 to a certain l-divisible subgroup $\mathbf{G}$ of the group $J[l^\infty]$ of l-power torsion points of J. This $\mathbf{G}$ is defined as follows. Let $\bar{f} = (t-1)^b$ be the largest power of $t - 1$ dividing $\chi \bmod l$, so that $\chi \bmod l$ has the factorisation

$$(\chi \bmod l) = \bar{f} \cdot \bar{f}^\perp$$

in coprime monic polynomials in $\mathbf{F}_l[t]$. Hensel's lemma implies that this factorisation can be lifted uniquely to a factorisation

$$\chi = f \cdot f^\perp,$$

where f and $f^\perp$ are coprime monic polynomials in $\mathbf{Z}_l[t]$. The Chinese remainder theorem gives a decomposition

$$\mathbf{Z}_l[t]/(\chi) \xrightarrow{\sim} \mathbf{Z}_l[t]/(f) \times \mathbf{Z}_l[t]/(f^\perp), \tag{3.13}$$

which in turn induces a decomposition

$$J[l^\infty] \cong \mathbf{G} \times \mathbf{G}^\perp$$

of l-divisible groups. We note that $\mathbf{G}$ is of rank b and that f is the characteristic polynomial of Frob_q on $\mathbf{G}$. Let a be a positive integer such that

$$t^a = 1 \quad \text{in } (\mathbf{F}_l[t]/\bar{f})^\times, \tag{3.14}$$

let h_a be the unique element of $\mathbf{Z}_l[t]/(f)$ such that

$$t^a - 1 = lh_a \in \mathbf{Z}_l[t]/(f), \tag{3.15}$$

and let k_a be an extension of degree a of k. All the points of $\mathbf{G}[l]$ are k_a-rational, and the b-dimensional $\mathbf{F}_l$-vector space $\mathbf{G}[l](k_a)$ is the generalised eigenspace corresponding to the eigenvalue 1 of Frob_q inside the $\mathbf{F}_l$-vector space of points of $J[l]$ over an algebraic closure of k_a. In particular, we have the identity

$$J[l](k) = \{x \in \mathbf{G}[l](k_a) \mid \mathrm{Frob}_q(x) = x\}.$$

As explained in § 3.8, the map

$$\begin{aligned} \mathbf{G}(k_a)/l\mathbf{G}(k_a) &\xrightarrow{\sim} \mathbf{G}[l](k_a) \\ x &\longmapsto h_a(\mathrm{Frob}_q)(x) \end{aligned}$$

is well-defined and equal to the Kummer isomorphism

$$K_l^{\mathbf{G}/k_a} \colon \mathbf{G}(k_a)/l\mathbf{G}(k_a) \xrightarrow{\sim} \mathbf{G}[l](k_a)$$

of order l.

We use the above results to generate uniformly random elements of the $\mathbf{F}_l$-vector space $\mathbf{G}[l](k_a)$. We factor $\#J(k_a)$ as

$$\#J(k_a) = l^{c_a} m_a$$

with $c_a \geq 0$, $m_a \geq 1$ and $l \nmid m_a$. Let e be the idempotent in $\mathbf{Z}_l[t]/(\chi)$ corresponding to the element $(1, 0)$ on the right-hand side of (3.13). Composing the maps

$$J(k_a) \xrightarrow{m_a} J[l^\infty](k_a) \xrightarrow{e(\mathrm{Frob}_q)} \mathbf{G}(k_a) \longrightarrow \mathbf{G}(k_a)/l\mathbf{G}(k_a) \xrightarrow{h_a(\mathrm{Frob}_q)} \mathbf{G}[l](k_a) \tag{3.16}$$

we get a surjective group homomorphism from $J(k_a)$ to $\mathbf{G}[l](k_a)$. We can use this map to convert uniformly random elements of $J(k_a)$ into uniformly random elements of $\mathbf{G}[l](k_a)$, provided we know e and h_a to sufficient l-adic precision. It is clear that to compute the Kummer map we only need to know the image of h_a in $\mathbf{Z}_l[t]/(f, l) = \mathbf{F}_l[t]/((t-1)^b)$. Since $\mathbf{G}(k_a)$ can be identified with a subgroup of $\#J(k_a)$, it is annihilated by l^{c_a}, and we have

$$J[l^\infty](k_a) = J[l^{c_a}](k_a) \quad \text{and} \quad \mathbf{G}(k_a) = \mathbf{G}[l^{c_a}](k_a).$$

This implies that it suffices to know e to precision $O(l^{c_a})$.

Let us check that there is a reasonably small a for which (3.14) holds. For any non-negative integer γ the identity

$$t^{l^\gamma} - 1 = (t-1)^{l^\gamma}$$

holds in $\mathbf{F}_l[t]$, and the right-hand side maps to zero in $\mathbf{F}_l[t]/(t-1)^b$ if and only if $l^\gamma \ge b$. Since l is a prime number, we conclude that the order of t in $\mathbf{F}_l[t]/((t-1)^b)$ equals l^γ, where γ is the least non-negative integer such that $l^\gamma \ge b$.

Algorithm 3.12 *(Computing the l-torsion of the Picard group).* Let X be a projective curve over a finite field k with q elements, let J be its Jacobian, and let l be a prime number different from the characteristic of k. Given the k-algebra $S_X^{(7)}$ and the characteristic polynomial χ of the Frobenius endomorphism of J over k, this algorithm outputs an $\mathbf{F}_l$-basis for $J[l](k) = (\operatorname{Pic} X)[l]$. The algorithm depends on a parameter $\alpha \in (0,1)$.

1. Factor $\chi \bmod l$ in $\mathbf{F}_l[t]$ as

$$(\chi \bmod l) = \bar f \cdot \bar f^\perp,$$

where $\bar f$ is the greatest power of $t-1$ dividing $\chi \bmod l$, say $\bar f = (t-1)^b$, and lift this to a factorisation

$$\chi = f \cdot f^\perp$$

in coprime monic polynomials in $\mathbf{Z}_l[t]$.

2. Compute the non-negative integer r defined by

$$r = \begin{cases} 0 & \text{if } b = 0; \\ b - 1 + \left\lceil \dfrac{\log \frac{1}{1-\alpha^{1/b}}}{\log l} \right\rceil & \text{if } b \ge 1. \end{cases}$$

3. Define $a = l^\gamma$, where γ is the least non-negative integer such that $l^\gamma \ge b$. Generate a finite extension k_a of degree a of k. Factor $\#J(k_a)$ as

$$\#J(k_a) = l^{c_a} m_a \quad \text{with } l \nmid m_a.$$

Compute the image of the idempotent e in $(\mathbf{Z}/l^{c_a}\mathbf{Z})[t]/(\chi)$ using the extended Euclidean algorithm, and compute the image of h_a in $\mathbf{F}_l[t]/((t-1)^b)$ using the definition (3.15) of h_a.

4. Generate r uniformly random elements of $J(k_a)$ as explained in § 3.5, and map them to elements $x_1, \ldots, x_r \in \mathbf{G}[l](k_a)$ using the homomorphism (3.16).

5. Using Algorithm 3.7, compute a basis for the kernel of the $\mathbf{F}_l$-linear map

$$\begin{aligned} \Sigma\colon \mathbf{F}_l^r &\longrightarrow \mathbf{G}[l](k_a) \\ (c_1, \ldots, c_r) &\longmapsto \sum_{i=1}^r c_i x_i. \end{aligned}$$

If the dimension of this kernel is greater than $r - b$, go to step 4.

6. Use the $\mathbf{F}_l$-linear relations between $x_1, \ldots, x_r$ computed in the previous step to find a subsequence $(y_1, \ldots, y_b)$ of $(x_1, \ldots, x_r)$ that is an $\mathbf{F}_l$-basis of $\mathbf{G}[l](k_a)$.
7. Let M be the matrix with respect to the basis $(y_1, \ldots, y_b)$ of the $\mathbf{F}_l$-linear automorphism of $\mathbf{G}[l](k_a)$ induced by the Frobenius endomorphism Frob_q of J over k. Compute M by computing $\mathrm{Frob}_q(y_i)$ for $i = 1, \ldots, b$ using Algorithm 3.2 and then applying Algorithm 3.7 to express the $\mathrm{Frob}_q(y_i)$ as linear combinations of the y_i.
8. Compute a basis for the kernel of $M - I$, where I is the $b \times b$ identity matrix. Map the basis elements to elements $z_1, \ldots, z_t$ of $\mathbf{G}[l](k_a)$ using the injective homomorphism
$$\begin{aligned} \mathbf{F}_l^b &\longrightarrow \mathbf{G}[l](k_a) \\ (a_1, \ldots, a_b) &\longmapsto \sum_{i=1}^b a_i y_i. \end{aligned}$$
Output $(z_1, \ldots, z_t)$.

Analysis. As we remarked earlier, the definition of a implies that a equals the order of t in $(\mathbf{F}_l[t]/(t-1)^b)^\times$; furthermore, $J[l](k)$ equals the kernel of $\mathrm{Frob}_q - \mathrm{id}$ on $\mathbf{G}[l](k_a)$. The elements $x_1, \ldots, x_r$ of $\mathbf{G}[l](k_a)$ are uniformly random by the fact that (3.16) is a homomorphism. By Lemma 3.10, they generate the b-dimensional $\mathbf{F}_l$-vector space $\mathbf{G}[l](k_a)$ with probability at least α. The definition of a also implies that
$$a \le \max\{1, 2g_X l - 1\},$$
while the "class number formula" (3.6) gives the upper bound
$$\begin{aligned} c_a &\le \frac{\log \# J(k_a)}{\log l} \\ &\le \frac{2g_X \log\bigl(1 + q^{a/2}\bigr)}{\log l}. \end{aligned}$$
This shows that c_a is bounded by a polynomial in g_X, $\log q$ and l. For fixed α we therefore reach step 6 in expected polynomial time in $\deg \mathcal{L}_X$, $\log q$ and l. In steps 6–8 we compute a basis for the kernel of $\mathrm{Frob}_q - \mathrm{id}$, which is $J[l](k)$. We conclude that the algorithm is correct and runs in probabilistic polynomial time in $\deg \mathcal{L}_X$, $\log q$ and l. ◇

Remark. The elements $z_j \in J[l](k_a)$ output by the preceding algorithm are in fact defined over k. In general, I do not know how to generate k-vector spaces (instead of k_a-vector spaces) representing them. However, if we know a k-rational point on X, then we can use Algorithm 2.13 to accomplish this.

4. Modular symbols

In this section we collect some results on the problem of computing the Hecke algebra $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$, defined in § I.2.2 for given positive integers $n \geq 1$ and $k \geq 2$. We also give applications to finding an Atkin–Lehner basis of $\mathrm{S}_k(\Gamma_1(n),\mathbf{C})$, to computing zeta functions of modular curves over finite fields and to finding cusp forms of weight k for $\Gamma_1(n)$ consisting of forms with integral q-expansion and small Petersson norm.

4.1. Computing Hecke algebras

The Hecke algebra can be computed by means of the technique of *modular symbols*, developed by Manin [73], Shokurov [103], Merel [79], Cremona [18] and others. We refer to Stein [104, Chapter 8] for more details. The results can be phrased as follows. Given integers $n \geq 1$ and $k \geq 2$, one can compute the Hecke algebra $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$ in the form of the multiplication table with respect to some $\mathbf{Z}$-basis $(t_1,\ldots,t_N)$ of $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n)),\mathbf{Z})$. This computation can be done in time polynomial in n and k. Furthermore, given a positive integer m one can compute the matrix of the Hecke operator T_m with respect to the basis $(t_1,\ldots,t_N)$ in time polynomial in n, k and m, and one can compute the diamond operators $\langle d\rangle$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ on this basis in time polynomial in n and k. Similar results as above hold with $\Gamma_1(n)$ replaced by $\Gamma_1(n;p)$, where p is a prime number possibly dividing n.

Since the exact result we need does not seem to have been published, let us sketch how to compute $\mathbf{T}(\mathrm{S}_k(\Gamma,\mathbf{Z}))$, where Γ is either $\Gamma_1(n)$ or $\Gamma_1(n;p)$. There is a certain $\mathbf{Q}$-vector space $\mathbf{S}_k(\Gamma,\mathbf{Q})^+$, the plus one quotient of the $\mathbf{Q}$-vector space of modular symbols of weight k for Γ, which is canonically isomorphic to $\mathrm{Hom}(\mathrm{S}_k(\Gamma,\mathbf{Q}),\mathbf{Q})$; see Stein [104, § 9.3]. Let m be the least positive integer that is larger than the degree of the line bundle of cusp forms of weight k on the modular curve in question. We compute the matrices of the Hecke operators $T_1, \ldots, T_m$ with respect to some $\mathbf{Q}$-basis of $\mathbf{S}_k(\Gamma,\mathbf{Q})^+$ in time polynomial in k and m. We then use the LLL lattice basis reduction algorithm (see for example Lenstra, Lenstra and Lovász [66] or Lenstra [68]) to find a $\mathbf{Z}$-basis for the Hecke algebra

$$\mathbf{T}(\mathbf{S}_k(\Gamma,\mathbf{Q})^+) \cong \mathbf{T}(\mathrm{S}_k(\Gamma,\mathbf{Z}))$$

and the corresponding multiplication table. Since we can compute the matrix of any Hecke operator t on $\mathbf{S}_k(\Gamma,\mathbf{Q})^+$, we can also express t on the $\mathbf{Z}$-basis of $\mathbf{T}(\mathrm{S}_k(\Gamma,\mathbf{Z}))$ found by the LLL algorithm.

Let $\mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n))$ be the $\mathbf{Z}$-module of cusp forms whose q-expansions at the cusp 0 have coefficients in $\mathbf{Z}$. We can compute $\mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n))$ itself by using that it is isomorphic (as a $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$-module) to the $\mathbf{Z}$-linear dual of $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$ via the perfect pairing

$$\begin{aligned} \mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z})) \times \mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n)) &\longrightarrow \mathbf{Z} \\ (t,f) &\longmapsto a_1(t^\vee f) \end{aligned} \tag{4.1}$$

from § I.2.4.

The *new quotient* of $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$ is the quotient of $\mathbf{T}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$ by the annihilator of $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$; we denote it by $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n),\mathbf{Z}))$. We can compute this quotient as follows. Using standard methods from linear algebra, we compute a basis of primitive forms of weight k for each $\Gamma_1(d)$, with $d \mid n$. This gives us the *Atkin–Lehner basis* of $\mathrm{S}_k(\Gamma_1(n))$. We then compute the annihilator of $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{C})$, and from this we compute $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{Z}))$. The whole computation takes time polynomial in n and k.

We define $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{Z}))$-modules

$$\mathrm{S}_k^{\mathrm{int,new}}(\Gamma_1(n)) = \mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n)) \cap \mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{C})$$

and

$$\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{Q}) = \mathrm{S}_k(\Gamma_1(n), \mathbf{Q}) \cap \mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{C}).$$

These can be computed from $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{Z}))$ in the same way as described above.

4.2. Computing the zeta function of a modular curve

Let $n \geq 5$ be an integer and p a prime number not dividing n. Applying the results of § 4.1 with $k = 2$ and using the isomorphism

$$\mathbf{T}_1(n) \xrightarrow{\sim} \mathbf{T}(\mathrm{S}_2(\Gamma_1(n), \mathbf{Z}))$$

from § I.2.3, we see that we can compute the Hecke algebra $\mathbf{T}_1(n)$ in time polynomial in n. It is well known that from the elements T_p and $\langle p \rangle$ of $\mathbf{T}_1(n)$ one can compute the characteristic polynomial χ of the Frobenius operator Frob_p on the l-adic Tate module

$$\mathrm{T}_l \mathrm{J}_1(n)_{\mathbf{F}_p} = \varprojlim_r \mathrm{J}_1(n)_{\mathbf{F}_p}[l^r](\overline{\mathbf{F}}_p),$$

where l is any prime number different from p. From the polynomial χ we get the zeta function of $\mathrm{X}_1(n)_{\mathbf{F}_p}$ using the formula

$$\mathrm{Z}_{\mathrm{X}_1(n)_{\mathbf{F}_p}} = \frac{t^{\deg \chi} \chi(1/t)}{(1-t)(1-pt)}.$$

Let us describe in some more detail how to compute χ. We know from § I.1.3 that $\mathbf{Q}_l \otimes_{\mathbf{Z}_l} \mathrm{T}_l \mathrm{J}_1(n)_{\mathbf{F}_p}$ is a free $\mathbf{Q}_l \otimes_{\mathbf{Z}} \mathbf{T}_1(n)$-module of rank 2 and that the characteristic polynomial of Frob_p on it equals $x^2 - T_p x + p\langle p \rangle \in \mathbf{T}_1(n)[x]$. This implies that the characteristic polynomial of Frob_p viewed as a $\mathbf{Q}_l$-linear map is

$$\chi = \mathrm{N}_{\mathbf{T}_1(n)[x]/\mathbf{Z}[x]}(x^2 - T_p x + p\langle p \rangle) \in \mathbf{Z}[x].$$

To compute the right-hand side, we apply the following standard algorithm for computing norms. We choose a $\mathbf{Z}$-basis of $\mathbf{T}_1(n)$; this can also be interpreted as a $\mathbf{Z}[x]$-basis of $\mathbf{T}_1(n)[x]$. We write M_{T_p} and $M_{\langle p \rangle}$ for the matrices of T_p and $\langle p \rangle$ with respect to the chosen basis. Then we compute χ as the determinant of the matrix $x^2 \cdot \mathrm{id} - x \cdot M_{T_p} + p \cdot M_{\langle p \rangle}$ with coefficients in $\mathbf{Z}[x]$.

4.3. Finding a basis of cusp forms with small Petersson norm

Let n and k be integers with $n \ge 1$ and $k \ge 2$. Let $\langle\ ,\ \rangle_{\Gamma_1(n)}$ denote the Petersson inner product on $\mathrm{S}_k(\Gamma_1(n), \mathbf{C})$, as defined in § II.2.1. We will explain how to find a $\mathbf{Q}$-basis of $\mathrm{S}_k(\Gamma_1(n), \mathbf{Q})$ consisting of forms with integral q-expansions at the cusp 0 and with "small" Petersson norms, using the results of § II.2.4 and § 4.1.

Let $\mathbf{T}'$ be the subring of $\mathrm{End}_{\mathbf{C}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{C}))$ generated by $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{Z}))$ and the $T_p^\vee$ for $p \mid n$. Then $\mathbf{T}'$ is a commutative free $\mathbf{Z}$-algebra of finite rank, containing $\mathbf{T}^{\mathrm{new}}(\mathrm{S}_k(\Gamma_1(n), \mathbf{Z}))$ as a subring of finite index. On $\mathbf{T}'$ there is an involution $t \mapsto t^\vee$ sending each Hecke operator to its dual; see § I.2.2. We equip $\mathbf{T}'$ with the modified trace pairing

$$\begin{aligned}\langle\ ,\ \rangle_{\mathbf{T}'} \colon \mathbf{T}' \times \mathbf{T}' &\to \mathbf{Z}\\ (t, u) &\mapsto \mathrm{tr}_{\mathbf{T}'/\mathbf{Z}}(t u^\vee).\end{aligned}$$

We put

$$S' = \{f \in \mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{C}) \mid a_1(tf) \in \mathbf{Z} \text{ for all } t \in \mathbf{T}'\}.$$

The pairing (4.1) induces a perfect pairing

$$\begin{aligned}\mathbf{T}' \times S' &\to \mathbf{Z}\\ (t, f) &\mapsto a_1(t^\vee f)\end{aligned}$$

and hence an isomorphism

$$\mathbf{T}' \xrightarrow{\sim} \mathrm{Hom}(S', \mathbf{Z}) \tag{4.2}$$

of $\mathbf{T}'$-modules, where the action of an element $t \in \mathbf{T}'$ on $\mathrm{Hom}(S', \mathbf{Z})$ is induced from the action of $t^\vee$ on S'. We extend the base field to $\mathbf{C}$ in (4.2) and decompose the right-hand side into simple $(\mathbf{T}' \otimes \mathbf{C})$-modules corresponding to primitive forms. This gives an isomorphism

$$\begin{aligned}\mathbf{T}' \otimes \mathbf{C} &\xrightarrow{\sim} \bigoplus_{f \in P_k(\Gamma_1(n))} \mathbf{C}\\ t &\longmapsto \bigl(e_f(t^\vee)\bigr)_f,\end{aligned} \tag{4.3}$$

where $P_k(\Gamma_1(n))$ is the set of primitive cusp forms of weight k for $\Gamma_1(n)$ and $e_f(t)$ denotes the eigenvalue of t on f. For every $t \in \mathbf{T}'$, the adjoint of t with respect to the inner product $\langle\ ,\ \rangle_{\Gamma_1(n)}$ equals $t^\vee$. This implies that

$$e_f(t^\vee) = \overline{e_f(t)} \quad \text{for all } f \in P_k(\Gamma_1(n)) \text{ and all } t \in \mathbf{T}'.$$

If R is a ring and M a free $\mathbf{R}$-module of finite rank, we write $\mathrm{tr}_R(t \mid M)$ for the trace of an endomorphism t of M. Then (4.2) implies

$$\begin{aligned}\langle t, u\rangle_{\mathbf{T}'} &= \mathrm{tr}_{\mathbf{Z}}(t u^\vee \mid \mathbf{T}')\\ &= \mathrm{tr}_{\mathbf{Z}}(t u^\vee \mid \mathrm{Hom}(S', \mathbf{Z}))\\ &= \mathrm{tr}_{\mathbf{Z}}(t^\vee u \mid S')\\ &= \mathrm{tr}_{\mathbf{C}}(t^\vee u \mid \mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n), \mathbf{C}))\\ &= \sum_{f \in P_k(\Gamma_1(n))} e_f(t^\vee u).\end{aligned}$$

This implies that under the isomorphism (4.3), the unique sesquilinear form on $\mathbf{T}'\otimes\mathbf{C}$ extending $\langle\ ,\ \rangle_{\mathbf{T}'}$ corresponds to the standard Hermitean inner product on $\bigoplus_f \mathbf{C}$. We conclude that the bilinear form $\langle\ ,\ \rangle_{\mathbf{T}'}$ on $\mathbf{T}'$ is symmetric and positive definite, and that the dual inner product

$$\langle\ ,\ \rangle_{S'}\colon S'\times S'\to\mathbf{R}$$

induced by (4.2) has the property that the set of primitive forms is orthonormal for the extension of $\langle\ ,\ \rangle_{S'}$ to a Hermitean inner product on $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$.

Algorithm 4.1 *(Find a small basis of* $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{Q})$*).* Given integers $n\ge 1$ and $k\ge 2$ as well as a real number $c>4/3$, this algorithm outputs a $\mathbf{Q}$-basis $(g_1,\ldots,g_N)$ of $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{Q})$ such that for each i we have $g_i\in\mathrm{S}_k^{\mathrm{int,new}}(\Gamma_1(n))$ and

$$\langle g_i,g_i\rangle_{\Gamma_1(n)}\le c^{N(N-1)/2}(A_{k,\epsilon}n^\epsilon\operatorname{vol}_{\Gamma_1(n)})^N\bigl(4\pi(D+1)\exp(4\pi(D+1))\bigr)^{N-1}$$

for any $\epsilon>0$, where $A_{k,\epsilon}>0$ is defined in Lemma II.2.1, $N=\dim_{\mathbf{C}}\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(n),\mathbf{C})$ and D is the degree of the line bundle $\omega^{\otimes k}(-\text{cusps})$ on the stack $\mathcal{M}_{\Gamma_1(n)}$.

1. Using modular symbols, compute a $\mathbf{Z}$-basis $(t_1,\ldots,t_N)$ of the ring $\mathbf{T}'$ defined above. We denote by $(f_1,\ldots,f_N)$ the dual basis of S'.
2. Compute the matrix M of the inner product $\langle\ ,\ \rangle_{\mathbf{T}'}$ with respect to $(t_1,\ldots,t_N)$.
3. Compute M^{-1}; this is the matrix of $\langle\ ,\ \rangle_{S'}$ with respect to $(f_1,\ldots,f_N)$.
4. Using the LLL algorithm, compute a basis of S' that is c-reduced with respect to $\langle\ ,\ \rangle_{S'}$.

Analysis. It is clear that for fixed c, the running time of the algorithm is polynomial in n and k. It remains to prove the upper bound on $\langle g_i,g_i\rangle_{\Gamma_1(n)}$. We note that

$$\det\bigl(\langle g_i,g_j\rangle_{S'}\bigr)_{i,j=1}^N=\det(M^{-1})\in\{1/m\mid m=1,2,3,\ldots\}.$$

The c-reducedness of the basis $(g_1,\ldots,g_N)$ implies

$$\begin{aligned}\prod_{i=1}^N\langle g_i,g_i\rangle_{S'}&\le c^{N(N-1)/2}\det\bigl(\langle g_i,g_j\rangle_{S'}\bigr)_{i,j=1}^N\\&\le c^{N(N-1)/2}.\end{aligned}$$

For $i=1,\ldots,N$, we now write g_i as a $\mathbf{C}$-linear combination

$$g_i=\sum_{f\in P_k(\Gamma_1(n))}\alpha_i^f f.$$

The upper bound on Petersson norms of primitive forms proved in Lemma II.2.1 implies that for every $\epsilon>0$ there is an explicitly computable real number $A_{k,\epsilon}$ such that

$$\begin{aligned}\langle g_i,g_i\rangle_{\Gamma_1(n)}&=\sum_{f\in P_k(\Gamma_1(n))}|\alpha_i|^2\langle f,f\rangle_{\Gamma_1(n)}\\&\le A_{k,\epsilon}n^\epsilon\operatorname{vol}_{\Gamma_1(n)}\sum_{f\in P_k(\Gamma_1(n))}|\alpha_i^f|^2\\&=A_{k,\epsilon}n^\epsilon\operatorname{vol}_{\Gamma_1(n)}\langle g_i,g_i\rangle_{S'}.\end{aligned}$$

This implies

$$\prod_{i=1}^{N} \langle g_i, g_i \rangle_{\Gamma_1(n)} \le c^{N(N-1)/2} (A_{k,\epsilon} n^\epsilon \operatorname{vol}_{\Gamma_1(n)})^N.$$

For each $j \neq i$, we bound $\langle g_j, g_j \rangle_{\Gamma_1(n)}$ from below as in Lemma II.2.2. This implies the claimed bound on $\langle g_i, g_i \rangle_{\Gamma_1(n)}$. ◇

Algorithm 4.2 *(Find a small basis of* $\mathrm{S}_k(\Gamma_1(n), \mathbf{Q})$*).* Given integers $n \ge 1$ and $k \ge 2$ as well as a real number $c > 4/3$, this algorithm outputs a $\mathbf{Q}$-basis $(h_1, \ldots, h_N)$ of $\mathrm{S}_k(\Gamma_1(n), \mathbf{Q})$ such that for all i we have $h_i \in \mathrm{S}_k^{\mathrm{int}}(\Gamma_1(n))$ and

$$\langle h_i, h_i \rangle_{\Gamma_1(n)} \le c^{N(N-1)/2} (A_{k,\epsilon} n^\epsilon \operatorname{vol}_{\Gamma_1(n)})^N \bigl(4\pi(D+1)\exp(4\pi(D+1))\bigr)^{N-1}$$

for any $\epsilon > 0$, where $A_{k,\epsilon} > 0$ is defined in Lemma II.2.1, $N = \dim_{\mathbf{C}} \mathrm{S}_k(\Gamma_1(n), \mathbf{C})$ and D is the degree of the line bundle $\omega^{\otimes k}(-\text{cusps})$ on $\mathcal{M}_{\Gamma_1(n)}$.

1. Using Algorithm 4.1, compute a $\mathbf{Q}$-basis B_d of $\mathrm{S}_k^{\mathrm{new}}(\Gamma_1(d), \mathbf{Q})$ for each divisor d of n.
2. Output the basis $B = \bigsqcup_{d|n} \bigsqcup_{e|n/d} (b_e^{n,d})^* B_d$ of $\mathrm{S}_k(\Gamma_1(n), \mathbf{Q})$.

Analysis. It is clear that for fixed c, the running time of the algorithm is polynomial in n and k. The claimed bounds on the Petersson norms of the b_i follow from those in Algorithm 4.1 together with the equality

$$\langle (b_e^{n,d})^* g, (b_e^{n,d})^* g \rangle_{\Gamma_1(n)} = \frac{\langle g, g \rangle_{\Gamma_1(n)}}{(ee')^{k/2}} \quad \text{for all } g \in \mathrm{S}_k(\Gamma_1(n), \mathbf{C}),$$

which follows from the definition of the Petersson inner product. ◇

5. Computing with vector space schemes and Galois representations

In this section we explain how to find the Galois representation attached to a finite-dimensional $\mathbf{F}$-vector space scheme over $\mathbf{Q}$, where $\mathbf{F}$ is a finite field. We also describe how to find the minimal non-trivial subrepresentations of such a Galois representation. Finally, we give an algorithm to compute the Frobenius conjugacy classes at prime numbers at which such a representation is unramified.

To solve the first two problems we need to be able to factor polynomials over number fields efficiently. There exist deterministic algorithms that accomplish this. For these we refer to Lenstra, Lenstra and Lovász [66], Lenstra [65], van Hoeij [109], Belabas [6], and Belabas et al. [7].

5.1. Computing Galois groups

We start by describing a well-known algorithm for computing the Galois group of a finite Galois extension of a number field; see for example Lenstra [67, Theorem 3.2].

Suppose we are given a Galois extension $K \subseteq L$ of number fields. By the primitive element theorem, we can choose an isomorphism

$$K[x]/(f) \xrightarrow{\sim} L,$$

over K, where $f \in K[x]$ is some monic irreducible polynomial. Because $K \subseteq L$ is a Galois extension, L is the splitting field of f over K. We compute all the roots of f in L by factoring f, and we fix one root α (say $x \bmod f$). Then the map

$$\begin{aligned} \mathrm{Gal}(L/K) &\to \{\text{roots of } f \text{ in } L\} \\ \sigma &\mapsto \sigma(\alpha) \end{aligned}$$

is a bijection. Since all the roots can be expressed as polynomials in α, we can, for each root β of f, compute the corresponding element of $\mathrm{Gal}(L/K)$ as a group of permutations of the roots of f. In other words, if $[L:K]=n$, we can give $\mathrm{Gal}(L/K)$ as a subgroup of order n in the symmetric group S_n.

Remark. Suppose the extension $K \subseteq L$ is given by the multiplication table of L with respect to some K-basis rather than by the minimal polynomial of a primitive element. It is well known that in this situation one can find a primitive element as a small linear combination of the given basis elements, because all elements of L that do not generate L over K lie in the union of the finitely many strict subfields of L containing K.

5.2. Representing Galois representations

Let $\mathbf{F}$ be a finite field, and let

$$\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F})$$

be a two-dimensional representation. Let K_ρ denote the finite Galois extension of $\mathbf{Q}$ such that ρ factors as

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K_\rho/\mathbf{Q}) \rightarrowtail \mathrm{GL}_2(\mathbf{F}).$$

For algorithmic purposes, the representation ρ can be described using the following data:

(1) the characteristic p of $\mathbf{F}$;

(2) the multiplication table of $\mathbf{F}$ with respect to some $\mathbf{F}_p$-basis of $\mathbf{F}$;

(3) the multiplication table of K_ρ with respect to some $\mathbf{Q}$-basis of K_ρ;

(4) the list of pairs $(M_\sigma, \rho(\sigma))$, where σ runs over $\mathrm{Gal}(K_\rho/\mathbf{Q})$ and M_σ is the matrix of σ with respect to the given $\mathbf{Q}$-basis of K_ρ.

5.3. Representing vector space schemes

Let k be a field, let $\mathbf{F}$ be a finite field, and let V be a finite $\mathbf{F}$-vector space scheme over k. We suppose given a closed immersion

$$\iota\colon V \to \mathbf{A}^1_k$$

over k, giving an $\mathbf{F}$-vector space scheme structure on the image of ι. This structure is given by the following data:

(1) the monic polynomial $P \in k[x]$ defining the image of ι;

(2) an element $S \in k[x_1, x_2]/(P(x_1), P(x_2))$ such that

$$P(S) = 0 \text{ in } k[x_1, x_2]/(P(x_1), P(x_2))$$

and such that the addition morphism

$$+\colon V \times_{\operatorname{Spec} k} V \to V$$

corresponds via ι to the k-algebra homomorphism

$$\begin{aligned} k[x]/(P) &\to k[x_1, x_2]/(P(x_1), P(x_2)) \\ x &\mapsto S; \end{aligned}$$

(3) for all $a \in \mathbf{F}$ an element $M_a \in k[x]/(P)$ with $P(M_a) = 0$ in $k[x]/(P)$ and such that the multiplication morphism

$$a\cdot\colon V \to V$$

corresponds via ι to the k-algebra homomorphism

$$\begin{aligned} k[x]/(P) &\to k[x]/(P) \\ x &\mapsto M_a. \end{aligned}$$

Let q be the "coordinate" of the zero section of V, i.e. the element $q \in k$ such that the ideal $(x - q)$ of $k[x]/(P)$ corresponds to the trivial subgroup scheme $\{0\}$ of V. This q can be extracted from P and M_0; namely, it is the unique root of P in k such that the map $x \mapsto M_0$ factors as

$$k[x]/(P) \to k \to k[x]/(P),$$

where the first map sends x to q.

Remark. It is not the case that an embedding ι as above exists for all k and V. For example, if V is the constant vector space scheme $\mathbf{F}$ and k is finite with $\#k < \#\mathbf{F}$, there is no such ι.

5.4. Finding minimal components of a vector space scheme

Let V be a finite $\mathbf{F}$-vector space scheme over $\mathbf{Q}$, represented as in § 5.3. We will give an algorithm to find the minimal non-trivial $\mathbf{F}$-vector space schemes contained in V. These correspond to the minimal elements (with respect to division) in the set of monic polynomials R with

$$(x-q) \mid R \mid P, \quad R \neq x-q$$

that have the property that the maps giving the $\mathbf{F}$-vector space scheme structure on V induce maps $\mathbf{Q}[x]/(R) \to \mathbf{Q}[x]/(R)$ and $\mathbf{Q}[x]/(R) \to \mathbf{Q}[x_1, x_2]/(R(x_1), R(x_2))$. The first step in the algorithm is to factor the polynomial P over $\mathbf{Q}$. As remarked before, there are (deterministic) algorithms for doing this that run in polynomial time in the degree of P and the largest among the heights of its coefficients.

Algorithm 5.1 *(Finding minimal components of a vector space scheme).* Let $\mathbf{F}$ be a finite field, and let V be a finite $\mathbf{F}$-vector space scheme over $\mathbf{Q}$. Given polynomials P, S, and M_a for $a \in \mathbf{F}$ describing V as in § 5.3, this algorithm outputs the polynomials R defining the minimal non-trivial $\mathbf{F}$-vector space schemes contained in V.

1. Factor P as a product

$$P = P_0 P_1 \ldots P_n,$$

where $P_0, \ldots, P_n$ are distinct monic irreducible elements of $\mathbf{Q}[x]$ and such that $P_0 = x - q$ with q as above.

2. Choose a generator a of the cyclic group $\mathbf{F}^\times$.

3. Put $T = \emptyset$.

4. For $i = 1, \ldots, n$:

5. Put $R = P_0 P_i$.

6. Replace R by the monic generator of the kernel of the ring homomorphism

$$\begin{aligned} \mathbf{Q}[x] &\to \mathbf{Q}[x_1, x_2]/(R(x_1), R(x_2)) \\ x &\mapsto S. \end{aligned}$$

Then replace R by the monic generator of the kernel of the ring homomorphism

$$\begin{aligned} \mathbf{Q}[x] &\to \mathbf{Q}[x]/(R) \\ x &\mapsto M_a. \end{aligned}$$

Repeat this step until R does not change anymore.

7. Remove all $R' \in T$ such that R strictly divides R', and add R to T.

8. Output the set T.

Analysis. It follows from the construction of T that its elements are, as required, the minimal elements of the set of polynomials R as above. Furthermore, R remains a divisor of P, so step 6 is executed at most $\deg P$ times. This shows that the algorithm runs in time polynomial in $\deg P = \#\mathbf{F}^{\dim_{\mathbf{F}} V}$ and in the largest among the heights of the coefficients of P. ◇

5.5. Computing Galois representations attached to vector space schemes

Let $\mathbf{F}$ be a finite field, and let V be a finite $\mathbf{F}$-vector space scheme over $\mathbf{Q}$. There is an associated Galois representation

$$\rho_V \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}} V(\overline{\mathbf{Q}}).$$

Let K_V denote the finite Galois extension of $\mathbf{Q}$ such that ρ_V factors as

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K_V/\mathbf{Q}) \rightarrowtail \mathrm{Aut}_{\mathbf{F}} V(\overline{\mathbf{Q}}).$$

We now assume V is two-dimensional over $\mathbf{F}$ and is given by polynomials P, S and M_a for $a \in \mathbf{F}$, as in § 5.4. The following algorithm, which is the same as the one described by Couveignes and Edixhoven in [17, § 14.7], computes ρ_V in this situation. It is based on the following observation. Under the usual correspondence between finite $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-sets and finite étale Q-algebras, let A be the $\mathbf{Q}$-algebra corresponding to $V(\overline{\mathbf{Q}})$, and let B be the $\mathbf{Q}$-algebra corresponding to $\mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(\overline{\mathbf{Q}}))$. Then there is an isomorphism

$$A \cong \mathbf{Q}[x]/(P)$$

of $\mathbf{Q}$-algebras, and the inclusion

$$\begin{aligned} \mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(\overline{\mathbf{Q}})) &\to V(\overline{\mathbf{Q}})^2 \\ \alpha &\mapsto (\alpha(1,0), \alpha(0,1)) \end{aligned}$$

induces a surjection

$$\mathbf{Q}[x_1, x_2]/(P(x_1), P(x_2)) \to B.$$

The natural right action of $\mathrm{GL}_2(\mathbf{F})$ on $\mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(\overline{\mathbf{Q}}))$ gives a left action of $\mathrm{GL}_2(\mathbf{F})$ on B.

The elements of $\mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(K_V)) = \mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(\overline{\mathbf{Q}}))$ correspond bijectively to the $\mathbf{Q}$-algebra homomorphisms $B \to K_V$. We fix one isomorphism

$$\phi \in \mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(K_V)).$$

Since V is generated by the image of ϕ and K_V is the splitting field of V, the $\mathbf{Q}$-algebra homomorphism $B \to K_V$ corresponding to ϕ is surjective. This means that the choice of ϕ gives an identification of K_V with a quotient of B.

Let $T \subseteq \mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(K_V))$ be the $\mathrm{Gal}(K_V/\mathbf{Q})$-orbit of ϕ, and let G_T be the subgroup of $\mathrm{GL}_2(\mathbf{F})$ consisting of elements that preserve T. Since $\mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(K_V))$ is a right $\mathrm{GL}_2(\mathbf{F})$-torsor, T is a right G_T-torsor, and the choice of ϕ gives an isomorphism

$$\begin{aligned} G_T &\xrightarrow{\sim} T \\ g &\longmapsto \phi \circ g. \end{aligned}$$

By definition, T is also a left $\mathrm{Gal}(K_V/\mathbf{Q})$-torsor, and ϕ gives an isomorphism

$$\begin{aligned} \mathrm{Gal}(K_V/\mathbf{Q}) &\xrightarrow{\sim} T \\ \sigma &\longmapsto \rho_V(\sigma) \circ \phi. \end{aligned}$$

Composing the second isomorphism with the inverse of the first, we get an embedding

$$\begin{aligned} \mathrm{Gal}(K_V/\mathbf{Q}) &\to \mathrm{GL}_2(\mathbf{F}) \\ \sigma &\mapsto \phi^{-1} \circ \rho_V(\sigma) \circ \phi \end{aligned} \tag{5.1}$$

whose image equals G_T.

Algorithm 5.2 *(Compute the Galois representation associated to a finite $\mathbf{Q}$-vector space scheme).* Given a finite field $\mathbf{F}$ and an $\mathbf{F}$-vector space scheme V over $\mathbf{Q}$ given by polynomials P, S and M_a for $a \in \mathbf{F}$ as in § 5.3, this algorithm outputs a Galois representation $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F})$ isomorphic to ρ_V, in the format described in § 5.2.

1. Compute the left action of $\mathrm{GL}_2(\mathbf{F})$ on the $\mathbf{Q}$-algebra $\mathbf{Q}[x_1, x_2]/(P(x_1), P(x_2))$, using S and the M_a with $a \in \mathbf{F}$.
2. Write $P = P_0 P_{\neq 0}$, with $P_0 = x - q$ as in § 5.4, and put $a = 1 - P_{\neq 0}/P_{\neq 0}(q) \in \mathbf{Q}[x]$.
3. Compute the element
$$b = \prod_{g \in \mathrm{GL}_2(\mathbf{F})} g \cdot (a \otimes 1) \in \mathbf{Q}[x_1, x_2]/(P(x_1), P(x_2)).$$
4. Compute the $\mathbf{Q}$-algebra
$$B = \big(\mathbf{Q}[x_1, x_2]/(P(x_1), P(x_2))\big)/(1 - b).$$
5. Find a maximal ideal I of B, and compute the field $K = B/I$.
6. Compute the left action of $\mathrm{GL}_2(\mathbf{F})$ on B, and find the subgroup $G_I \subset \mathrm{GL}_2(\mathbf{F})$ that stabilises I.
7. For all $g \in G_I$, compute the matrix $\sigma(g)$ of the automorphism of K induced by g.
8. Output K and the list of pairs $(\sigma(g), g)$ with $g \in G_I$.

Analysis. The definition of a implies that the canonical isomorphism

$$\mathbf{Q}[x]/(P) \xrightarrow{\sim} \mathbf{Q}[x]/(P_0) \times \mathbf{Q}[x]/(P_{\neq 0})$$

sends a to $(0, 1)$; in other words, a is the idempotent in $\mathbf{Q}[x]/(P)$ that, as a function on V, is 1 on $V \setminus \{0\}$ and 0 on $\{0\}$. By definition, $b \in \mathbf{Q}[x_1, x_2]/(P(x_1), P(x_2))$ is the idempotent that, as a function of $V(\overline{\mathbf{Q}})^2$, is 1 on $\mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(\overline{\mathbf{Q}}))$ and 0 on its complement, so B is the same $\mathbf{Q}$-algebra as in the discussion preceding the algorithm. The choice of a quotient $B \to K$ identifies K with K_V and fixes an element $\phi \in \mathrm{Isom}_{\mathbf{F}}(\mathbf{F}^2, V(K))$. The group G_I computed in step 6 consists of the elements of $\mathrm{GL}_2(\mathbf{F})$ that respect the $\mathrm{Gal}(K/\mathbf{Q})$-orbit T of ϕ, so it is equal to G_T. The definition of the action of G_I on K implies that under the representation (5.1), each element g in the image G_T corresponds to the element $\sigma(g) \in \mathrm{Gal}(K_V/\mathbf{Q})$ on the left-hand side. This shows that the output is correct. Finally, the algorithm runs in (deterministic) polynomial time in $\#\mathbf{F}$ and the largest among the heights of the coefficients of P. ◇

5.6. Twisting representations by characters

Algorithm 5.3 *(Twist a representation by a character).* Given a finite field $\mathbf{F}$, a positive integer n, a homomorphism

$$\tilde{\chi}\colon (\mathbf{Z}/n\mathbf{Z})^\times \to \mathbf{F}^\times$$

and a two-dimensional representation

$$\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{F})$$

in the format of § 5.2, this algorithm outputs the twisted representation

$$\begin{aligned}\rho'\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) &\to \mathrm{GL}_2(\mathbf{F})\\ \sigma &\mapsto \chi(\sigma)\rho(\sigma)\end{aligned}$$

in the format of § 5.2; here $\chi\colon \mathrm{Gal}(\mathbf{Q}(\zeta_n)/\mathbf{Q}) \to \mathbf{F}^\times$ is the character corresponding to $\tilde{\chi}$.

1. Compute a compositum L of K_ρ and $\mathbf{Q}(\zeta_n)$.
2. Compute $\mathrm{Gal}(L/\mathbf{Q})$ as described in § 5.1.
3. For each $\tau \in \mathrm{Gal}(L/\mathbf{Q})$, compute the element $a_\tau \in (\mathbf{Z}/n\mathbf{Z})^\times$ such that $\tau(\zeta_n) = \zeta_n^{a_\tau}$, the restriction σ_τ of σ to K_ρ, and

$$\rho'(\tau) = \tilde{\chi}(a_\tau)\rho(\sigma_\tau) \in \mathrm{GL}_2(\mathbf{F}).$$

4. Compute the subgroup $\ker\rho'$ of $\mathrm{Gal}(L/\mathbf{Q})$.
5. Compute $K_{\rho'}$ as the fixed field of $\ker\rho'$.
6. For each $\sigma \in \mathrm{Gal}(K_{\rho'}/\mathbf{Q}) = \mathrm{Gal}(L/\mathbf{Q})/\ker\rho'$, output the matrix of σ on $K_{\rho'}$ and the element $\rho'(\sigma)$.

Analysis. It is clear that the algorithm is correct. It runs in (deterministic) polynomial time in n, $\#\mathbf{F}$ and the largest among the heights of the coefficients in the multiplication table of ρ. ◇

5.7. Finding the Frobenius conjugacy class

Let $K \subseteq L$ be a Galois extension of number fields, and let $\mathfrak{p}$ be a prime of K such that the extension is unramified at $\mathfrak{p}$. We will now describe how to identify the Frobenius conjugacy class at $\mathfrak{p}$ inside $\mathrm{Gal}(L/K)$.

For simplicity we restrict ourselves to the case where $K = \mathbf{Q}$, so $\mathfrak{p}$ is a rational prime p. First we compute an order O in L that is maximal at p. There are well-known ways to do this; see for example Buchmann and Lenstra [11, Algorithm 6.1]. We then compute the finite $\mathbf{F}_p$-algebra

$$A = O/(p),$$

together with the Frobenius automorphism

$$\begin{aligned} \mathrm{Frob}_p \colon A &\to A \\ a &\mapsto a^p \end{aligned}$$

and the action of $\mathrm{Gal}(L/\mathbf{Q})$ on A.

One way to continue would be to find the primary decomposition of A. Instead of describing this approach, we give a deterministic way to find the Frobenius conjugacy class, due to H. W. Lenstra. For each $\sigma \in \mathrm{Gal}(L/\mathbf{Q})$, we compute the ideal I_σ of A generated by the image of the $\mathbf{F}_p$-linear map $\mathrm{Frob}_p - (\sigma \bmod p)$. Now A/I_σ is the largest quotient of A on which $\sigma \bmod p$ acts as Frob_p. This means that σ is in the Frobenius conjugacy class if and only if the ideal I_σ is strictly smaller than A.

Chapter V

Computing modular Galois representations

1. Introduction

Let n and k be positive integers, and let l be a prime number. Let f be a modular form of weight k for $\Gamma_1(n)$ over a finite field $\mathbf{F}$ of characteristic l that is an eigenform for the Hecke operators T_p with p prime and $\langle d\rangle$ with $d \in (\mathbf{Z}/n\mathbf{Z})^\times$, with eigenvalues a_p and $\epsilon(d)$, respectively.

The goal of this chapter is to give an algorithm for computing the semi-simple two-dimensional representation

$$\rho_f\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}} W_f$$

associated to f by Theorem I.3.3. This ρ_f is uniquely defined by the following properties: ρ_f is unramified outside nl, and the characteristic polynomial of the Frobenius conjugacy class at a prime $p \nmid nl$ equals $t^2 - a_p t + \epsilon(p)p^{k-1} \in \mathbf{F}[t]$.

Let K_f denote the finite Galois extension of $\mathbf{Q}$ such that ρ_f factors as

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K_f/\mathbf{Q}) \rightarrowtail \mathrm{Aut}_{\mathbf{F}} W_f.$$

By "computing ρ_f" we mean producing the following data:

(1) the multiplication table of K_f with respect to some $\mathbf{Q}$-basis $(b_1, \ldots, b_r)$ of K_f;

(2) for every $\sigma \in \mathrm{Gal}(K_f/\mathbf{Q})$, the matrix of σ with respect to the basis $(b_1, \ldots, b_r)$ and the matrix of $\rho_f(\sigma)$ with respect to some fixed $\mathbf{F}$-basis of W_f.

Moreover, we want to do this *efficiently*. Ideally, we would have an algorithm that computes these data in polynomial time in n, k and $\#\mathbf{F}$. Unfortunately, several difficulties present themselves that prevent us from stating a completely general result.

First, our approach only leads to *probabilistic* algorithms, due to the fact that it is based on the algorithms in Chapter IV.

Second, we will give a bound for the expected running time of our algorithm that depends on certain real numbers that are defined as follows. For every smooth, proper and geometrically connected curve X over $\mathbf{Q}$, we define

$$\gamma(X) = \sum_{p\,\mathrm{prime}} \gamma(X_{\mathbf{W}_p}) \log p, \tag{1.1}$$

where $\mathbf{W}_p$ is the field of fractions of the ring of Witt vectors of $\overline{\mathbf{F}}_p$, and $\gamma(X_{\mathbf{W}_p})$ is the real number defined in § III.4.2. The bound for the running time contains a term that is linear in $\gamma(\mathrm{X}_1(n'))$, where n' equals n or nl, depending on k, and where $\mathrm{X}_1(n')$ is the coarse moduli space defined in § I.1.1. We therefore need a bound on $\gamma(\mathrm{X}_1(n'))$ that is polynomial in n. The problem is that we do not have enough information about the semi-stable reduction of $\mathrm{X}_1(n)$ at primes p such that p^2 divides n to find such a bound.

Third, we recall from § I.3.6 that if the desired Galois representation is irreducible, then it is realised, up to a twist by a character, as a simple constituent of $J[\mathfrak{m}](\overline{\mathbf{Q}})$, where J is the Jacobian of the modular curve $\mathrm{X}_1(n')$ for a certain n', and $\mathfrak{m}$ is a maximal ideal of the Hecke algebra $\mathbf{T}_1(n') \subseteq J[\mathfrak{m}]$. The expected running time of our algorithm is polynomial in the degree of $J[\mathfrak{m}]$ over $\mathbf{Q}$, which is problematic if $J[\mathfrak{m}](\overline{\mathbf{Q}})$ is composed of many copies of the representation. If we restrict ourselves to those cases in which the "simplicity" phenomenon described in § I.3.7 holds (for example, $2 \le k \le l-1$), then the running time is polynomial in n, k and $\#\mathbf{F}$. We would be able to state the same conclusion in general if an absolute bound were known on the dimension of the $\mathbf{F}$-vector space scheme $J[\mathfrak{m}]$. Extensive computations of Hecke algebras on spaces of cusp forms of prime weight by Kilford and Wiese [58] with Hecke algebras have not revealed any cases where the multiplicity is greater than two, but it is unknown as of this writing whether the multiplicity can be greater than two.

Taking into account these restrictions, we do have the following result. For clarity, we state a result that is slightly weaker than what we actually prove in this chapter.

Theorem 1.1. *Let a be a positive integer. There is a probabilistic algorithm that, given a squarefree positive integer b coprime to a, an integer $k \ge 2$, a finite field $\mathbf{F}$ of characteristic greater than k and a Hecke eigenform f of weight k for $\Gamma_1(ab)$ over $\mathbf{F}$, computes the Galois representation ρ_f in expected time polynomial in b and $\#\mathbf{F}$.*

2. Reduction to torsion subschemes in Jacobians of modular curves

Let n and k be positive integers, let l be a prime number, and let f be an eigenform of weight k for $\Gamma_1(n)$ over a finite field $\mathbf{F}$ of characteristic l.

We assume that l does not divide n. As explained in § I.3.3, this is not a real restriction, since we can always find a form for $\Gamma_1(m)$, with $m \mid n$ and $l \nmid m$, whose attached Galois representation is isomorphic to ρ_f.

2.1. Reduction to irreducible representations

In the algorithm that we describe in this chapter the case where ρ_f is absolutely irreducible is treated in an essentially different way than the case where it is not, with almost all the work going into the former case. We therefore begin by deciding whether ρ_f is absolutely irreducible.

In general, ρ_f is absolutely irreducible if and only if it is irreducible after extension of scalars to a quadratic extension of $\mathbf{F}$. If $l > 2$, then the fact that any complex conjugation has two distinct eigenvalues implies the stronger statement that if ρ_f is irreducible over $\mathbf{F}$, then it is absolutely irreducible. After replacing $\mathbf{F}$ by a quadratic extension if $l = 2$, the question is therefore equivalent to the question whether ρ_f is reducible.

We recall from § I.3.4 that if ρ_f is reducible, it is of the form $\epsilon_1\chi_l^i \oplus \epsilon_2\chi_l^j$, where ϵ_1 and ϵ_2 are characters of conductors n_1 and n_2, respectively, such that $n_1 n_2 \mid n$ and $\epsilon_1\epsilon_2 = \epsilon$, and where $i + j = k - 1$ in $\mathbf{Z}/(l-1)\mathbf{Z}$. Conversely, given such ϵ_1 and ϵ_2, the Eisenstein series $E_{k'}^{\epsilon_1,\epsilon_2}$, where $3 \le k' \le l+1$ and $k' \equiv k \pmod{l-1}$, has $\epsilon_1 \oplus \epsilon_2\chi_l^{k-1}$ as its associated Galois representation. By Theorem I.3.5, we can therefore decide whether ρ_f is reducible by comparing the eigenvalues of the Hecke operators on f (or the coefficients of the q-expansion of f) to the coefficients of these Eisenstein series. Moreover, if ρ_f is reducible, it is straightforward to write down ρ_f in the desired form.

2.2. Reduction to torsion in Jacobians

From now on we consider the case where the representation attached to f is absolutely irreducible; in particular, f is a cusp form. We have seen in § I.3.3 that there exist integers j and $\tilde{k}$ such that

$$0 \le j \le l-1, \quad 1 \le \tilde{k} \le l+1 \quad \text{and} \quad \tilde{k} \equiv k + 2j \pmod{l-1}$$

and an eigenform $\tilde{f}$ of weight $\tilde{k}$ for $\Gamma_1(n)$ over $\mathbf{F}$ such that the eigenvalues of the Hecke operators on f and on $\tilde{f}$ are related by the formula

$$T_p\tilde{f} = (p \bmod l)^j a_p \tilde{f} \text{ for } p \ne l \text{ prime} \quad \text{and} \quad \langle d\rangle\tilde{f} = \epsilon(d)\tilde{f} \text{ for } d \in (\mathbf{Z}/n\mathbf{Z})^\times.$$

The representation

$$\rho_{\tilde{f}}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}} W_{\tilde{f}}$$

attached to $\tilde{f}$ is irreducible, too, and the above equation shows that we can compute ρ_f as the twist of $\rho_{\tilde{f}}$ by the $(-j)$-th power of the l-cyclotomic character.

If $\tilde{k} = 1$, we can find an eigenform of weight l for $\Gamma_1(n)$ over a quadratic extension $\mathbf{F}'$ of $\mathbf{F}$ whose associated Galois representation is isomorphic to $\mathbf{F}' \otimes_{\mathbf{F}} \rho_f$; see Edixhoven [31, proof of Proposition 2.7]. As a representation over $\mathbf{F}$, this $\mathbf{F}' \otimes_{\mathbf{F}} \rho_f$ is a direct sum of two copies of ρ_f, and we can use § IV.5.4 to extract ρ_f. This reduces the problem to the case where $\tilde{k} \ge 2$.

We let $\mathbf{F}_{\tilde{f}}$ denote the field generated by the eigenvalues of the Hecke operators on $\tilde{f}$. As in § I.3.6, we write

$$n' = \begin{cases} n & \text{if } \tilde{k} = 2; \\ nl & \text{if } 3 \le \tilde{k} \le l+1. \end{cases}$$

We consider the Abelian variety $\mathrm{J}_1(n')_{\mathbf{Q}}$ and the Hecke algebra

$$\mathbf{T}_1(n') \subseteq \mathrm{End}\,\mathrm{J}_1(n')_{\mathbf{Q}}$$

defined in § I.1.3. We have seen in § I.3.6 that there exists a surjective ring homomorphism

$$e_{\tilde{f}}\colon \mathbf{T}_1(n') \to \mathbf{F}_{\tilde{f}}$$

that sends each T_p for p prime and each $\langle d\rangle_n$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ to the eigenvalue of the corresponding operator on $\tilde{f}$ and, in case $\tilde{k} > 2$, sends $\langle d\rangle_l$ for $d \in (\mathbf{Z}/l\mathbf{Z})^\times$ to $d^{\tilde{k}-2}$. If $\mathfrak{m}_{\tilde{f}}$ denotes the kernel of $e_{\tilde{f}}$, the representation

$$\rho_{\mathrm{J}_1(n')[\mathfrak{m}_{\tilde{f}}]}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}_{\mathbf{F}_{\tilde{f}}}(\mathrm{J}_1(n')[\mathfrak{m}_{\tilde{f}}](\overline{\mathbf{Q}}))$$

is non-zero and all its simple constituents are isomorphic to $\rho_{\tilde{f}}$. Usually, $\rho_{\mathrm{J}_1(n')[\mathfrak{m}_{\tilde{f}}]}$ itself is already simple, as explained in § I.3.7. This reduces the problem to computing Galois representations of the form $\rho_{\mathrm{J}_1(n')[\mathfrak{m}_{\tilde{f}}]}$.

3. Galois representations in torsion of Jacobians: notation and overview

In this section we will explain the strategy for computing Galois representations of the form $\rho_{J[\mathfrak{m}]}$, where J is the Jacobian of a modular curve and $\mathfrak{m}$ is a maximal ideal of the corresponding Hecke algebra. Details will be given in the next sections.

3.1. The situation

Let n be a positive integer, let l be a prime number not dividing n, and let n' be either n or nl. We abbreviate

$$X = \mathrm{X}_1(n')_{\mathbf{Z}[1/nl]}, \quad J = \mathrm{J}_1(n')_{\mathbf{Z}[1/nl]}, \quad \mathbf{T} = \mathbf{T}_1(n').$$

We write g for the genus of the fibres of X, which equals the dimension of the fibres of J.

Let O denote the rational cusp of X corresponding to the Néron polygon $E_{n'}$ with n' sides, consecutively labelled by $\mathbf{Z}/n'\mathbf{Z}$, with the embedding $\mathbf{Z}/n'\mathbf{Z} \to E_{n'}$ sending $a \in \mathbf{Z}/n'\mathbf{Z}$ to the point 1 of the copy of $\mathbf{P}^1$ labelled a.

Let $\mathbf{F}$ be a finite field of characteristic l, and let $e\colon \mathbf{T} \to \mathbf{F}$ be a surjective ring homomorphism. Let $\mathfrak{m}$ be the kernel of e, and let $J[\mathfrak{m}]$ be the maximal closed subscheme of J annihilated by $\mathfrak{m}$; this is a finite étale covering of $\operatorname{Spec}\mathbf{Z}[1/nl]$ (see § I.3.6). We write $\deg J[\mathfrak{m}]$ for the degree of this covering.

3.2. Stratifications and the scheme $D_\mathfrak{m}$

We consider the d-th symmetric power $\mathrm{Sym}^d X$ of X over $\operatorname{Spec}\mathbf{Z}[1/nl]$, which is by definition the quotient of $X \times X \times \ldots \times X$ (d factors) by the symmetric group S_d; this exists because X is projective over $\operatorname{Spec}\mathbf{Z}[1/nl]$. This scheme represents the functor of effective divisors of degree d on X. The choice of O in § 3.1 gives proper morphisms

$$\mathrm{Sym}^d X \to J \quad (0 \le d \le g)$$

sending an effective divisor D of degree d to the class of the divisor $D - d\cdot O$. For $d = g$, this morphism is birational. We let J_d denote the image of $\mathrm{Sym}^d X$ in J, so that we have a chain of closed immersions

$$\{0\} = J_0 \subseteq J_1 \subseteq \ldots \subseteq J_g.$$

We define the *stratification* of $J[\mathfrak{m}]$ as the chain of finite $\mathbf{Z}[1/nl]$-schemes

$$\{0\} = J[\mathfrak{m}]_0 \subseteq J[\mathfrak{m}]_1 \subseteq \ldots \subseteq J[\mathfrak{m}]_g = J[\mathfrak{m}],$$

where

$$J[\mathfrak{m}]_d = J[\mathfrak{m}] \cap J_d.$$

Note that the $J[\mathfrak{m}]_i$ are not necessarily flat over $\mathbf{Z}[1/nl]$, except for $J[\mathfrak{m}]_0$ and $J[\mathfrak{m}]$ itself. We define the *generic stratification type of* $J[\mathfrak{m}]$ to be the non-decreasing sequence of positive integers

$$\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}}) = \bigl(1 = \deg(J[\mathfrak{m}]_0)_{\mathbf{Q}}, \deg(J[\mathfrak{m}]_1)_{\mathbf{Q}}, \ldots, \deg(J[\mathfrak{m}]_g)_{\mathbf{Q}} = \deg J[\mathfrak{m}]\bigr).$$

Similarly, for every prime number p not dividing nl, we define the *stratification type of* $J[\mathfrak{m}]$ *modulo* p as

$$\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p}) = \bigl(1 = \deg(J[\mathfrak{m}]_0)_{\mathbf{F}_p}, \ldots, \deg(J[\mathfrak{m}]_g)_{\mathbf{F}_p} = \deg J[\mathfrak{m}]\bigr).$$

For every $x \in J[\mathfrak{m}](\overline{\mathbf{Q}})$, we let D_x denote the O-normalised representative of x as defined in § IV.2.9. Let K be the splitting field of $J[\mathfrak{m}]_{\mathbf{Q}}$ inside $\overline{\mathbf{Q}}$, and let R be the integral closure of $\mathbf{Z}[1/nl]$ in K. Since $J[\mathfrak{m}]$ is finite étale over $\mathbf{Z}[1/nl]$, we can interpret the D_x as the set of R-points of a closed subscheme

$$D_{\mathfrak{m}} \rightarrowtail \operatorname{Sym}^g X$$

that is finite étale over $\mathbf{Z}[1/nl]$. The morphism $\operatorname{Sym}^g X \to J$ restricts to an isomorphism

$$c_{\mathfrak{m}}\colon D_{\mathfrak{m}} \xrightarrow{\sim} J[\mathfrak{m}].$$

Similarly, for every prime number $p \nmid nl$ and every $x \in J[\mathfrak{m}](\overline{\mathbf{F}}_p)$ we define $d_x^{\mathbf{F}_p}$ as the least integer d for which x is in $(J_d)_{\mathbf{F}_p}$, and we write $D_x^{\mathbf{F}_p}$ for the O-normalised representative of x. We view the $D_x^{\mathbf{F}_p}$ as the $\overline{\mathbf{F}}_p$-points of a closed subscheme

$$D_{\mathfrak{m}}^{\mathbf{F}_p} \rightarrowtail \operatorname{Sym}^g X_{\mathbf{F}_p}.$$

We have isomorphisms

$$(D_{\mathfrak{m}})_{\mathbf{Q}} \xrightarrow{\sim} J[\mathfrak{m}]_{\mathbf{Q}}$$

and

$$D_{\mathfrak{m}}^{\mathbf{F}_p} \xrightarrow{\sim} J[\mathfrak{m}]_{\mathbf{F}_p} \quad (p \nmid nl \text{ prime}),$$

making $(D_{\mathfrak{m}})_{\mathbf{Q}}$ and $D_{\mathfrak{m}}^{\mathbf{F}_p}$ into $\mathbf{F}$-vector space schemes over $\operatorname{Spec}\mathbf{Q}$ and over $\operatorname{Spec}\mathbf{F}_p$, respectively. However, the subschemes $D_{\mathfrak{m}}^{\mathbf{F}_p}$ and $(D_{\mathfrak{m}})_{\mathbf{F}_p}$ of $\operatorname{Sym}^g X_{\mathbf{F}_p}$ do not in general coincide for every prime number p. The reason for this is that if K is a number field an element $x \in J(K)$ that is not in some J_d may still specialise modulo a prime of K to a point that *is* in J_d. In other words, if $\mathcal{L}$ is a line bundle on X over some number field, there may exist an integer d such that the reduction of $\mathcal{L}(dO)$ modulo p has non-zero global sections while the same does not hold for $\mathcal{L}(dO)$ over $\mathbf{Q}$. We will come back to this phenomenon in § 4.3 below.

3.3. Overview of the algorithm

The goal is to find an explicit representation for the finite $\mathbf{F}$-vector space scheme $J[\mathfrak{m}]_{\mathbf{Q}}$ over $\mathbf{Q}$. The basic strategy is to choose a suitable closed immersion

$$\iota\colon J[\mathfrak{m}]_{\mathbf{Q}} \to \mathbf{A}^1_{\mathbf{Q}}$$

of $\mathbf{Q}$-schemes. The meaning of "suitable" will become clear in Section 5 below. The image of ι is defined by some monic polynomial $P_\iota \in \mathbf{Q}[x]$ of degree equal to $\deg J[\mathfrak{m}]$. As explained in § IV.5.3, the $\mathbf{F}$-vector space scheme structure on the image of ι is given by polynomials S and M_a for $a \in \mathbf{F}$ describing the addition and scalar multiplication.

The question is now how to find $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$, or equivalently the polynomials P_ι, S and M_a. The approach that we will take here is due to Couveignes, and comes down to approximating $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$, either over the complex numbers or modulo many small prime numbers. For pressing reasons of space, time and technicalities, I limit myself the second approach. In view of the results of Couveignes, Edixhoven et al. [17], however, it may be expected that the approximations can also be done over the complex numbers. This would lead to deterministic variants of the results of this chapter.

Let $\tilde S \in \mathbf{Q}[x_1, x_2]$ be the unique representative of $S \in \mathbf{Q}[x_1, x_2]/(P_\iota(x_1), P_\iota(x_2))$ that has degree less than $\deg P_\iota$ in both x_1 and x_2. Similarly, for $a \in \mathbf{F}$ let $\tilde M_a \in \mathbf{Q}[x]$ be the unique representative of $M_a \in \mathbf{Q}[x]/(P_\iota)$ that has degree less than $\deg P_\iota$.

Definition. The *height* of $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$ is the maximum of the logarithmic heights of the coefficients of P_ι, $\tilde S$ and the $\tilde M_a$ for $a \in \mathbf{F}$.

Let p be a prime number not dividing nl. As will be explained in Section 4 below, we can represent $J_{\mathbf{F}_p}$ in a way that is well suited for computations, and we can compute the action of the Hecke algebra on $J_{\mathbf{F}_p}$. Given $\mathbf{T}_1(n')$ and $\mathfrak{m}$, we can find a finite extension k_p of $\mathbf{F}_p$ over which $J[\mathfrak{m}]_{\mathbf{F}_p}$ splits. Using our algorithms for computing in J, we then find the $\mathbf{F}$-vector space $J[\mathfrak{m}]_{\mathbf{F}_p}(k_p)$. From this we compute $D_{\mathfrak{m}}^{\mathbf{F}_p}$, which for all p outside some finite set equals $(D_{\mathfrak{m}})_{\mathbf{F}_p}$. It can be checked whether p is in this set; this is rather non-trivial and will be explained in § 4.3 below.

The closed embedding ι is constructed as follows. Once we know $D_{\mathfrak{m}}$ for a suitable prime number p, we use this to choose a non-constant rational function

$$\psi\colon X \to \mathbf{P}^1_{\mathbf{Q}}$$

and a rational map

$$\lambda\colon \mathbf{P}^g_{\mathbf{Q}} \dashrightarrow \mathbf{A}^1_{\mathbf{Q}}$$

such that the composed map

$$\operatorname{Sym}^g X_{\mathbf{Q}} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}} \overset{\lambda}{\dashrightarrow} \mathbf{A}^1_{\mathbf{Q}}$$

gives a well-defined closed immersion of $(D_{\mathfrak{m}})_{\mathbf{Q}}$ into $\mathbf{A}^1_{\mathbf{Q}}$. We then define ι as the composition of the maps

$$J[\mathfrak{m}]_{\mathbf{Q}} \xrightarrow[\sim]{c_{\mathfrak{m}}^{-1}} (D_{\mathfrak{m}})_{\mathbf{Q}} \rightarrowtail \operatorname{Sym}^g X_{\mathbf{Q}} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}} \overset{\lambda}{\dashrightarrow} \mathbf{A}^1_{\mathbf{Q}}.$$

After choosing ι, we compute the reductions of the $\mathbf{F}$-vector space scheme $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$ modulo sufficiently many small prime numbers. We then reconstruct $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$ from these reductions. If necessary, we extract one simple component from $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$ as described in § IV.5.4. The corresponding two-dimensional irreducible Galois representation can then be computed as described in § IV.5.2.

4. Computations modulo prime numbers

4.1. Representing modular curves over finite fields

As the basis for our algorithm, we use the methods for computing in Picard groups of projective curves over finite fields explained in Chapter IV. For any prime number p not dividing nl, we consider the curve $X_{\mathbf{F}_p}$ over the field $\mathbf{F}_p$. We take the projective embedding of $X_{\mathbf{F}_p}$ given by the line bundle

$$\mathcal{L} = \omega^{\otimes 2}$$

of modular forms of weight 2. This line bundle satisfies the essential inequality $\deg \mathcal{L} \geq 2g + 1$ of § IV.2.1 because $\deg \mathcal{L}$ equals $2g - 2$ plus the number of cusps (this follows from formulae for these quantities or from the existence of the *Kodaira–Spencer isomorphism* between the sheaf of differentials and the sheaf of cusp forms of weight 2), and X has at least three cusps (this also follows from well-known formulae). This choice of $\mathcal{L}$ implies

$$S_{X_{\mathbf{F}_p}} = \bigoplus_{i=0}^{\infty} \mathrm{M}_{2i}(\Gamma_1(n'), \mathbf{F}_p). \tag{4.1}$$

In other words, $S_{X_{\mathbf{F}_p}}$ is the algebra of modular forms of even weight for $\Gamma_1(n')$ over $\mathbf{F}_p$. We represent such forms by q-expansions at the rational cusp 0 of $X_{\mathbf{F}_p}$ (see § I.2.4) up to sufficient order.

The first thing to do is finding the data needed to represent $X_{\mathbf{F}_p}$ in the form needed for the algorithms of Chapter IV. As explained in § IV.4.1, we can use modular symbols to compute the Hecke algebra

$$\mathbf{T}_1(n') \cong \mathbf{T}(\mathrm{S}_2(\Gamma_1(n'), \mathbf{Z}))$$

on cusp forms of weight 2, together with the diamond operators in $\mathbf{T}_1(n')$, in time polynomial in n'. Furthermore, for any positive integer m we can compute the element T_m of $\mathbf{T}_1(n')$ in time polynomial in m. We can then compute a basis of q-expansions for the space

$$\mathrm{S}_2(\Gamma_1(n'), \mathbf{F}_p) = \mathrm{Hom}_{\mathbf{Z}\text{-mod}}(\mathbf{T}_1(n'), \mathbf{F}_p)$$

of cusp forms. Furthermore, we can compute a basis of q-expansions for the space of Eisenstein series of weight 2 for $\Gamma_1(n')$ using the formulae from § II.2.3. The fact that $\deg \mathcal{L} \geq 2g + 1$ implies that the $\mathbf{F}_p$-algebra $S_{X_{\mathbf{F}_p}}$ is generated by the homogeneous elements of degree 1, so by multiplying q-expansions of forms of weight 2 we can compute $S^{(h)}_{X_{\mathbf{F}_p}}$ for any positive integer h in time polynomial in n', h and $\log p$. Taking $h = 7$ will be sufficient for all the algorithms that we will need.

We also recall from § IV.4.2 that we can compute the zeta function of $X_{\mathbf{F}_p}$, in the form of its numerator $L_{X_{\mathbf{F}_p}}$, in time polynomial in n' and p.

Remark. The fact that the running time of the algorithm to compute the zeta function is exponential in $\log p$ is the reason why we need small primes p.

Once we have computed $S^{(7)}_{X_{\mathbf{F}_p}}$ and the zeta function of $X_{\mathbf{F}_p}$, we can use the algorithms for computing in the Jacobian of a curve over a finite field that were described in Chapter IV.

4.2. Computing the action of the Hecke algebra

We will now explain how to compute the action of the Hecke algebra $\mathbf{T}_1(n')$ on $J_{\mathbf{F}_p}$. For all $d \in (\mathbf{Z}/n'\mathbf{Z})^\times$, the automorphism r_d of $X = \mathrm{X}_1(n')$ induces the diamond operator $\langle d\rangle$ on $\mathrm{M}_k(\Gamma_1(n'), \mathbf{F}_p)$ for all k. Since we know the action of the Hecke algebra on the $\mathrm{M}_k(\Gamma_1(n'), \mathbf{F}_p)$, we can compute the map $r_d^\#\colon S^{(h)}_{X_{\mathbf{F}_p}} \xrightarrow{\sim} S^{(h)}_{X_{\mathbf{F}_p}}$ using (4.1). This means that we can compute $\operatorname{Pic} r_d$ and $\operatorname{Alb} r_d$ by means of Algorithms IV.2.14 and IV.2.15 (where we use O as the rational point needed by the latter algorithm) in time polynomial in n' and $\log p$.

Now let r be a prime number different from p. We consider the maps

$$q_1, q_2\colon \mathrm{X}_1(n'; r)_{\mathbf{F}_p} \to X_{\mathbf{F}_p}$$

defining the Hecke operator T_r. We denote by ω the line bundle of modular forms of weight 1 on $X_{\mathbf{F}_p}$. We make $X_{\mathbf{F}_p}$ into a projective curve via the line bundle $q_1^*\omega$. The assumption that $r \neq p$ implies that the canonical map

$$\phi^*\colon q_2^*\omega \xrightarrow{\sim} q_1^*\omega$$

from § I.2.2 is an isomorphism. From the formulae for the effect of q_1^* and q_2^* on q-expansions given in § I.2.4, we can compute the maps

$$q_1^\#, q_2^\#\colon S^{(h)}_{X_{\mathbf{F}_p}} \to S^{(h)}_{\mathrm{X}_1(n';r)_{\mathbf{F}_p}}$$

for any h. This allows us to compute the maps

$$\operatorname{Pic} q_1, \operatorname{Pic} q_2\colon J_{\mathbf{F}_p} \to \mathrm{J}_1(n'; r)_{\mathbf{F}_p}$$

and

$$\operatorname{Alb} q_1, \operatorname{Alb} q_2\colon \mathrm{J}_1(n'; r)_{\mathbf{F}_p} \to J_{\mathbf{F}_p}.$$

In particular, we can compute T_r. The expected running time is polynomial in n', r and $\log p$.

Let Frob_p and Ver_p denote the Frobenius and Verschiebung endomorphisms of J. Let k be a finite extension of degree d of $\mathbf{F}_p$, and let x be an element of $J(k)$. From $\mathrm{Frob}_p \mathrm{Ver}_p = p$ and $\mathrm{Frob}_p^d(x) = x$ it follows that we can compute Ver_p on x using the formula

$$\mathrm{Ver}_p(x) = p \,\mathrm{Frob}_p^{d-1}(x).$$

Since we can compute Frob_p, Ver_p and $\langle p\rangle$, we can compute T_p using the Eichler–Shimura relation

$$T_p = \mathrm{Frob}_p + \langle p\rangle \mathrm{Ver}_p$$

from § I.1.4. The running time is polynomial in n' and $\log p$.

4.3. Good prime numbers

Definition. We say that a prime number p is $\mathfrak{m}$-*good* if the following two conditions hold:

(1) p does not divide nl;

(2) the stratification type $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$ modulo p is equal to the generic stratification type $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}})$. (Recall that in general we have a pointwise inequality $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p}) \geq \operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}})$.)

Otherwise we say that p is $\mathfrak{m}$-*bad*. We define a positive integer $B_\mathfrak{m}$ by

$$B_\mathfrak{m} = \prod_{p\ \mathfrak{m}\text{-bad}} p.$$

We note that $\deg J[\mathfrak{m}]_{\mathbf{F}_p,d} \geq \deg J[\mathfrak{m}]_d$ for all prime numbers $p \nmid nl$, with equality for all d if and only if p is $\mathfrak{m}$-good. We also note that in this case we can identify $D_\mathfrak{m}^{\mathbf{F}_p}$ with the fibre over $\mathbf{F}_p$ of the closed subscheme $D_\mathfrak{m}$ of $\operatorname{Sym}^g X_{\mathbf{Z}[1/nl]}$. This fact enables us to compute the reduction of the closed subscheme $D_\mathfrak{m}$ of $\operatorname{Sym}^g X$ modulo $\mathfrak{m}$-good prime numbers.

The following algorithm computes the $\mathbf{F}$-vector space of points of $D_\mathfrak{m}^{\mathbf{F}_p}$ over a suitable finite extension k of $\mathbf{F}_p$ over which $D_\mathfrak{m}^{\mathbf{F}_p}$ splits.

Algorithm 4.1 *(Compute $D_\mathfrak{m}^{\mathbf{F}_p}$ and the stratification type of $J[\mathfrak{m}]_{\mathbf{F}_p}$).* Let the notation be as above. Given the ring homomorphism $e\colon \mathbf{T} \to \mathbf{F}$ and a prime number $p \nmid nl$, this algorithm outputs the following information:

(1) the stratification type $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$;

(2) a finite extension k of $\mathbf{F}_p$ such that the points of $J[\mathfrak{m}]_{\mathbf{F}_p}$ are k-rational;

(3) the $\mathbf{F}_p$-algebra $S^{(7)}_{X_{\mathbf{F}_p}}$ and the k-algebra $S^{(7)}_{X_k} = k \otimes_{\mathbf{F}_p} S^{(7)}_{X_{\mathbf{F}_p}}$;

(4) the $\mathbf{F}$-vector space $D_\mathfrak{m}^{\mathbf{F}_p}(k)$, given by the positive integer

$$d = \dim_\mathbf{F} J[\mathfrak{m}](k)$$

and a list of pairs $(v, \Gamma(X_k, \mathcal{L}^{\otimes 2}(-D_{x(v)})))$, where v runs over $\mathbf{F}^d$ and $x(v)$ is the image of v under a fixed $\mathbf{F}$-linear isomorphism

$$\mathbf{F}^d \xrightarrow{\sim} D_\mathfrak{m}^{\mathbf{F}_p}(k).$$

1. Compute the $\mathbf{F}_p$-algebra $S^{(7)}_{X_{\mathbf{F}_p}}$ using modular symbols; see Section IV.4 and § 3.3.

2. Compute the polynomial $L_{X/\mathbf{F}_p} \in \mathbf{Z}[t]$ (the numerator of the zeta function of $X_{\mathbf{F}_p}$) as described in § IV.4.2.

3. Compute the order a of t in the group $\big(\mathbf{F}[t]/(t^2 - a_p t + \epsilon(p)p)\big)^\times$.

4. Generate a finite extension k of $\mathbf{F}_p$ with $[k : \mathbf{F}_p] = a$.

5. Compute the k-algebra $S^{(7)}_{X_k} = k \otimes_{\mathbf{F}_p} S^{(7)}_{X_{\mathbf{F}_p}}$.

6. Compute the polynomial $L_{X/k} \in \mathbf{Z}[u]$ (the numerator of the zeta function of X_k) as the resultant of $L_{X/\mathbf{F}_p} \in \mathbf{Z}[t]$ and $t^a - u$.

7. Compute an $\mathbf{F}_l$-basis for the l-torsion subgroup $J[l](k)$ using Algorithm IV.3.12.

8. Compute generators $t_1, \ldots, t_m$ of the $\mathbf{T}$-module $\mathfrak{m}/l\mathbf{T}$.

9. Compute the matrices of the t_i with respect to the basis of $J[l](k)$ computed in step 7, using the algorithms from §§ IV.2.11 and IV.3.7.

10. Compute an $\mathbf{F}_l$-basis for $J[\mathfrak{m}](k)$, which is the intersection of the kernels of the t_i.

11. Choose a primitive element γ of $\mathbf{F}$ over $\mathbf{F}_l$, choose a lift of γ to $\mathbf{T}$, and use this lift to compute the matrix of γ with respect to the $\mathbf{F}_l$-basis of $J[\mathfrak{m}](k)$ computed in the previous step. Use this matrix to extract an $\mathbf{F}$-basis $(b_1, \ldots, b_d)$ of $J[\mathfrak{m}](k)$.

12. For each $v = (v_1, \ldots, v_d) \in \mathbf{F}^d$, compute $x(v) = \sum_{i=1}^d v_i b_i \in J[\mathfrak{m}](k)$, and compute $d_{x(v)}$ and $\Gamma(X_{\mathbf{F}_p}, \mathcal{L}^{\otimes 2}(-D_{x(v)}))$ as described in § IV.2.9. In particular, this gives the $\mathbf{F}$-linear isomorphism

$$\mathbf{F}^d \xrightarrow{\sim} D_{\mathfrak{m}}^{\mathbf{F}_p}(k)$$

in the form of the list of pairs $(v, \Gamma(X_k, \mathcal{L}^{\otimes 2}(-D_{x(v)})))$, where v runs over $\mathbf{F}^d$.

13. Compute $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$ from the $d_{x(v)}$.

14. Output all the required information.

Analysis. By the Eichler–Shimura relation, the Frobenius automorphism of $J[\mathfrak{m}]_{\mathbf{F}_p}$ satisfies

$$\mathrm{Frob}_p^2 - a_p \mathrm{Frob}_p + \epsilon(p)p = 0,$$

where a_p and $\epsilon(p)$ are the images of T_p and $\langle p \rangle$ under the quotient map $\mathbf{T} \to \mathbf{F}$. This implies that there is a (unique) $\mathbf{F}$-algebra homomorphism

$$\mathbf{F}[t]/(t^2 - a_p t + \epsilon(p)p) \longrightarrow \operatorname{End}_{\mathbf{F}} J[\mathfrak{m}](\overline{\mathbf{F}}_p)$$

mapping t to Frob_p. The definition of a as the order of t in $\bigl(\mathbf{F}[t]/(t^2 - a_p t + \epsilon(p)p)\bigr)^\times$ therefore implies that Frob_p^a acts trivially on $J_{\mathbf{F}_p}[\mathfrak{m}]$, so that the points of $J_{\mathbf{F}_p}[\mathfrak{m}]$ are k-rational. The claim that we can compute $L_{X/k} \in \mathbf{Z}[u]$ as the resultant of $L_{X/\mathbf{F}_p} \in \mathbf{Z}[t]$ and $t^a - u$ follows from the fact that the roots of $L_{X/k}$ are the a-th powers of the roots of $L_{X/\mathbf{F}_p}$. The rest of the algorithm clearly does what it is supposed to do. Using the fact that a is at most $\#\mathbf{F}^2 - 1$, it is straightforward to check that the expected running time is polynomial in n, $\deg J[\mathfrak{m}]$ and p. ◇

Remarks. (1) In the above algorithm, some steps can be omitted if we are only interested in part of the output.

(2) The reason that the running time is polynomial in p and not $\log p$ is that we have to compute the polynomial $L_{X/\mathbf{F}_p}$.

Using the preceding algorithm, we can compute $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}})$ provided we know a bound for the number of $\mathfrak{m}$-bad primes. The following algorithm makes this precise.

Algorithm 4.2 *(Compute the generic stratification type).* Let the notation be as above. Given the ring homomorphism $e\colon \mathbf{T} \to \mathbf{F}$, this algorithm outputs the following information:

(1) the generic stratification type $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}})$ of $J[\mathfrak{m}]$;

(2) a non-empty set P of $\mathfrak{m}$-good prime numbers;

(3) for each $p \in P$ a finite extension k_p of $\mathbf{F}_p$ such that $D_{\mathfrak{m}}^{\mathbf{F}_p}$ splits over k_p and the $\mathbf{F}$-vector space $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$ (given as in Algorithm 4.1).

1. Find a positive integer $B \geq B_{\mathfrak{m}}$, where $B_{\mathfrak{m}}$ is the product of all $\mathfrak{m}$-bad primes, as defined above; see § 6.8 below.

2. Using (for example) the sieve of Eratosthenes, compute the smallest prime number β such that the set Q of prime numbers p with $p \leq \beta$ and $p \nmid nl$ satisfies

$$\prod_{p \in Q} p > B.$$

3. For all $p \in Q$, compute $X_{\mathbf{F}_p}$ and its zeta function using modular symbols, and then compute $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$, a splitting field k_p for $J[\mathfrak{m}]_{\mathbf{F}_p}$ and the $\mathbf{F}$-vector space $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$ using Algorithm 4.1.

4. Among the sequences $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$ for $p \in Q$, there is a pointwise minimum. Output this minimum together with the set P consisting of the $p \in Q$ for which it is attained, and output k_p and $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$ for all $p \in P$.

Analysis. The prime number theorem implies that the bound β computed in step 2 is at most $c \log B$ for some positive real number c. The choice of Q implies that Q contains at least one $\mathfrak{m}$-good prime. It is now clear that the algorithm is correct and runs in expected time polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B$. ◇

Remarks. (1) The form of the prime number theorem that is used here is

$$\sum_{p \leq x \text{ prime}} \log p \sim x \quad \text{as } x \to \infty.$$

In fact, it would be enough to use the following bound, proved by Chebyshev: there is a real number $c > 0$ such that

$$\sum_{p \leq x \text{ prime}} \log x \geq cx \quad \text{for all } x \geq 2.$$

(2) It seems reasonable to expect that in many cases the generic stratification type will be $(1, 1, \ldots, 1, \deg J[\mathfrak{m}])$. If this is in fact the case, then we find this out as soon as we encounter one prime number p such that the stratification type modulo p equals $(1, 1, \ldots, 1, \deg J[\mathfrak{m}])$.

Once we know the generic stratification type, Algorithm 4.1 allows us to check whether a prime number p is $\mathfrak{m}$-good in expected time polynomial in p. This implies

that for any positive integer C, we can compute a set P of $\mathfrak{m}$-good prime numbers such that
$$\prod_{p\in P} p > C$$
in expected time polynomial in $\log C$. Furthermore, for an $\mathfrak{m}$-good prime number p, the $\mathbf{F}$-vector space scheme $D_{\mathfrak{m}}^{\mathbf{F}_p}$ equals the reduction $(D_{\mathfrak{m}})_{\mathbf{F}_p}$ of $D_{\mathfrak{m}}$ modulo p. We can therefore compute the $\mathbf{F}$-vector space of points of $(D_{\mathfrak{m}})_{\mathbf{F}_p}$, again using Algorithm 4.1.

5. Choosing a suitable embedding

In this section we explain how to choose a closed immersion
$$\iota\colon J[\mathfrak{m}]_{\mathbf{Q}} \to \mathbf{A}^1_{\mathbf{Q}}.$$
We define the closed subscheme $D_{\mathfrak{m}} \rightarrowtail \operatorname{Sym}^g X$ and the isomorphism $c_{\mathfrak{m}}\colon D_{\mathfrak{m}} \xrightarrow{\sim} J[\mathfrak{m}]$ as in § 3.2. We are going to choose a non-constant rational function
$$\psi\colon X_{\mathbf{Q}} \to \mathbf{P}^1_{\mathbf{Q}}$$
such that the map
$$\psi_*\colon \operatorname{Sym}^g X_{\mathbf{Q}} \longrightarrow \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}}.$$
is a closed immersion on $(D_{\mathfrak{m}})_{\mathbf{Q}}$. We then use the isomorphism
$$\operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}}$$
given by the elementary symmetric functions. Next we choose a rational map
$$\lambda\colon \mathbf{P}^g_{\mathbf{Q}} \dashrightarrow \mathbf{A}^1_{\mathbf{Q}}$$
such that λ is well-defined and injective on the image of $(D_{\mathfrak{m}})_{\mathbf{Q}}$ under the composed map
$$(D_{\mathfrak{m}})_{\mathbf{Q}} \rightarrowtail \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}}.$$
We define ι as the composed map
$$\iota\colon J[\mathfrak{m}]_{\mathbf{Q}} \xrightarrow[\sim]{c_{\mathfrak{m}}^{-1}} (D_{\mathfrak{m}})_{\mathbf{Q}} \rightarrowtail \operatorname{Sym}^g X_{\mathbf{Q}} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}} \overset{\lambda}{\dashrightarrow} \mathbf{P}^1_{\mathbf{Q}}.$$

To choose ψ and λ, we make use of an auxiliary prime number p. This p is required to satisfy the following conditions:

(1) $p \nmid nl$;

(2) $p > 2\left(\deg \omega^{\otimes w}(-\text{cusps}) + \binom{\deg J[\mathfrak{m}]}{2}\right)$;

(3) $p > 2\left(\deg J[\mathfrak{m}] + \binom{\deg J[\mathfrak{m}]}{2}\right)$;

(4) p is $\mathfrak{m}$-good.

Algorithm 5.1 *(Choosing an auxiliary prime number).* Given the ring homomorphism $e\colon \mathbf{T} \to \mathbf{F}$ and the generic stratification type $J[\mathfrak{m}]$, this algorithm outputs a prime number p satisfying the above conditions (1)–(4), a splitting field k_p for $D_{\mathfrak{m}}^{\mathbf{F}_p}$ over $\mathbf{F}_p$ and the set $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$.

1. Let p be the least prime number satisfying the above conditions (1)–(3).
2. Compute $S^{(7)}_{X_{\mathbf{F}_p}}$ and $L_{X_{\mathbf{F}_p}}$ using modular symbols.
3. Using Algorithm 4.1, check whether p is $\mathfrak{m}$-good; if so, compute a splitting field k_p for $D_{\mathfrak{m}}^{\mathbf{F}_p}$ over $\mathbf{F}_p$ and the set $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$, output these data, and stop.
4. Replace p by the next prime number not dividing nl, and go to step 2.

Analysis. The prime number theorem implies that the prime number p output by the algorithm is bounded by a linear function of $\log B_{\mathfrak{m}}$, n^2 and $(\deg J[\mathfrak{m}])^2$; compare the remark following Algorithm 4.2. The expected running time of the algorithm is polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$. ◇

We will take the rational function ψ on $X_{\mathbf{Q}}$ of the form

$$\psi = \alpha/\beta \quad \text{with } \alpha, \beta \in \mathrm{S}_w^{\mathrm{int}}(\Gamma_1(n')),$$

where w is an integer with $1 \le w \le 12$ chosen such that

(1) the line bundle $\omega^{\otimes w}$ of modular forms of weight w on the moduli stack $\mathcal{M}_{\Gamma_1(n')}$ over $\operatorname{Spec} \mathbf{Z}$ descends to the coarse moduli space $\mathrm{X}_1(n')$;

(2) the line bundle $\omega^{\otimes w}(-\mathrm{cusps})$ on $\mathrm{X}_1(n')$ has degree $\deg \omega^{\otimes w}(-\mathrm{cusps}) \ge 2g+1$ on the fibres, and has non-negative degree on each irreducible component of each fibre.

We can always take $w = 12$.

Algorithm 5.2 *(Choosing the map ψ).* Given an auxiliary prime number p as output by Algorithm 5.1, a splitting field k_p of $J[\mathfrak{m}]_{\mathbf{F}_p}$ over $\mathbf{F}_p$ and the $\mathbf{F}$-vector space $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$, this algorithm outputs cusp forms

$$\alpha, \beta \in \mathrm{S}_w^{\mathrm{int}}(\Gamma_1(n'))$$

with the following properties:

(1) α and β have no common zeroes as sections of the line bundle $\omega^w(-\mathrm{cusps})$ on the curve $X_{\mathbf{Q}} = \mathrm{X}_1(n')_{\mathbf{Q}}$;

(2) the rational function $\psi = \alpha/\beta\colon X_{\mathbf{Q}} \to \mathbf{P}^1_{\mathbf{Q}}$ (which is well defined because of (1)) induces a closed immersion $D_{\mathfrak{m}} \rightarrowtail \mathbf{P}^g_{\mathbf{Q}}$;

(3) the logarithms of the Petersson norms of both α and β are bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$.

1. Using Algorithm IV.4.2, find a $\mathbf{Q}$-basis $(b_1, \ldots, b_N)$ of $\mathrm{S}_w(\Gamma_1(n'), \mathbf{Q})$ consisting of forms with integral q-expansion at the cusp 0 and with small Petersson norm.
2. Using modular symbols, compute $\mathrm{S}_w(\Gamma_1(n'), \mathbf{F}_p)$ and $\mathrm{M}_{w+2}(\Gamma_1(n'), \mathbf{F}_p)$ using q-expansions to precision greater than $\deg \omega^{2w+2}(-\mathrm{cusps})$.

3. Choose uniformly random elements

$$\bar\alpha = \sum_{i=1}^{N} \bar\alpha_i (b_i \bmod p), \quad \bar\beta = \sum_{i=1}^{N} \bar\beta_i (b_i \bmod p) \quad \text{in } \mathrm{S}_w(\Gamma_1(n'), \mathbf{F}_p).$$

4. Compute the image of the multiplication map

$$(\mathbf{F}_p\bar\alpha + \mathbf{F}_p\bar\beta) \otimes_{\mathbf{F}_p} \mathrm{M}_{w+2}(\Gamma_1(n'), \mathbf{F}_p) \longrightarrow \mathrm{S}_{2w+2}(\Gamma_1(n'), \mathbf{F}_p).$$

Check whether this image is the full space $\mathrm{S}_{2w+2}(\Gamma_1(n'), \mathbf{F}_p)$. If not, go to step 3.

5. We now have a morphism

$$\psi_* = (\bar\alpha/\bar\beta)_* \colon \mathrm{Sym}^g X_{\mathbf{F}_p} \to \mathrm{Sym}^g \mathbf{P}^1_{\mathbf{F}_p}.$$

Compute the images under ψ_* of the elements $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p) = (D_{\mathfrak{m}})_{\mathbf{F}_p}(k_p)$ as homogeneous polynomials of degree g using Algorithm IV.2.8. If two of these polynomials are the same (up to multiplication by elements of $k^\times$), go to step 3.

6. Output the cusp forms

$$\alpha = \sum_{i=1}^{N} \alpha_i b_i \quad \text{and} \quad \beta = \sum_{i=1}^{N} \beta_i b_i,$$

where $\alpha_1, \ldots, \alpha_N$ are integers with $|\alpha_i| \le p/2$ and $(\alpha_i \bmod p) = \bar\alpha_i$ and similarly for the β_i.

Analysis. Because of our choice of p, the probability that uniformly random $\bar\alpha$ and $\bar\beta$ do not have any common zeroes and that $\psi_* = \bar\alpha/\bar\beta$ is injective on the set $D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p)$, which has cardinality $\deg J[\mathfrak{m}]$, is at least $1/2$; see Khuri-Makdisi [57, Proposition 4.3]. This implies that the expected running time of the algorithm is bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$. The correctness of the check that α and β do not have any common zeroes follows from Lemma IV.2.3. Finally, the bound on the Petersson norms of α and β follows from the triangle inequality and the choice of basis for $\mathrm{S}_w^{\mathrm{int}}(\Gamma_1(n'))$. ◇

After we have chosen a rational function ψ, the next problem is to choose a suitable rational map $\lambda\colon \mathbf{P}^g_{\mathbf{Q}} \dashrightarrow \mathbf{A}^1_{\mathbf{Q}}$.

Algorithm 5.3 *(Choosing the map λ).* Given an auxiliary prime number p as output by Algorithm 5.1, a splitting field k_p of $J[\mathfrak{m}]_{\mathbf{F}_p}$ over $\mathbf{F}_p$, and the set of k_p-valued points of the image of the map

$$D_{\mathfrak{m}}^{\mathbf{F}_p} \rightarrowtail \mathrm{Sym}^g X_{\mathbf{F}_p} \xrightarrow{\psi_*} \mathrm{Sym}^g \mathbf{P}^1_{\mathbf{F}_p}$$

where ψ is a rational function ψ as output by Algorithm 5.2 (with the same auxiliary prime number), this algorithm outputs a $2 \times (g+1)$-matrix Λ over $\mathbf{Z}$ with coprime entries, such that the rational map

$$\lambda\colon \mathbf{P}^g_{\mathbf{Q}} \dashrightarrow \mathbf{A}^1_{\mathbf{Q}}$$

given by Λ induces a well-defined closed immersion on the image of $D_{\mathfrak{m}}$ under the map

$$\operatorname{Sym}^g X_{\mathbf{Q}} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}},$$

and such that the coefficients of Λ are bounded by a linear function of n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$.

1. Choose a random $2 \times (g+1)$-matrix $\Lambda_{\mathbf{F}_p}$ over $\mathbf{F}_p$.
2. Check whether the rational map $\mathbf{P}^g_{\mathbf{F}_p} \dashrightarrow \mathbf{A}^1_{\mathbf{F}_p}$ defined by $\Lambda_{\mathbf{F}_p}$ is well-defined and injective on the image of the map

$$D_{\mathfrak{m}}^{\mathbf{F}_p} \rightarrowtail \operatorname{Sym}^g X_{\mathbf{F}_p} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{F}_p} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{F}_p}.$$

   If not, go back to step 1.
3. Take a lift Λ of $\Lambda_{\mathbf{F}_p}$ with integer coefficients of absolute value at most $p/2$.
4. Divide Λ by the greatest common divisor of its entries, and output the result.

Analysis. The choice of the auxiliary prime number p implies that a randomly chosen λ satisfies the imposed conditions with probability at least $1/2$. It is now straightforward to check that the expected running time of the algorithm is polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$. ◇

Remark. Another possibility, described in [17, § 8.2], is to choose some small positive integer m and to take λ to be the map defined by viewing elements of $\mathbf{P}^g_{\mathbf{Q}}$ as polynomials of degree g and evaluating these in m.

6. Height bounds and bad prime numbers

The following definition says which prime numbers p can be used in our computation after maps ψ and λ have chosen as in Section 5.

Definition. We say that a prime number p is $(\mathfrak{m}, \psi)$-*good* if p is $\mathfrak{m}$-good and the map ψ is defined modulo p, i.e. α and β have no common zeroes as sections of the line bundle $\omega^{\otimes w}(-\text{cusps})$ on $\mathrm{X}_1(n')_{\mathbf{F}_p}$. We say that p is $(\mathfrak{m}, \psi, \lambda)$-*good* if in addition the map λ is well-defined on the image of the morphism

$$D_{\mathfrak{m}}^{\mathbf{F}_p} = (D_{\mathfrak{m}})_{\mathbf{F}_p} \rightarrowtail \operatorname{Sym}^g X_{\mathbf{F}_p} \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{F}_p} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{F}_p}.$$

The antonyms of $(\mathfrak{m}, \psi)$-good and $(\mathfrak{m}, \psi, \lambda)$-good are $(\mathfrak{m}, \psi)$-*bad* and $(\mathfrak{m}, \psi, \lambda)$-*bad*, respectively.

We define positive integers $B_{\mathfrak{m},\psi}$ and $B_{\mathfrak{m},\psi,\lambda}$ by

$$B_{\mathfrak{m},\psi} = \prod_{p\ (\mathfrak{m},\psi)\text{-bad}} p \quad \text{and} \quad B_{\mathfrak{m},\psi,\lambda} = \prod_{p\ (\mathfrak{m},\psi,\lambda)\text{-bad}} p,$$

where p runs over the finite set of $(\mathfrak{m}, \psi)$-bad prime numbers and that of $(\mathfrak{m}, \psi, \lambda)$-bad prime numbers, respectively.

If p is a $(\mathfrak{m}, \psi)$-good prime number, we can compute the reduction modulo p of the image of $D_\mathfrak{m}$ under the map

$$\operatorname{Sym}^g X \xrightarrow{\psi_*} \operatorname{Sym}^g \mathbf{P}^1 \xrightarrow{\sim} \mathbf{P}^g$$

as a set of k-valued points, where k is a finite extension of $\mathbf{F}_p$ as in Algorithm 4.1, using Algorithm IV.2.8. If p is in addition $(\mathfrak{m}, \psi, \lambda)$-good, then the monic polynomial defining the image of $D_\mathfrak{m}$ under the map

$$\iota\colon J[\mathfrak{m}]_\mathbf{Q} \to \mathbf{A}^1_\mathbf{Q}$$

can be reduced modulo p, and we can compute this reduction.

In order to reconstruct the $\mathbf{F}$-vector space scheme $\iota(J[\mathfrak{m}])$ from its reductions modulo prime numbers, we need an upper bound on the height of $\iota(J[\mathfrak{m}])$, which was defined in § 3.3. We will show that when the maps ψ and λ are chosen as in Algorithms 5.2 and 5.3, this height is bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$. We also need upper bounds on the integers $B_\mathfrak{m}$, $B_{\mathfrak{m},\psi}$ and $B_{\mathfrak{m},\psi,\lambda}$. To prove that our algorithm for computing modular Galois representations runs in expected time polynomial in the input size, we need to show that $\log B_\mathfrak{m}$, $\log B_{\mathfrak{m},\psi}$ and $\log B_{\mathfrak{m},\psi,\lambda}$ are bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$.

Remarks. (1) The upper bounds for $B_\mathfrak{m}$ and for the height need to be made explicit to ensure the correctness of Algorithms 4.2 above and 7.2 below, respectively. In contrast, the upper bounds for $B_{\mathfrak{m},\psi}$ and $B_{\mathfrak{m},\psi,\lambda}$ are not needed as input for the algorithm.

(2) It would certainly have been possible, but probably not very enlightening, to write down bounds that are polynomials in n and $\deg J[\mathfrak{m}]$ with real coefficients. I have aimed at a balance by giving formulae involving non-explicit constants that can, however, easily be approximated using a computer. There is at this moment still an exception, namely a bound on $\gamma(\mathrm{X}_1(n))$ in the case where n is not squarefree.

6.1. Height bounds

We suppose that maps ψ and λ as above have been chosen. Our next goal is to derive a bound on the height of $\iota(J[\mathfrak{m}])$. We start with some general observations on the behaviour of heights with respect to symmetric functions and linear maps.

If λ is an $m \times n$-matrix over $\overline{\mathbf{Q}}$, we define the height $h(\lambda)$ of λ as the height of the point of $\mathbf{P}^{mn+m+n}(\mathbf{Q})$ whose projective coordinates are the coefficients of λ.

Lemma 6.1. *Let n be a positive integer, and let $\Sigma^n\colon (\mathbf{P}^1_\mathbf{Q})^n \to \mathbf{P}^n_\mathbf{Q}$ be the symmetrisation map defined by*

$$\Sigma^n((x_1 : y_1), \ldots, (x_n : y_n)) = (\sigma_0 : \sigma_1 : \ldots : \sigma_n),$$

where

$$\sigma_k = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ \#I = k}} \prod_{i \in I} x_i \cdot \prod_{j \notin I} y_j.$$

Then the inequality

$$h_{\mathbf{P}^n}(\Sigma^n(p_1,\ldots,p_n)) \le \log\binom{n}{\lfloor n/2\rfloor} + \sum_{i=1}^n h_{\mathbf{P}^1}(p_i)$$
$$\le n\log 2 + \sum_{i=1}^n h_{\mathbf{P}^1}(p_i)$$

holds for all $p_1,\ldots,p_n \in \mathbf{P}^1(\overline{\mathbf{Q}})$.

Proof. Let K be a number field, let $p_1, \ldots, p_n$ be in $\mathbf{P}^1(K)$, and write $p_i = (x_i : y_i)$ with $x_i, y_i \in K$. For any valuation v of K, the triangle inequality implies that the elements $\sigma_k \in \overline{\mathbf{Q}}$ satisfy

$$|\sigma_k|_v \le c_v \max_{\substack{I\subseteq\{1,\ldots,n\}\\ \#I=k}} \prod_{i\in I}|x_i|_v \cdot \prod_{j\notin I}|y_j|_v,$$

where

$$c_v = \begin{cases} 1 & \text{if } v \text{ is ultrametric;}\\ \binom{n}{\lfloor n/2\rfloor}^{[K_v:\mathbf{R}]} & \text{if } v \text{ is Archimedean.}\end{cases}$$

This implies

$$\max_k |\sigma_k|_v \le c_v \prod_{i=1}^n \max\{|x_i|_v, |y_i|_v\}.$$

Taking logarithms, summing over all v and dividing by $[K:\mathbf{Q}]$ we get the first inequality. The second follows by applying the elementary inequality $\binom{n}{\lfloor n/2\rfloor} \le 2^n$. □

Lemma 6.2. *Let* $\lambda\colon \mathbf{P}^n_{\mathbf{Q}} \dashrightarrow \mathbf{P}^m_{\mathbf{Q}}$ *be a rational map given by a non-zero matrix* $(a_{i,j})_{i,j}$ *over* $\mathbf{Q}$. *Then for any* $p \in \mathbf{P}^n(\overline{\mathbf{Q}})$ *such that* λ *is defined at* p, *we have*

$$h_{\mathbf{P}^m}(\lambda(p)) \le \log n + h(\lambda) + h_{\mathbf{P}^n}(p).$$

Proof. Choose a number field K with $p \in \mathbf{P}^n(K)$, and write $p = (x_0 : \ldots : x_n)$. Put $b_i = \sum a_{i,j}x_j$ for $i = 0, \ldots, m$, so that $\lambda(p) = (b_0 : \ldots : b_m)$. Then for any valuation v of K,

$$|b_i|_v = \left|\sum_{j=0}^n a_{i,j}x_j\right|_v \le d_v \max_j |a_{i,j}x_j|_v,$$

with

$$d_v = \begin{cases} 1 & \text{if } v \text{ is ultrametric;}\\ n^{[K_v:\mathbf{R}]} & \text{if } v \text{ is Archimedean.}\end{cases}$$

Therefore,

$$\max_i |b_i|_v \le d_v \max_{i,j}|a_{i,j}|_v \max_k |x_k|_v.$$

Taking logarithms, summing over all v and dividing by $[K:\mathbf{Q}]$, we obtain the desired inequality. □

Lemma 6.3. *Let $A = (a_{i,j})_{i,j=1}^n$ and $y = (y_i)_{i=1}^n$ be elements of $\mathrm{GL}_n(\overline{\mathbf{Q}})$ and $\overline{\mathbf{Q}}^n$, respectively. Then the unique solution $x = (x_i)_{i=1}^n \in \overline{\mathbf{Q}}^n$ of $Ax = y$ satisfies*

$$\max_{1\le i\le n} h(x_i) \le 2n^2 b + n\log n,$$

where b is the maximum of all the $h(y_i)$ and $h(a_{i,j})$.

Proof. This follows from Cramer's rule and bounds on the height of the determinant of an invertible matrix in terms of the heights of its coefficients; for details we refer to Couveignes, Edixhoven et al. [17, § 4.2]. □

We now derive bounds for the heights of the points $\iota(v)$ with $v \in J[\mathfrak{m}](\overline{\mathbf{Q}})$. We consider the divisor $D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})$ corresponding to v under the isomorphism

$$c_{\mathfrak{m}}\colon D_{\mathfrak{m}}(\overline{\mathbf{Q}}) \xrightarrow{\sim} J[\mathfrak{m}](\overline{\mathbf{Q}}),$$

and we write

$$D = D_1 + \cdots + D_g \quad \text{with } D_i \in X_{\mathbf{Q}}(\overline{\mathbf{Q}}).$$

From the definition of ι we see that

$$\iota(v) = \lambda(\Sigma^g(\psi(D_1),\ldots,\psi(D_g))).$$

By Lemmata 6.2 and 6.1 we get

$$\begin{aligned} h(\iota(v)) &\le \log g + h(\lambda) + h_{\mathbf{P}^g}\bigl(\Sigma^g(\psi(D_1),\ldots,\psi(D_g))\bigr) \\ &\le \log g + h(\lambda) + g\log 2 + \sum_{i=1}^g h_{\mathbf{P}^1}(\psi(D_i)). \end{aligned} \tag{6.1}$$

This shows that bounding $h(\iota(v))$ essentially comes down to bounding the $h_{\mathbf{P}^1}(\psi(D_i))$; we will study these in § 6.2 below.

As in § 3.3, we let $P_\iota \in \mathbf{Q}[x]$ denote the monic polynomial defining $\iota(J[\mathfrak{m}]_{\mathbf{Q}})$ in $\mathbf{A}^1_{\mathbf{Q}}$. We let $h(P_\iota)$ denote the maximum of the logarithmic heights of the coefficients of P_ι. Since the coefficients of P_ι are the (dehomogenised) elementary symmetric polynomials in the $\iota(v)$ for $v \in J[\mathfrak{m}](\overline{\mathbf{Q}})$, a second application of Lemma 6.1 yields

$$\begin{aligned} h(P_\iota) &\le \deg J[\mathfrak{m}]\log 2 + \sum_{D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} h(\iota(c_{\mathfrak{m}}(D))) \\ &\le \deg J[\mathfrak{m}]\log 2 + \deg J[\mathfrak{m}]\,(g\log 2 + \log g + h(\lambda)) \\ &\qquad + \sum_{D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} \sum_{i=1}^g h_{\mathbf{P}^1}(\psi(D_i)) \\ &= \deg J[\mathfrak{m}]\bigl((g+1)\log 2 + \log g + h(\lambda)\bigr) + \sum_{D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} \sum_{i=1}^g h_{\mathbf{P}^1}(\psi(D_i)). \end{aligned} \tag{6.2}$$

Finally, we give bounds on the heights of the polynomials defining the addition and scalar multiplication in terms of the $h(\iota(v))$. Here we follow [17, § 14.5]. We put $r = \deg P_\iota = \deg J[\mathfrak{m}]$, and we write

$$S = \sum_{i,j=0}^{r-1} s_{i,j} x_1^i x_2^j \in \mathbf{Q}[x_1, x_2]/(P_\iota(x_1), P_\iota(x_2)),$$

$$M_a = \sum_{i=0}^{r-1} m_i^a x^i \in \mathbf{Q}[x]/(P_\iota) \quad \text{for all } a \in \mathbf{F}^\times.$$

Then we have by definition

$$\sum_{i,j=0}^{r-1} s_{i,j} \iota(v)^i \iota(w)^j = \iota(v+w) \quad \text{for all } v, w \in J[\mathfrak{m}](\overline{\mathbf{Q}})$$

and

$$\sum_{i=0}^{r-1} m_i^a \iota(v)^i = \iota(av) \quad \text{for all } v \in J[\mathfrak{m}](\overline{\mathbf{Q}}), a \in \mathbf{F}.$$

Lemma 6.3 now implies

$$\max_{i,j} h(s_{i,j}) \le 2r^2(r-1)^2 \max_v h(\iota(v)) + r^2 \log r^2$$

and

$$\max_i h(m_i^a) \le 2r(r-1) \max_v h(\iota(v)) + r \log r \quad \text{for all } a \in \mathbf{F}.$$

6.2. Relating heights to Arakelov intersection numbers

We now study the $h_{\mathbf{P}^1}(\psi(D_i))$ in more detail. Let K be any number field with the following properties:

(1) all the D_i for $D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})$ are K-rational;

(2) $X_{\mathbf{Q}} \times \operatorname{Spec} K$ has a regular and semi-stable model

$$\pi\colon \mathcal{X} \to \operatorname{Spec} \mathbf{Z}_K.$$

Then there exists a morphism

$$\phi\colon \mathcal{X}_\psi \to \mathcal{X},$$

obtained by successively blowing up in closed points, such that ψ extends to a morphism

$$\tilde{\psi}\colon \mathcal{X}_\psi \to \mathbf{P}^1_{\mathbf{Z}_K}.$$

The arithmetic surface $\mathcal{X}_\psi$ is regular, but not necessarily semi-stable (if we blow up double points of the fibres, we get exceptional divisors of multiplicity greater than 1). By taking the Zariski closure, we extend the points $D_i \in X(K)$ and the divisors $D \in D_{\mathfrak{m}}(K)$ to Cartier divisors on $\mathcal{X}$.

We now temporarily assume that $\psi(D_i) \neq \infty$; however, the height bound that we are going to deduce will also hold in the case $\psi(D_i) = \infty$, for reasons that will be indicated below. By definition, we have

$$h_{\mathbf{P}^1}(\psi(D_i)) = \frac{1}{[K : \mathbf{Q}]} \sum_v \log \max\{1, |\psi(D_i)|_v\},$$

where v runs over all the places of K. Now we note that for each finite place v of K, we have the equality

$$\log \max\{1, |\psi(D_i)|_v\} = (\log \# k_v)(\tilde{\psi} \circ D_i \,.\, \infty)_v, \tag{6.3}$$

where k_v is the residue field of v and $(\tilde{\psi} \circ D_i \,.\, \infty)_v$ denotes the local intersection number of the sections $\tilde{\psi} \circ D_i$ and ∞ of $\mathbf{P}^1_{\mathbf{Z}_K}$ at v. Furthermore, from the elementary inequality

$$\max\{1, t\} \le \sqrt{1 + t^2} \quad (t \ge 0)$$

it follows that for each infinite place v of K we have the inequality

$$\log \max\{1, |\psi(D_i)|_v\} \le [K_v : \mathbf{R}] \left(\frac{1}{2} - 2\pi \operatorname{gr}_{\mathbf{P}^1(K_v)}(\psi(D_i)_v, \infty) \right). \tag{6.4}$$

Here $\operatorname{gr}_{\mathbf{P}^1}$ is the Green function for the Fubini–Study $(1,1)$-form on $\mathbf{P}^1(K_v)$ as in § III.1.1:

$$\operatorname{gr}_{\mathbf{P}^1(K_v)}(z, \infty) = \frac{1}{4\pi} - \frac{1}{4\pi} \log(1 + |z|^2). \tag{6.5}$$

It follows from (6.3) and (6.4) that the height of the point $\psi(D_i) \in \mathbf{P}^1(\overline{\mathbf{Q}})$ can be bounded in terms of the degree of the metrised line bundle $(\tilde{\psi} \circ D_i)^* \mathcal{O}_{\mathbf{P}^1}(\infty)$ on $\operatorname{Spec} \mathbf{Z}_K$ as follows:

$$h_{\mathbf{P}^1}(\psi(D_i)) \le \frac{1}{[K : \mathbf{Q}]} \deg(\tilde{\psi} \circ D_i)^* \mathcal{O}_{\mathbf{P}^1}(\infty) + \frac{1}{2}. \tag{6.6}$$

Here the line bundle $\mathcal{O}_{\mathbf{P}^1}(\infty)$ on $\mathbf{P}^1_{K_v}$ is endowed with the metric defined by

$$\log |1|_{\mathcal{O}_{\mathbf{P}^1}(\infty)}(z) = 2\pi \operatorname{gr}_{\mathbf{P}^1}(z, \infty).$$

The inequality (6.6) is also valid in the case where $\psi(D_i) = \infty$, and is in fact an equality in this case; this can be seen by computing the degree via the global section z of $\mathcal{O}_{\mathbf{P}^1}(\infty)$ (which vanishes only along 0) instead of the section 1. This means that we can now dispense with our temporary assumption that $\psi(D_i) \neq \infty$.

On each of the Riemann surfaces

$$\mathfrak{X}_v = X(K_v),$$

we define a smooth real-valued function ϕ_v by

$$\begin{aligned} \phi_v(x) &= \log |1|_{\mathcal{O}_{\mathfrak{X}_v}(\psi^{-1}\infty)}(x) - \log |1|_{\mathcal{O}_{\mathbf{P}^1}(\infty)}(\psi(x)) \\ &= 2\pi \operatorname{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(x, \psi^{-1}\infty) - 2\pi \operatorname{gr}_{\mathbf{P}^1}(\psi(x), \infty), \end{aligned}$$

where $\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}$ denotes the canonical Green function of the Riemann surface $\mathfrak{X}_v$. Then ϕ_v satisfies the differential equation

$$\begin{aligned}2i\partial\bar\partial\phi_v &= 2\pi(\delta_{\psi^{-1}\infty} - (\deg\psi)\mu^{\mathrm{can}}_{\mathfrak{X}_v}) - 2\pi(\delta_{\psi^{-1}\infty} - \psi^*\mu_{\mathbf{P}^1})\\ &= 2\pi\psi^*\mu_{\mathbf{P}^1} - 2\pi(\deg\psi)\mu^{\mathrm{can}}_{\mathfrak{X}_v}\end{aligned}$$

with the normalising condition

$$\int_{\mathfrak{X}_v}\phi_v\mu^{\mathrm{can}}_{\mathfrak{X}_v} = -2\pi\int_{x\in\mathfrak{X}_v}\mathrm{gr}_{\mathbf{P}^1}(\psi(x),\infty)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(x).$$

From the definition of $\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}$ it now follows that we can express ϕ_v as

$$\begin{aligned}\phi_v(x) &= \int_{y\in\mathfrak{X}_v}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(x,y)\bigl(2\pi\psi^*\mu_{\mathbf{P}^1}(y) - 2\pi(\deg\psi)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(y)\bigr)\\ &\qquad - 2\pi\int_{y\in\mathfrak{X}_v}\mathrm{gr}_{\mathbf{P}^1}(\psi(y),\infty)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(y)\\ &= 2\pi\int_{y\in\mathfrak{X}_v}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(x,y)\psi^*\mu_{\mathbf{P}^1}(y) - 2\pi\int_{y\in\mathfrak{X}_v}\mathrm{gr}_{\mathbf{P}^1}(\psi(y),\infty)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(y).\end{aligned}$$

Using the formula (6.5) for $\mathrm{gr}_{\mathbf{P}^1}$, we see that

$$\begin{aligned}\phi_v(x) &= 2\pi\int_{y\in\mathfrak{X}_v}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(x,y)\psi^*\mu_{\mathbf{P}^1}(y) - \frac{1}{2} + \frac{1}{2}\int_{y\in\mathfrak{X}_v}\log(1+|\psi(y)|^2)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(y)\\ &\le 2\pi\deg\psi\sup_{\mathfrak{X}_v\times\mathfrak{X}_v}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v} - \frac{1}{2} + \frac{1}{2}\int_{y\in\mathfrak{X}_v}\log(1+|\psi(y)|^2)\mu^{\mathrm{can}}_{\mathfrak{X}_v}(y).\end{aligned}\tag{6.7}$$

By the definition of ϕ_v, we may rewrite the degree appearing on the right-hand side of (6.6) as

$$\begin{aligned}\deg(\tilde\psi\circ D_i)^*\mathcal{O}_{\mathbf{P}^1}(\infty) &= \deg D_i^*(\tilde\psi^*\mathcal{O}_{\mathbf{P}^1}(\infty))\\ &= \deg D_i^*\mathcal{O}_{\mathcal{X}_\psi}(\tilde\psi^{-1}\infty) + \sum_{v\in K_{\mathrm{inf}}}[K_v:\mathbf{R}]\phi_v(D_{i,v})\\ &= (D_i\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi} + \sum_{v\in K_{\mathrm{inf}}}[K_v:\mathbf{R}]\phi_v(D_{i,v}),\end{aligned}\tag{6.8}$$

where $\mathcal{O}_{\mathcal{X}_\psi}(\tilde\psi^{-1}\infty)$ is metrised in the standard way and $(D_i\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi}$ denotes the Arakelov intersection product of divisors on the regular arithmetic surface $\mathcal{X}_\psi$. Combining (6.6), (6.8), (6.7) and the fact that all the $\mathfrak{X}_v$ are isomorphic to the Riemann surface

$$\mathfrak{X} = X(\mathbf{C}),$$

we now deduce the following bound for the height of $\psi(D_i)$:

$$\begin{aligned}h_{\mathbf{P}^1}(\psi(D_i)) \le \frac{1}{[K:\mathbf{Q}]}&(D_i\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi} + 2\pi\deg\psi\sup_{\mathfrak{X}\times\mathfrak{X}}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}}\\ &+ \frac{1}{2}\int_{y\in\mathfrak{X}}\log(1+|\psi(y)|^2)\mu^{\mathrm{can}}_{\mathfrak{X}}(y).\end{aligned}$$

Substituting this in the bound (6.2) for $h(P_\iota)$ and simplifying, we get

$$\begin{aligned} h(P_\iota) \le \deg J[\mathfrak{m}]\biggl(&(g+1)\log 2 + \log g + h(\lambda) + 2\pi g \deg\psi \sup_{\mathfrak{X}\times\mathfrak{X}} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}} \\ &+ \frac{g}{2}\int_{\mathfrak{X}} \log(1+|\psi|^2)\mu_{\mathfrak{X}}^{\mathrm{can}}\biggr) \\ &+ \frac{1}{[K:\mathbf{Q}]}\sum_{D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} (D\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi}. \end{aligned} \tag{6.9}$$

6.3. Specialisation to our choice of ψ

We recall that we have chosen ψ of the form

$$\psi = \alpha/\beta \quad \text{with } \alpha,\beta\in \mathrm{S}_w^{\mathrm{int}}(\Gamma_1(n')).$$

We view α and β as rational sections of the line bundle $\omega^{\otimes w}(-\text{cusps})$ on $\mathcal{X}$. We write the divisor of α (in the classical sense, i.e. without "infinite" components) as $\operatorname{div}^+\alpha - \operatorname{div}^-\alpha$, where $\operatorname{div}^\pm\alpha$ are effective divisors on $\mathcal{X}$ having no prime divisors in common. Since $\alpha\in \mathrm{S}_w^{\mathrm{int}}(\Gamma_1(n'))$, the support of $\operatorname{div}^-\alpha$ is contained in the set of irreducible components of fibres of $\mathcal{X}$ that do not meet the cusp O. We define $\operatorname{div}^\pm\beta$ similarly, and we do the same for $\operatorname{div}^\pm\tilde\psi$ on $\mathcal{X}_\psi$. Noting that

$$\operatorname{div}\tilde\psi = \phi^{-1}\operatorname{div}\alpha - \phi^{-1}\operatorname{div}\beta$$

we see that

$$\begin{aligned} \tilde\psi^{-1}\infty &= \operatorname{div}^-\tilde\psi \\ &\le \phi^{-1}(\operatorname{div}^+\beta + \operatorname{div}^-\alpha). \end{aligned} \tag{6.10}$$

We put any admissible metric $|\ \ |_\omega$ on the line bundle ω on the Riemann surface $\mathfrak{X}$. This also gives an admissible metric $|\ \ |_{\omega^{\otimes w}}$ on $\omega^{\otimes w}$. Multiplication by β gives an isomorphism

$$\mathcal{O}_{\mathcal{X}}\biggl(\operatorname{div}^+\beta - \operatorname{div}^-\beta + \sum_{v\in K_{\mathrm{inf}}} a_\beta \mathfrak{X}_v\biggr) \xrightarrow{\sim} \omega^{\otimes w}(-\text{cusps}) \tag{6.11}$$

of admissible line bundles on $\mathcal{X}$, where

$$a_\beta = -\int_{\mathfrak{X}} \log|\beta|_{\omega^{\otimes w}}\mu_{\mathfrak{X}}^{\mathrm{can}}.$$

Now let D be an element of $D_{\mathfrak{m}}(\overline{\mathbf{Q}})$. By the isomorphism (6.11), the inequality (6.10) and the projection formula for ϕ, the intersection number $(D\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi}$ occurring in (6.9) can be bounded as

$$\begin{aligned} (D\,.\,\tilde\psi^{-1}\infty)_{\mathcal{X}_\psi} &\le (D\,.\,\phi^{-1}(\operatorname{div}^+\beta+\operatorname{div}^-\alpha))_{\mathcal{X}_\psi} \\ &= (D\,.\,\phi^*(\omega^{\otimes w}(\operatorname{div}^-\alpha + \operatorname{div}^-\beta - \text{cusps})))_{\mathcal{X}_\psi} - g[K:\mathbf{Q}]a_\beta \\ &= (D\,.\,\omega^{\otimes w}(\operatorname{div}^-\alpha+\operatorname{div}^-\beta - \text{cusps}))_{\mathcal{X}} - g[K:\mathbf{Q}]a_\beta, \end{aligned} \tag{6.12}$$

where the divisor D in the last expression is to be interpreted as the Zariski closure in $\mathcal{X}$ of the divisor D on X_K. We write

$$\operatorname{div}\alpha = H_\alpha + \sum_{\mathfrak{p}} \sum_{V\in W_\mathfrak{p}} n_{\mathfrak{p},V} V, \tag{6.13}$$

where H_α is an effective horizontal divisor, the $n_{\mathfrak{p},V}$ are integers, $\mathfrak{p}$ runs over the closed points of $\operatorname{Spec}\mathbf{Z}_K$ and $W_\mathfrak{p}$ is the set of irreducible components of the fibre $\mathcal{X}_{k(\mathfrak{p})}$. In particular, we get

$$\operatorname{div}^+\alpha = H_\alpha + \sum_{\mathfrak{p}} \sum_{\substack{V\in W_\mathfrak{p}\\ n_{\mathfrak{p},V}>0}} n_{\mathfrak{p},V} V \quad\text{and}\quad \operatorname{div}^-\alpha = \sum_{\mathfrak{p}} \sum_{\substack{V\in W_\mathfrak{p}\\ n_{\mathfrak{p},V}<0}} -n_{\mathfrak{p},V} V.$$

Let $\mathfrak{p}$ be a closed point of $\operatorname{Spec}\mathbf{Z}_K$. The $n_{\mathfrak{p},V}$ with $V\in W_\mathfrak{p}$ satisfy the equations

$$\sum_{V'\in W_\mathfrak{p}} n_{\mathfrak{p},V'}(V\,.\,V')_{\mathcal{X}} = b_V \quad \text{for all } V\in W_\mathfrak{p},$$

where

$$b_V = (\omega^{\otimes w}(-\text{cusps}) - H_\alpha\,.\,V)_{\mathcal{X}}.$$

We recall from Section 5 that we have chosen w such that $\omega^{\otimes w}(-\text{cusps})$ has non-negative degree on each irreducible component of each fibre. This implies

$$\sum_{\substack{V\in W_\mathfrak{p}\\ b_V>0}} b_V \le \deg\omega^{\otimes w}(-\text{cusps}).$$

Let p be the residue characteristic of $\mathfrak{p}$, let $e(\mathfrak{p})$ be the ramification index of $\mathfrak{p}$ over p, and let $\mathbf{W}_p$ be the field of fractions of the ring of Witt vectors of $\overline{\mathbf{F}}_p$. Applying Lemma III.4.1, we obtain

$$\max_{V\in W_\mathfrak{p}} n_{\mathfrak{p},V} - \min_{V\in W_\mathfrak{p}} n_{\mathfrak{p},V} \le 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\deg\omega^{\otimes w}(-\text{cusps}),$$

where $\gamma(X_{\mathbf{W}_p})$ is the real number defined in § III.4.2. In particular, since there is at least one V for which $n_{\mathfrak{p},V}\ge 0$ (the one intersecting O), we see that

$$-\min_{V\in W_\mathfrak{p}} n_{\mathfrak{p},V} \le 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\deg\omega^{\otimes w}(-\text{cusps}). \tag{6.14}$$

Taking the sum over all $\mathfrak{p}$, we get the inequality

$$\begin{aligned}(D\,.\operatorname{div}^-\alpha)_{\mathcal{X}} &\le g\sum_{\mathfrak{p}}\Bigl(-\min_{V\in W_\mathfrak{p}} n_{\mathfrak{p},V}\Bigr)\log\#k(\mathfrak{p})\\ &\le 2g\deg\omega^{\otimes w}(-\text{cusps})\sum_{\mathfrak{p}} e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\log\#k(\mathfrak{p}).\end{aligned}$$

Grouping the $\mathfrak{p}$ by their residue characteristics, using the relation

$$\sum_{\mathfrak{p}|p} e(\mathfrak{p}) \log \# k(\mathfrak{p}) = [K:\mathbf{Q}] \log p$$

and using the definition (1.1) of $\gamma(X)$, we obtain

$$(D \,.\, \operatorname{div}^{-} \alpha)_{\mathcal{X}} \leq 2g\gamma(X) \deg \omega^{\otimes w}(-\text{cusps}).$$

The same inequality holds with α replaced by β. Substituting this in (6.12), we conclude that

$$\begin{aligned}(D \,.\, \tilde{\psi}^{-1}\infty)_{\mathcal{X}_\psi} \leq (D \,.\, \omega^{\otimes w}(-\text{cusps}))_{\mathcal{X}} + 4g[K:\mathbf{Q}]\gamma(X) \deg \omega^{\otimes w}(-\text{cusps}) \\ + g[K:\mathbf{Q}] \int_{\mathfrak{X}} \log |\beta|_{\omega^{\otimes w}} \mu_{\mathfrak{X}}^{\text{can}}.\end{aligned}$$

We substitute this into (6.9); we also rewrite the integral occurring in that inequality as

$$\int_{\mathfrak{X}} \log(1+|\psi|^2)\mu_{\mathfrak{X}}^{\text{can}} = \int_{\mathfrak{X}} \log(|\alpha|^2_{\omega^{\otimes w}} + |\beta|^2_{\omega^{\otimes w}})\mu_{\mathfrak{X}}^{\text{can}} - \int_{\mathfrak{X}} \log|\beta|^2_{\omega^{\otimes w}} \mu_{\mathfrak{X}}^{\text{can}}.$$

This gives

$$\begin{aligned}h(P_\iota) \leq \frac{1}{[K:\mathbf{Q}]} &\sum_{D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} (D \,.\, \omega^{\otimes w}(-\text{cusps}))_{\mathcal{X}} \\ &+ \deg J[\mathfrak{m}]\Bigg((g+1)\log 2 + \log g + h(\lambda) + 2\pi g \deg \psi \sup_{\mathfrak{X}\times\mathfrak{X}} \operatorname{gr}_{\mathfrak{X}}^{\text{can}} \\ &+ \frac{g}{2}\int_{\mathfrak{X}} \log(|\alpha|^2_{\omega^{\otimes w}} + |\beta|^2_{\omega^{\otimes w}})\mu_{\mathfrak{X}}^{\text{can}} + 4g\gamma(X)\deg \omega^{\otimes w}(-\text{cusps})\Bigg).\end{aligned}$$

Now let $j_{\mathcal{X}}$ denote the canonical morphism $\mathcal{X} \to \mathbf{P}^1_{\mathbf{Z}_K}$. Multiplication by the discriminant modular form Δ gives an isomorphism

$$\Delta \colon \mathcal{O}_{\mathcal{X}}\Bigg(j^*_{\mathcal{X}}\infty + \sum_{v \in K_{\inf}} a_\Delta \mathfrak{X}_v\Bigg) \xrightarrow{\sim} \omega^{\otimes 12}$$

of admissible line bundles on $\mathcal{X}$, where

$$a_\Delta = -\int_{\mathfrak{X}} \log|\Delta|_{\omega^{\otimes 12}} \mu_{\mathfrak{X}}^{\text{can}}. \tag{6.15}$$

This implies that

$$(D \,.\, \omega)_{\mathcal{X}} = \frac{1}{12}(D \,.\, j^*_{\mathcal{X}}\infty)_{\mathcal{X}} + \frac{g[K:\mathbf{Q}]}{12} a_\Delta.$$

We can now rewrite the above inequality as

$$
\begin{aligned}
h(P_\iota) \le \frac{1}{[K:\mathbf{Q}]} & \sum_{D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} \Big(D \cdot \frac{w}{12} j_{\mathcal{X}}^* \infty - \text{cusps}\Big)_{\mathcal{X}} \\
& + \deg J[\mathfrak{m}] \Bigg((g+1)\log 2 + \log g + h(\lambda) + 2\pi g \deg \psi \sup_{\mathfrak{X}\times\mathfrak{X}} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}} \\
& \qquad + \frac{g}{2}\int_{\mathfrak{X}} \log(|\alpha|^2_{\omega^{\otimes w}} + |\beta|^2_{\omega^{\otimes w}})\mu_{\mathfrak{X}}^{\mathrm{can}} \\
& \qquad - \frac{gw}{12}\int_{\mathfrak{X}} \log|\Delta|_{\omega^{\otimes 12}} \mu_{\mathfrak{X}}^{\mathrm{can}} \\
& \qquad + 4g\gamma(X) \deg \omega^{\otimes w}(-\text{cusps})\Bigg).
\end{aligned} \tag{6.16}
$$

6.4. Bounds on the integrals

We choose a real number $\epsilon \in (0,1)$, we write $B_\infty(\epsilon)$ for the standard disc of area ϵ around the unique cusp ∞ of $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$ as in § II.1.2, and we define Y_0 as the complement of $B_\infty(\epsilon)$ in $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$. Furthermore, we define a compact subset Y of $\mathfrak{X}$ as the inverse image of Y_0 under the map

$$
\begin{aligned}
\Gamma_1(n')\backslash\mathbf{H} &\to \mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H} \\
\Gamma_1(n')z &\mapsto \mathrm{SL}_2(\mathbf{Z})z.
\end{aligned}
$$

Then the complement of Y is the disjoint union of the discs $B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$, with $\mathfrak{c}$ running over the cusps of $\Gamma_1(n')$, and with

$$
\epsilon_{\mathfrak{c}} = m_{\mathfrak{c}}\epsilon,
$$

where $m_{\mathfrak{c}}$ is the ramification index at $\mathfrak{c}$.

Let $\mu_{\mathbf{H}}$ be the $(1,1)$-form on $\mathfrak{X}$ (with singularities at the cusps) induced from the standard volume form on the hyperbolic plane $\mathbf{H}$, and let $F_{\Gamma_1(n')}$ be the function defined on $\Gamma_1(n')\backslash\mathbf{H}$ by

$$
F_{\Gamma_1(n')}(z) = \sum_{f\in B} (\Im z)^2 |f(z)|^2,
$$

where B is any orthonormal basis for the space of holomorphic cusp forms of weight 2 for $\Gamma_1(n')$ with respect to the Petersson inner product. In §§ II.4.2 and II.4.3, we saw that there is an effectively computable real number $C(\epsilon)$ such that

$$
0 \le F_{\Gamma_1(n')}(z) \le \begin{cases} C(\epsilon) & \text{if } z \in Y; \\ C(\epsilon)(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z))^2 \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}}(z)) & \text{if } z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}}), \end{cases}
$$

where $y_{\mathfrak{c}}$ is the function defined in § II.1.2. In particular, this gives an upper bound for $\sup_{\mathfrak{X}} F_{\Gamma_1(n')}$; see also Lemma II.4.1.

In the inequality (6.16), we replace the chosen admissible metric on each of the $\omega^{\otimes i}$ with the Petersson metric $|\ |_{i,\mathrm{Pet}}$ defined in § II.2.1. Outside the zeroes and poles of α, β and Δ, we have the equality

$$\frac{(|\alpha|^2_{\omega^{\otimes w}} + |\beta|^2_{\omega^{\otimes w}})^{1/w}}{|\Delta|^{1/12}_{\omega^{\otimes 12}}} = \frac{(|\alpha|^2_{w,\mathrm{Pet}} + |\beta|^2_{w,\mathrm{Pet}})^{1/w}}{|\Delta|^{1/12}_{12,\mathrm{Pet}}}.$$

From this we see that the value of the expression on the third line in (6.16) is unaffected by the change of metrics. Applying Jensen's inequality on convex functions gives

$$\begin{aligned}\int_{\mathfrak{X}} \log(|\alpha|^2_{w,\mathrm{Pet}} + |\beta|^2_{w,\mathrm{Pet}})\mu^{\mathrm{can}}_{\mathfrak{X}} &\le \log \int_{\mathfrak{X}} (|\alpha|^2_{w,\mathrm{Pet}} + |\beta|^2_{w,\mathrm{Pet}})\mu^{\mathrm{can}}_{\mathfrak{X}} \\ &= \log\Bigl(g^{-1} \int_{\mathfrak{X}} (|\alpha|^2_{w,\mathrm{Pet}} + |\beta|^2_{w,\mathrm{Pet}}) F_{\Gamma_1(n')}\mu_{\mathbf{H}}\Bigr) \\ &\le \log(\langle \alpha, \alpha\rangle_{\Gamma_1(n)} + \langle \beta, \beta\rangle_{\Gamma_1(n')}) \\ &\qquad + \log \sup_{\mathfrak{X}} (g^{-1} F_{\Gamma_1(n')}).\end{aligned}$$

The choice of α and β in Algorithm 5.2 implies that the logarithms of the Petersson norms of α and β are bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$.

It remains to find a lower bound for the integral

$$\int_{\mathfrak{X}} \log |\Delta|_{12,\mathrm{Pet}} \mu^{\mathrm{can}}_{\mathfrak{X}} = g^{-1} \int_{\Gamma_1(n')\backslash\mathbf{H}} \log |\Delta|_{12,\mathrm{Pet}} F_{\Gamma_1(n')}\mu_{\mathbf{H}}.$$

One can check numerically that

$$\log |\Delta|_{12,\mathrm{Pet}}(z) < 0 \quad \text{for all } z \in \mathbf{H}.$$

We therefore obtain

$$\begin{aligned}\int_{\Gamma_1(n')\backslash\mathbf{H}} \log |\Delta|_{12,\mathrm{Pet}} F_{\Gamma_1(n')}\mu_{\mathbf{H}} &\ge \int_Y \log |\Delta|_{12,\mathrm{Pet}} C(\epsilon)\mu_{\mathbf{H}} \\ &\quad + \sum_{\mathfrak{c}} \int_{B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})} \log |\Delta|_{12,\mathrm{Pet}} C(\epsilon)(\epsilon_{\mathfrak{c}} y_{\mathfrak{c}}(z))^2 \exp(4\pi/\epsilon_{\mathfrak{c}} - 4\pi y_{\mathfrak{c}}(z))\mu_{\mathbf{H}}.\end{aligned}$$

Rewriting this as an integral on $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$, using the identity

$$y_{\mathfrak{c}}(z) = \frac{y_\infty(z)}{m_{\mathfrak{c}}} \quad \text{for } z \in B_{\mathfrak{c}}(\epsilon_{\mathfrak{c}})$$

and putting

$$N(n') = [\mathrm{SL}_2(\mathbf{Z}) : \Gamma_1(n')],$$

we get

$$\begin{aligned}\int_{\Gamma_1(n')\backslash\mathbf{H}} \log |\Delta|_{12,\mathrm{Pet}} F_{\Gamma_1(n')}\mu_{\mathbf{H}} &\ge \frac{C(\epsilon)N(n')}{2} \int_{Y_0} \log |\Delta|_{12,\mathrm{Pet}}\mu_{\mathbf{H}} \\ &\quad + C(\epsilon)\epsilon^2 \int_{B_\infty(\epsilon)} \log |\Delta|_{12,\mathrm{Pet}} y_\infty^2 \sum_{\mathfrak{c}} \exp\left(\frac{4\pi}{m_{\mathfrak{c}}}(1/\epsilon - y_\infty)\right)\mu_{\mathbf{H}}.\end{aligned}$$

Because the modular form Δ has a simple zero at ∞, the real-valued function

$$z \mapsto \log \frac{|\Delta(z)|_{12,\mathrm{Pet}}}{|q_\infty(z)|} = 2\pi y_\infty(z) + \log|\Delta(z)|_{12,\mathrm{Pet}}$$

on the disc $B_\infty(\epsilon)$ extends to a superharmonic function on the compactification $\overline{B_\infty(\epsilon)}$. The minimum principle for superharmonic functions implies

$$\log|\Delta(z)|_{12,\mathrm{Pet}} \geq 2\pi/\epsilon - 2\pi y_\infty(z) + \inf_{y_\infty(w)=1/\epsilon} \log|\Delta(w)|_{12,\mathrm{Pet}}.$$

From this we get

$$\begin{aligned}\int_{B_\infty(\epsilon)} &\log|\Delta|_{12,\mathrm{Pet}} y_\infty^2 \sum_{\mathfrak{c}} \exp\left(\frac{4\pi}{m_{\mathfrak{c}}}(1/\epsilon - y_\infty)\right)\mu_{\mathbf{H}} \\ &\geq \sum_{\mathfrak{c}} \int_{1/\epsilon}^\infty \left(\inf_{y_\infty(w)=1/\epsilon} \log|\Delta(w)|_{12,\mathrm{Pet}} + 2\pi/\epsilon - 2\pi y\right) \exp\left(\frac{4\pi}{m_{\mathfrak{c}}}(1/\epsilon - y)\right) dy.\end{aligned}$$

Evaluating the integral using standard methods and noting that $\sum_{\mathfrak{c}} m_{\mathfrak{c}} = N(n')/2$, we get

$$\begin{aligned}\int_{B_\infty(\epsilon)} &\log|\Delta|_{12,\mathrm{Pet}} y_\infty^2 \sum_{\mathfrak{c}} \exp\left(\frac{4\pi}{m_{\mathfrak{c}}}(1/\epsilon - y_\infty)\right)\mu_{\mathbf{H}} \\ &\geq -\frac{1}{8\pi}\sum_{\mathfrak{c}} m_{\mathfrak{c}}^2 + \frac{N(n')}{8\pi} \inf_{y_\infty(z)=1/\epsilon} \log|\Delta|_{12,\mathrm{Pet}}(z).\end{aligned}$$

From the above bounds we conclude that

$$\int_{\mathfrak{X}} \log|\Delta|_{12,\mathrm{Pet}} \mu_{\mathfrak{X}}^{\mathrm{can}} \geq C(\epsilon)\left(\frac{N(n')}{2}\left(\frac{\epsilon^2}{4\pi} + \int_{Y_0}\mu_{\mathbf{H}}\right)\inf_{Y_0}\log|\Delta|_{12,\mathrm{Pet}} - \frac{\epsilon^2}{8\pi}\sum_{\mathfrak{c}} m_{\mathfrak{c}}^2\right).$$

6.5. Bounds on $\mathfrak{m}$-bad prime numbers in terms of cohomology

We will derive a bound for the product $B_{\mathfrak{m}}$ of the $\mathfrak{m}$-bad prime numbers in terms of the first cohomology group of a certain line bundle that can be constructed on X after a suitable base change.

Let x be an element of $J[\mathfrak{m}](\overline{\mathbf{Q}})$, and let $K_x \subset \overline{\mathbf{Q}}$ be the field of definition of x. Then K_x is isomorphic to the residue field of the closed point of $J[\mathfrak{m}]_{\mathbf{Q}}$ corresponding to x. Let T_x be the spectrum of the integral closure of $\mathbf{Z}[1/nl]$ in K_x; this is a finite étale $\mathbf{Z}[1/nl]$-scheme by the fact that $J[\mathfrak{m}]$ is étale over $\mathbf{Z}[1/nl]$. The point $x \in J(K_x)$ defines a line bundle $\mathcal{L}_x$ of degree 0 on X_{K_x}. As in § 3.2, we define d_x as the least $d \geq 0$ such that $x \in J[\mathfrak{m}]_d(\overline{\mathbf{Q}})$. We choose a non-zero global section s of $\mathcal{L}_x(d_x O)$ and define

$$D_x^0 = \operatorname{div} s.$$

Then s is unique up to multiplication by elements of $K_x^\times$, and D_x^0 is independent of the choice of s. By taking the Zariski closure, we extend D_x^0 to a horizontal divisor on the proper smooth curve

$$X_{T_x} = X \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} T_x$$

over T_x.

For any prime number $p \nmid nl$, we write

$$\begin{aligned} X_{T_x,p} &= X_{T_x} \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p \\ &= X \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} (T_x \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p). \end{aligned}$$

The fact that T_x is étale over $\mathbf{Z}[1/nl]$ implies that $T_x \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p$ is the disjoint union of the spectra of the residue fields of T_x that have characteristic p. By the definition of $\mathfrak{m}$-bad prime numbers, a prime number $p \nmid nl$ is $\mathfrak{m}$-bad if and only if

$$\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O)) \neq 0 \quad \text{for some } x \in J[\mathfrak{m}](\overline{\mathbf{Q}}).$$

This gives the implication

$$p \nmid nl \text{ is } \mathfrak{m}\text{-bad} \implies \#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O)) \geq p \quad \text{for some } x \in J[\mathfrak{m}](\overline{\mathbf{Q}}).$$

Whether $\#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O)) \geq p$ depends only on the $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-orbit of $x \in J[\mathfrak{m}](\overline{\mathbf{Q}})$ (equivalently, on the closed point of $J[\mathfrak{m}]_{\mathbf{Q}}$ corresponding to x). This means that

$$p \nmid nl \text{ is } \mathfrak{m}\text{-bad} \implies \prod_{x \in J[\mathfrak{m}](\overline{\mathbf{Q}})} \#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O))^{1/[K_x:\mathbf{Q}]} \geq p.$$

We therefore get

$$\log B_{\mathfrak{m}} \leq \log nl + \sum_{x \in J[\mathfrak{m}](\overline{\mathbf{Q}})} \frac{1}{[K_x : \mathbf{Q}]} \sum_{p \nmid nl \text{ prime}} \log \#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O)).$$

In order to bound the terms $\#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O))$, we will apply a suitable base extension and then choose an effective divisor R_x of degree $g - d_x$ in such a way that $\mathrm{H}^0(X_{T_x}, \mathcal{O}(D_x^0 + R_x - O))$ is still zero.

For a given $\epsilon > 0$, we extend the base T_x as follows. We first suppose $g \geq 2$. Because O is a cusp of X, and because $\Omega_{X/\mathbf{Q}}^{\otimes 12}$ has a non-zero section whose divisor is supported in the cusps (namely the cusp form Δ), the class $[(2g-2)O - \Omega_{X/\mathbf{Q}}]$ in $J(\mathbf{Q})$ is a torsion element by the Manin–Drinfeld theorem; see Drinfeld [29, Theorem 1]. By Lemma III.2.1, we can therefore choose a finite extension K'_x of K_x and an effective divisor R_x of degree $g - d_x$ on $X_{K'_x}$ with the property that the Néron–Tate height of the point $[R_x - (g - d_x)O] \in J(K'_x) \subset J(\overline{\mathbf{Q}})$ satisfies

$$\mathrm{h}_J^{\mathrm{NT}}([R_x - (g - d_x)O]) < (g - d_x)^2 \frac{\Omega^2_{X/\mathbf{Q},\mathrm{a}}}{2g - 2} + \epsilon \tag{6.17}$$

and such that the line bundle $\mathcal{O}(D_x^0 + R_x - O)$ on $X_{K'_x}$ has no non-zero global sections. In the case $g = 1$, we take R_x to be the zero divisor for all $x \neq 0$, and we take R_0 to be any non-zero torsion point; this implies that $\mathrm{h}_{J/\mathbf{Q}}^{\mathrm{NT}}([R_x - (g - d_x)O]) = 0$. We then define T'_x as the spectrum of the integral closure of $\mathbf{Z}[1/nl]$ in K'_x. Then T'_x is a finite locally free T_x-scheme of rank $[K'_x : K_x]$. We write

$$X_{T'_x} = X \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} T'_x,$$

and for every prime number $p \nmid nl$ we write

$$X_{T'_x,p} = X_{T'_x} \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p.$$

In contrast to $T_x \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p$, the finite $\mathbf{F}_p$-scheme $T'_x \times_{\operatorname{Spec}\mathbf{Z}[1/nl]} \operatorname{Spec}\mathbf{F}_p$ is not necessarily reduced. We therefore use the following results on proper smooth curves over Artin rings.

Lemma 6.4. *Let X be a proper smooth curve over a local Artinian ring A such that the special fibre of X is geometrically connected of genus g. For every relative Cartier divisor D on X, we have*

$$\operatorname{length}_A \mathrm{H}^0(X, \mathcal{O}_X(D)) - \operatorname{length}_A \mathrm{H}^1(X, \mathcal{O}_X(D)) = (1 - g + \deg D) \operatorname{length}_A A,$$

where $\deg D$ is the degree of D on the special fibre of X.

Proof. This can be proved starting from the case $D = 0$ by adding and subtracting effective divisors and using the long exact cohomology sequences associated to short exact sequences of the form

$$0 \longrightarrow \mathcal{L}(-E) \longrightarrow \mathcal{L} \longrightarrow i_* i^* \mathcal{L} \longrightarrow 0,$$

where $\mathcal{L}$ is a line bundle, E is an effective divisor and $i\colon E \rightarrowtail X$ is the inclusion map. Alternatively, the lemma can be deduced as a special case of the very general version of the Riemann–Roch theorem given by Berthelot in [101, exposé VIII, théorème 3.6]. □

Lemma 6.5. *Let k be a field, let A be a finite Artinian k-algebra, and let X be a proper smooth curve over A such that the fibres of X are geometrically connected of the same genus g. For every line bundle $\mathcal{L}$ on X that is of degree $g-1$ on the fibres, we have the equality*

$$\dim_k \mathrm{H}^0(X, \mathcal{L}) = \dim_k \mathrm{H}^1(X, \mathcal{L}).$$

Proof. This follows from Lemma 6.4 and the equality

$$\dim_k M = \sum_B [k(B) : k] \operatorname{length}_B (M \otimes_A B)$$

for any finitely generated A-module M, where B runs over the local factors of A and $k(B)$ is the residue field of B. □

We now bound $\#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O))$ as follows. Let p be a prime number not dividing nl. The fact that T'_x is locally free of rank $[K'_x : K_x]$ over T_x implies that

$$\#\mathrm{H}^0(X_{T_x,p}, \mathcal{O}(D_x^0 - O)) = \#\mathrm{H}^0(X_{T'_x,p}, \mathcal{O}(D_x^0 - O))^{1/[K'_x:K_x]}.$$

Since R_x is effective, we have

$$\#\mathrm{H}^0(X_{T'_x,p}, \mathcal{O}(D_x^0 - O)) \le \#\mathrm{H}^0(X_{T'_x,p}, \mathcal{O}(D_x^0 + R_x - O)).$$

It follows from Lemma 6.5 that

$$\#\mathrm{H}^0(X_{T'_x,p}, \mathcal{O}(D_x^0 + R_x - O)) = \#\mathrm{H}^1(X_{T'_x,p}, \mathcal{O}(D_x^0 + R_x - O)).$$

The compatibility of the formation of H^1 with base change implies that

$$\prod_{p\nmid nl \text{ prime}} \#\mathrm{H}^1(X_{T'_x,p}, \mathcal{O}(D_x^0 + R_x - O)) = \#\mathrm{H}^1(X_{T'_x}, \mathcal{O}(D_x^0 + R_x - O)).$$

Combining the above (in)equalities, we get

$$\log B_{\mathfrak{m}} \le \log nl + \sum_{x\in J[\mathfrak{m}](\overline{\mathbf{Q}})} \frac{1}{[K'_x : \mathbf{Q}]} \log \#\mathrm{H}^1(X_{T'_x}, \mathcal{O}(D_x^0 + R_x - O)). \tag{6.18}$$

6.6. Bounds from arithmetic intersection theory

We are now going to use arithmetic intersection theory to bound the right-hand side of the inequality (6.18), as well as certain intersection numbers that we will use to bound the right-hand side of (6.16). For any $\epsilon > 0$, we consider finite extensions K'_x and divisors R_x as in § 6.5. After extending these, we may assume all the K'_x are equal to a number field K over which $J[\mathfrak{m}]_{\mathbf{Q}}$ splits and such that $X_{\mathbf{Q}} \times \operatorname{Spec} K$ has a regular and semi-stable model

$$\pi\colon \mathcal{X} \to \operatorname{Spec} \mathbf{Z}_K.$$

In our notation below, we will identify the admissible line bundle Ω_π with an Arakelov divisor denoted by the same symbol.

Now let x be a non-zero element of $J[\mathfrak{m}](\overline{\mathbf{Q}}) = J[\mathfrak{m}](K)$. We extend the divisor D_x^0 to a section of π by taking the Zariski closure, and we abbreviate

$$D'_x = D_x^0 + R_x.$$

We note that

$$\log \#\mathrm{H}^1(X_{T'_x}, \mathcal{O}(D_x^0 + R_x - O)) \le \log \#\mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x - O)). \tag{6.19}$$

We apply Faltings's arithmetic Riemann–Roch formula from Section III.2 to the admissible line bundle $\mathcal{O}_{\mathcal{X}}(D'_x - O)$ on $\mathcal{X}$, where the metrics at the infinite places are chosen using the correspondence between Arakelov divisors and admissible line bundles as in Section III.2. This gives the equation

$$\deg \lambda_\pi \mathcal{O}_{\mathcal{X}}(D'_x - O) = \frac{1}{2}(D'_x - O \,.\, D'_x - O - \Omega_\pi)_{\mathcal{X}} + \deg \lambda_\pi \mathcal{O}_{\mathcal{X}}. \tag{6.20}$$

As in Section III.2, the left-hand side of (6.20) can be written as

$$\begin{aligned}\deg \lambda_\pi \mathcal{O}_{\mathcal{X}}(D'_x - O) = \deg\bigl[&\det \pi_*\mathcal{O}_{\mathcal{X}}(D'_x - O) \otimes \det \pi_*\Omega_\pi(-D'_x + O)\bigr] \\ &- \log \#\mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x - O)),\end{aligned}$$

where the metric on $\det \pi_* \mathcal{O}_{\mathcal{X}}(D'_x - O) \otimes \det \pi_* \Omega_\pi(-D'_x + O)$ is given by Faltings's axioms.

Next we want to apply the Faltings–Hriljac formula from § III.2.2 to the right-hand side of (6.20). We therefore choose a divisor Ψ on $\mathcal{X}$ that is a rational linear combination of irreducible components of fibres of π above closed points of $\operatorname{Spec} \mathbf{Z}_K$, with the property that

$$(D'_x - gO - \Psi \,.\, E)_{\mathcal{X}} = 0 \quad \text{for every vertical divisor } E.$$

Such a Ψ is unique up to addition of rational multiples of fibres. We can now rewrite the intersection number $(D'_x - O \,.\, D'_x - O - \Omega_\pi)_{\mathcal{X}}$ as

$$\begin{aligned}(D'_x - O \,.\, D'_x - O - \Omega_\pi)_{\mathcal{X}} = (D'_x &- gO - \Psi \,.\, D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}} \\ &+ (g-1)^2 (O \,.\, O)_{\mathcal{X}} - (g-1)(O \,.\, \Omega_\pi)_{\mathcal{X}} \\ &+ (\Psi \,.\, D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}}.\end{aligned} \tag{6.21}$$

Applying the Faltings–Hriljac formula to the first term on the right-hand side gives

$$(D'_x - gO - \Psi \,.\, D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}} = -[K : \mathbf{Q}] \langle [D'_x - gO], [D'_x + (g-2)O - \Omega_\pi] \rangle^{\mathrm{NT}}_{J/\mathbf{Q}},$$

where $[E]$ denotes the class of E in $J(K) \subset J(\overline{\mathbf{Q}})$ for any Arakelov divisor E on $\mathcal{X}$ that has degree 0 on the generic fibre. The fact that the class of $D^0_x - d_x O$ in $J(\overline{\mathbf{Q}})$ is a torsion element, together with the Manin–Drinfeld theorem, implies that

$$\begin{aligned}\langle [D'_x - gO], [D'_x + (g-2)O - \Omega_\pi] \rangle^{\mathrm{NT}}_{J/\mathbf{Q}} &= \langle [R_x - (g - d_x)O], [R_x - (g - d_x)O] \rangle^{\mathrm{NT}}_{J/\mathbf{Q}} \\ &= \mathrm{h}^{\mathrm{NT}}_{J/\mathbf{Q}}([R_x - (g - d_x)O]).\end{aligned}$$

We now turn to the second and third terms on the right-hand side of (6.21). The adjunction formula implies that these can be simplified to

$$(g-1)^2 (O \,.\, O)_{\mathcal{X}} - (g-1)(O \,.\, \Omega_\pi)_{\mathcal{X}} = -g(g-1)(O \,.\, \Omega_\pi)_{\mathcal{X}}.$$

For $\log \# \mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x - O))$ we now get the following expression:

$$\begin{aligned}\log \# \mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x - O)) = \deg\bigl[&\det \pi_* \mathcal{O}_{\mathcal{X}}(D'_x - O) \otimes \det \pi_* \Omega_\pi(-D'_x + O)\bigr] \\ &- \deg \lambda_\pi \mathcal{O}_{\mathcal{X}} + \frac{g(g-1)}{2} (O \,.\, \Omega_\pi)_{\mathcal{X}} \\ &+ \frac{[K : \mathbf{Q}]}{2} \mathrm{h}^{\mathrm{NT}}_{J/\mathbf{Q}}([R_x - (g - d_x)O]) \\ &- \frac{1}{2} (\Psi \,.\, D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}}.\end{aligned} \tag{6.22}$$

A similar computation starting with the admissible line bundle $\mathcal{O}_{\mathcal{X}}(D'_x)$ gives

$$\begin{aligned}(D'_x \,.\, O)_{\mathcal{X}} + \log \# \mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x)) = \deg\bigl[&\det \pi_* \mathcal{O}_{\mathcal{X}}(D'_x) \otimes \det \pi_* \Omega_\pi(-D'_x)\bigr] \\ &- \deg \lambda_\pi \mathcal{O}_{\mathcal{X}} + \frac{g(g-1)}{2} (O \,.\, \Omega_\pi)_{\mathcal{X}} \\ &+ \frac{[K : \mathbf{Q}]}{2} \mathrm{h}^{\mathrm{NT}}_{J/\mathbf{Q}}([R_x - (g - d_x)O]) \\ &- \frac{1}{2} (\Psi \,.\, D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}}.\end{aligned} \tag{6.23}$$

We will now give term-by-term bounds of the right-hand sides of (6.22) and (6.23). For the terms

$$\deg\bigl[\det \pi_*\mathcal{O}_{\mathcal{X}}(D'_x - O) \otimes \det \pi_*\Omega_\pi(-D'_x + O)\bigr]$$

and

$$\deg\bigl[\det \pi_*\mathcal{O}_{\mathcal{X}}(D'_x) \otimes \det \pi_*\Omega_\pi(-D'_x)\bigr]$$

we are going to use the following four lemmata.

Lemma 6.6. *Let $\pi\colon X \to B$ be a proper and flat morphism, with B a Dedekind scheme. For any relative effective Cartier divisor E on X such that $\pi_*\mathcal{O}_X(E)$ is locally free of rank 1, the canonical morphism $\mathcal{O}_B \to \pi_*\mathcal{O}_X(E)$ is an isomorphism.*

Proof. Our assumptions imply that the canonical morphism identifies $\mathcal{O}_B$ with a submodule of $\pi_*\mathcal{O}_X(E)$, and that $(\pi_*\mathcal{O}_X(E))/\mathcal{O}_B$ is a torsion module. This means that for any section s of $\pi_*\mathcal{O}_X(E)$ on an affine open subset $U \subset B$ there exists a non-zero $a \in \mathcal{O}_B(U)$ such that $as \in \mathcal{O}_B(U)$. But then we may view $s = (as)/a$ as the pull-back to $\pi^{-1}U$ of a rational function on U, and since E is horizontal, either s is the zero function or its divisor is effective, i.e. $s \in \mathcal{O}_B(U)$. □

Lemma 6.7. *Let K be a number field, let B be the spectrum of its ring of integers, and let $\pi\colon X \to B$ be a semi-stable arithmetic surface with fibres of genus $g \geq 1$. Let $S\colon B \to X$ be a section of π whose image is a Cartier divisor on X. For any effective horizontal Cartier divisor E on X having degree g on the generic fibre and such that $\pi_*\mathcal{O}_X(E - S) = 0$, the equality*

$$\begin{aligned}\deg\bigl[\det \pi_*\mathcal{O}_X(E - S) \otimes \det \pi_*\Omega_\pi(S - E)\bigr] &= \deg\bigl[\det \pi_*\mathcal{O}_X(E) \otimes \det \pi_*\Omega_\pi(-E)\bigr] \\ &\quad + 2\pi \sum_{v \in K_{\mathrm{inf}}} [K_v : \mathbf{R}]\, \mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(E_v, S_v)\end{aligned} \tag{6.24}$$

holds; here the line bundles between square brackets have been metrised using Serre duality and Faltings's axioms for the metrisation of the determinant of cohomology, and $\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}$ denotes the canonical Green function of the Riemann surface $\mathfrak{X}_v$.

Proof. Since $\pi_*\mathcal{O}_X(E-S)$ vanishes by assumption, $\pi_*\mathcal{O}_X(E)$ is locally free of rank 1 and hence canonically isomorphic to $\mathcal{O}_B$ by the previous lemma. Furthermore, by the (classical) Riemann–Roch formula on the generic fibre and the fact that $\pi_*\Omega_\pi(S-E)$ and $\pi_*\Omega_\pi(-E)$ are locally free, we see that these last two modules are both zero. Therefore the modules $\pi_*\mathcal{O}_X(E-S)$, $\pi_*\mathcal{O}_X(E)$, $\pi_*\Omega_\pi(S-E)$ and $\pi_*\Omega_\pi(-E)$ all have a canonically trivial determinant. In particular, there are canonical isomorphisms between these determinants, giving us the top, bottom and right isomorphisms in the diagram

$$\begin{array}{ccc}\det \pi_*\mathcal{O}_X(E - S) \otimes \det \pi_*\Omega_\pi(S - E) & \xrightarrow{\sim} & \det \pi_*\Omega_\pi(S - E) \\ {\scriptstyle\iota}\big\downarrow & & \big\uparrow{\scriptstyle\sim} \\ \det \pi_*\mathcal{O}_X(E) \otimes \det \pi_*\Omega_\pi(-E) & \xrightarrow{\sim} & \det \pi_*\Omega_\pi(-E).\end{array}$$

There is a unique isomorphism ι making this diagram commutative, which is, however, not an isometry. In fact, the terms of the equality (6.24) involving Green functions arise from the norms of the isomorphisms

$$\begin{aligned}\iota_v\colon \det \mathrm{H}^0(\mathfrak{X}_v, \mathcal{O}_{\mathfrak{X}_v}(E_v - S_v)) \otimes \det \mathrm{H}^0(\mathfrak{X}_v, \Omega^1_{\mathfrak{X}_v}(S_v - E_v))& \\ \xrightarrow{\sim} \det \mathrm{H}^0(\mathfrak{X}_v, \mathcal{O}_{\mathfrak{X}_v}(E_v)) \otimes \det \mathrm{H}^0(\mathfrak{X}_v, \Omega^1_{\mathfrak{X}_v}(E_v))&\end{aligned}$$

induced by ι at the infinite places of K. For every such infinite place v, Faltings's axioms for the metrisation of the determinant of cohomology provide a canonical isometry

$$\begin{aligned}\mathrm{Hom}_{\bar{K}_v}\bigl(\det \mathrm{H}^0(\mathfrak{X}_v, \mathcal{O}_{\mathfrak{X}_v}(E_v - S_v)) \otimes \det \mathrm{H}^0(\mathfrak{X}_v, \Omega^1_{\mathfrak{X}_v}(S_v - E_v)),& \\ \det \mathrm{H}^0(\mathfrak{X}_v, \mathcal{O}_{\mathfrak{X}_v}(E_v)) \otimes \det \mathrm{H}^0(\mathfrak{X}_v, \Omega^1_{\mathfrak{X}_v}(E_v))\bigr) \cong \mathcal{O}_{\mathfrak{X}_v}(E_v)[S_v],&\end{aligned}$$

under which the isomorphism ι_v corresponds to the basis element 1 of $\mathcal{O}_{\mathfrak{X}_v}(E_v)[S_v]$, the fibre of the line bundle $\mathcal{O}_{\mathfrak{X}_v}(E_v)$ at the point S_v. The norm of this element can be expressed in terms of the canonical Green function $\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}$ of $\mathfrak{X}_v$ by

$$\log|1|_{\mathcal{O}_{\mathfrak{X}_v}(E_v)}(S_v) = 2\pi\, \mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(E_v, S_v),$$

so that the norm of the isomorphism ι_v equals $\exp\bigl(2\pi\, \mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}_v}(E_v, S_v)\bigr)$. From this the equality (6.24) follows. □

Lemma 6.8. *Let $\mathfrak{X}$ be a compact connected Riemann surface of genus $g \ge 1$, and let $P_1, \ldots, P_g$ be points on $\mathfrak{X}$ such that $\mathrm{H}^0(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}(P_1 + \ldots + P_g)) = \mathbf{C}$. Let $(\alpha_1, \ldots, \alpha_g)$ be any basis for $\mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}})$ such that*

$$\alpha_i \in \mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)) \setminus \mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_i - \cdots - P_g)).$$

For each i let $Q_i^1, \ldots, Q_i^{g+i-2}$ be the zeroes (counted with multiplicities) of α_i, viewed as a global section of $\Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)$. Then the norm of the canonical generator $1 \otimes 1$ of the one-dimensional complex vector space

$$\lambda(\mathcal{O}_{\mathfrak{X}}(P_1 + \cdots + P_g)) = \det \mathrm{H}^0(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}(P_1 + \cdots + P_g)) \otimes \det \mathrm{H}^1(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}(P_1 + \cdots + P_g)),$$

metrised according to Faltings's axioms for the metrisation of the determinant of cohomology, satisfies

$$\begin{aligned}-\log\|1 \otimes 1\|_{\lambda(\mathcal{O}_{\mathfrak{X}}(P_1 + \cdots + P_g))} = -\frac{1}{2}\log\det\langle\alpha_i, \alpha_j\rangle_{i,j=1}^g + 2\pi \sum_{i=1}^g \sum_{j=1}^{g+i-2} \mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}}(P_i, Q_i^j)& \\ + \sum_{i=1}^g \int_{\mathfrak{X}} \bigl(\log|\alpha_i|_{\Omega^1_{\mathfrak{X}}}\bigr)\mu^{\mathrm{can}}_{\mathfrak{X}},&\end{aligned}$$

where $\langle\ ,\ \rangle$ is the inner product on $\mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}})$ defined in § III.1.1.

Proof. For each i, the fact that $(\alpha_1, \ldots, \alpha_i)$ is a basis for $\mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g))$ implies that

$$\det \mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)) = \mathbf{C} \cdot \alpha_1 \wedge \ldots \wedge \alpha_i.$$

Now each of the canonical isomorphisms

$$\begin{aligned} &\det \mathrm{H}^0(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}(P_i + \cdots + P_g)) \otimes \det \mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_i - \cdots - P_g)) \\ &\quad \xrightarrow{\sim} \det \mathrm{H}^0(\mathfrak{X}, \mathcal{O}_{\mathfrak{X}}(P_{i+1} + \cdots + P_g)) \otimes \det \mathrm{H}^0(\mathfrak{X}, \Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)) \\ &\qquad\qquad \otimes \Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)[P_i]^\vee, \end{aligned}$$

where $[P_i]$ denotes the fibre at the point P_i, is an isometry by Faltings's axioms. From this we obtain

$$-\log \|1 \otimes (\alpha_1 \wedge \ldots \wedge \alpha_{i-1})\| = -\log \|1 \otimes (\alpha_1 \wedge \ldots \wedge \alpha_i)\| + h_i(P_i),$$

where h_i is the function defined by

$$h_i = \log |\alpha_i|_{\Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)}.$$

By induction this implies

$$-\log \|1 \otimes 1\| = -\log \|1 \otimes (\alpha_1 \wedge \ldots \wedge \alpha_g)\| + \sum_{i=1}^{g} h_i(P_i).$$

One of Faltings's axioms relates the metric on $\lambda(\mathcal{O}_{\mathfrak{X}})$ to the inner product $\langle\ ,\ \rangle$ as follows:

$$\|1 \otimes (\alpha_1 \wedge \ldots \wedge \alpha_g)\|_{\lambda(\mathcal{O}_{\mathfrak{X}})} = \sqrt{\det\bigl(\langle \alpha_i, \alpha_j \rangle\bigr)_{i,j=1}^{g}}.$$

As for the functions h_i, the admissibility of the metric on $\Omega^1_{\mathfrak{X}}(-P_{i+1} - \cdots - P_g)$ implies that

$$\begin{aligned} \frac{1}{\pi i} \partial\bar{\partial} h_i &= (g + i - 2)\mu^{\mathrm{can}}_{\mathfrak{X}} - \sum_{j=1}^{g+i-2} \delta_{Q_i^j} \\ &= -\sum_{j=1}^{g+i-2} 2i\partial\bar{\partial} \operatorname{gr}^{\mathrm{can}}_{\mathfrak{X}}(\ , Q_i^j), \end{aligned}$$

from which it follows that

$$h_i = 2\pi \sum_{j=1}^{g+i-2} \operatorname{gr}^{\mathrm{can}}_{\mathfrak{X}}(\ , Q_i^j) + \int_{\mathfrak{X}} h_i \mu^{\mathrm{can}}_{\mathfrak{X}}.$$

From the definition of the metric on $\Omega^1_{\mathfrak{X}}(-P_{i+1} - \ldots - P_g)$ we also have the expression

$$\begin{aligned} h_i &= \log |\alpha_i|_{\Omega^1_{\mathfrak{X}}} - \sum_{j=i+1}^{g} \log |1|_{\mathcal{O}_{\mathfrak{X}}(P_j)} \\ &= \log |\alpha_i|_{\Omega^1_{\mathfrak{X}}} - 2\pi \sum_{j=i+1}^{g} \operatorname{gr}^{\mathrm{can}}_{\mathfrak{X}}(\ , P_j). \end{aligned}$$

Substituting this into the integral on the right-hand side of the previous expression for h_i and using the normalisation of the Green function, we get

$$h_i = 2\pi \sum_{j=1}^{g+i-2} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}}(\ \ , Q_i^j) + \int_{\mathfrak{X}} \big(\log |\alpha_i|_{\Omega^1_{\mathfrak{X}}}\big) \mu_{\mathfrak{X}}^{\mathrm{can}},$$

which finishes the proof of the lemma. $\square$

Lemma 6.9. *Let K be a number field, let B be the spectrum of its ring of integers, and let $\pi\colon X \to B$ be a semi-stable arithmetic surface with fibres of genus $g \geq 1$. For any effective horizontal Cartier divisor E on X having degree g on the generic fibre and such that $\pi_*\mathcal{O}_X(E)$ is locally free of rank 1, the inequality*

$$\begin{aligned}&\deg\big[\det \pi_*\mathcal{O}_X(E) \otimes \det \pi_*\Omega_\pi(-E)\big] \\ &\quad\leq g \sum_{v \in K_{\mathrm{inf}}} [K_v : \mathbf{R}] \bigg(3\pi(g-1) \sup_{\mathfrak{X}_v \times \mathfrak{X}_v} \mathrm{gr}_{\mathfrak{X}_v}^{\mathrm{can}} + \sup_{\langle \alpha, \alpha \rangle = 1} \int_{\mathfrak{X}_v} \big(\log |\alpha|_{\Omega^1_{\mathfrak{X}_v}}\big) \mu_{\mathfrak{X}_v}^{\mathrm{can}} \bigg)\end{aligned}$$

holds, where the second supremum is taken over all global holomorphic 1-forms α on $\mathfrak{X}_v$ having norm 1 with respect to the inner product $\langle\ ,\ \rangle$.

Proof. By Lemma 6.6 the sheaf $\pi_*\mathcal{O}_X(E)$ is canonically isomorphic to $\mathcal{O}_B$. Furthermore, $\pi_*\Omega_\pi(-E)$ vanishes by the (classical) Riemann–Roch formula applied to the generic fibre of π. This implies that $\det \pi_*\mathcal{O}_X(E) \otimes \det \pi_*\Omega_\pi(-E)$ is trivialised by the section $1 \otimes 1$. The degree of this line bundle can be expressed as

$$\deg\big[\det \pi_*\mathcal{O}_X(E) \otimes \det \pi_*\Omega_\pi(-E)\big] = -\sum_{v \in K_{\mathrm{inf}}} [K_v : \mathbf{R}] \log \|1 \otimes 1\|_{\lambda(\mathcal{O}_{\mathfrak{X}_v}(E_v))}.$$

We now apply Lemma 6.8, in which we may take for $(\alpha_1, \ldots, \alpha_g)$ an orthonormal basis by means of the Gram–Schmidt process. This gives

$$\begin{aligned}-\log \|1 \otimes 1\|_{\lambda(\mathcal{O}_{\mathfrak{X}_v}(E_v))} &= 2\pi \sum_{i=1}^{g} \sum_{j=1}^{g+i-2} \mathrm{gr}_{\mathfrak{X}_v}^{\mathrm{can}}(P_i, Q_i^j) + \sum_{i=1}^{g} \int_{\mathfrak{X}_v} \big(\log |\alpha_i|_{\Omega^1_{\mathfrak{X}_v}}\big) \mu_{\mathfrak{X}_v}^{\mathrm{can}} \\ &\leq 2\pi \cdot \frac{3g(g-1)}{2} \sup_{\mathfrak{X}_v \times \mathfrak{X}_v} \mathrm{gr}_{\mathfrak{X}_v}^{\mathrm{can}} + g \sup_{\langle \alpha, \alpha \rangle = 1} \int_{\mathfrak{X}_v} \big(\log |\alpha|_{\Omega^1_{\mathfrak{X}_v}}\big) \mu_{\mathfrak{X}_v}^{\mathrm{can}},\end{aligned}$$

from which the bound of the lemma follows. $\square$

For the second term on the right-hand sides of (6.22) and (6.23), we have the following bound. By the definition of the Faltings height in Section III.2, this term can be written as

$$-\deg \lambda_\pi \mathcal{O}_{\mathcal{X}} = -[K : \mathbf{Q}] h_{\mathrm{Faltings}}(\mathrm{X}_1(n)).$$

It follows from a result of Bost for Abelian varieties that if Y is any curve of genus g_Y over a number field, then

$$h_{\mathrm{Faltings}}(Y) \geq -g_Y \log(\pi\sqrt{2});$$

see Autissier [4, théorème 3.1]. In particular, this gives

$$-\deg\lambda_\pi\mathcal{O}_{\mathcal{X}} \le [K:\mathbf{Q}]g\log(\pi\sqrt{2}). \tag{6.25}$$

We next rewrite the intersection number in the third term of (6.22) and (6.23) as

$$(O\,.\,\Omega_\pi)_{\mathcal{X}} = [K:\mathbf{Q}]\deg O^*\Omega_{\mathrm{X}_1(n')/\mathbf{Z}}, \tag{6.26}$$

where the degree on the right-hand side is taken on $\operatorname{Spec}\mathbf{Z}$. We recall that this makes sense since the image of O lies in the smooth locus of $\mathrm{X}_1(n')$.

By our choice of the divisor R_x, the Néron–Tate height occurring in (6.22) and (6.23) can be bounded as

$$\mathrm{h}^{\mathrm{NT}}_{J/\mathbf{Q}}([R_x-(g-d_x)O]) < (g-d_x)^2\frac{\Omega^2_{X/\mathbf{Q},\mathrm{a}}}{2g-2}+\epsilon, \tag{6.27}$$

if $g\ge 2$, where $\Omega^2_{X/\mathbf{Q},\mathrm{a}}$ denotes the self-intersection of the relative dualising sheaf of X in the sense of Zhang. As mentioned before, the left-hand side vanishes if $g=1$.

Finally, we consider the fifth term in (6.22) and (6.23). We write

$$\Psi = \sum_{\mathfrak{p}}\sum_{V\in W_{\mathfrak{p}}} a_{\mathfrak{p},V}V,$$

where $\mathfrak{p}$ runs over the set of closed points of $\operatorname{Spec}\mathbf{Z}_K$ and $W_{\mathfrak{p}}$ is the set of irreducible components of the fibre $\mathcal{X}_{k(\mathfrak{p})}$; then

$$-\frac{1}{2}(\Psi\,.\,D'_x+(g-2)O-\Omega_\pi)_{\mathcal{X}} = -\frac{1}{2}\sum_{\mathfrak{p}}\sum_{V\in W_{\mathfrak{p}}} a_{\mathfrak{p},V}(V\,.\,D'_x+(g-2)O-\Omega_\pi)_{\mathcal{X}}.$$

Let $\mathfrak{p}$ be a closed point of $\operatorname{Spec}\mathbf{Z}_K$, let p be the residue characteristic of $\mathfrak{p}$, and let $e(\mathfrak{p})$ be the ramification index of $\mathfrak{p}$ over p. We write $\mathbf{W}_p$ for the field of fractions of the ring of Witt vectors of $\bar{\mathbf{F}}_p$, and we define a real number $\gamma(X_{\mathbf{W}_p})$ as in § III.4.2. We first assume that $g=1$. Then we get the inequality

$$\begin{aligned} a_{\mathfrak{p},V}(V\,.\,D'_x-O-\Omega_\pi)_{\mathcal{X}} &= a_{\mathfrak{p},V}(V\,.\,D'_x)_{\mathcal{X}} - a_{\mathfrak{p},V}(V\,.\,O+\Omega_\pi)_{\mathcal{X}}\\ &\le \max_{V'} a_{\mathfrak{p},V'}\cdot(V\,.\,D'_x)_{\mathcal{X}} - \min_{V'} a_{\mathfrak{p},V'}\cdot(V\,.\,O+\Omega_\pi)_{\mathcal{X}}. \end{aligned}$$

Taking the sum over all $\mathfrak{p}$ and V and using that both D'_x and $O+\Omega_\pi$ have degree 1 on each fibre, we get

$$(\Psi\,.\,D'_x+(g-2)O-\Omega_\pi)_{\mathcal{X}} \le \sum_{\mathfrak{p}}(\max_V a_{\mathfrak{p},V}-\min_V a_{\mathfrak{p},V})\log\#k(\mathfrak{p}).$$

Similarly, if $g\ge 2$, we get the inequality

$$\begin{aligned} a_{\mathfrak{p},V}(V\,.\,D'_x+(g-2)O-\Omega_\pi)_{\mathcal{X}} &= a_{\mathfrak{p},V}(V\,.\,D'_x+(g-2)O)_{\mathcal{X}} - a_{\mathfrak{p},V}(V\,.\,\Omega_\pi)_{\mathcal{X}}\\ &\le \max_{V'} a_{\mathfrak{p},V'}(V\,.\,D'_x+(g-2)O)_{\mathcal{X}}\\ &\quad - \min_{V'} a_{\mathfrak{p},V'}(V\,.\,\Omega_\pi)_{\mathcal{X}} \end{aligned}$$

for all $\mathfrak{p}$ and V, where we have used that $(V\,.\,\Omega_\pi)_{\mathcal{X}} \geq 0$. This implies

$$(\Psi\,.\,D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}} \leq (2g-2)\sum_{\mathfrak{p}}(\max_V a_{\mathfrak{p},V} - \min_V a_{\mathfrak{p},V})\log\#k(\mathfrak{p}).$$

Next we bound the expressions $\max_V a_{\mathfrak{p},V} - \min_V a_{\mathfrak{p},V}$. The definition of Ψ is equivalent to

$$\sum_{V'} a_{\mathfrak{p},V'}(V\,.\,V')_{\mathcal{X}} = (V\,.\,D'_x - gO) \quad \text{for all } \mathfrak{p} \text{ and } V.$$

Applying Lemma III.4.1 gives

$$\max_V a_{\mathfrak{p},V} - \min_V a_{\mathfrak{p},V} \leq 2e(\mathfrak{p})g\gamma(X_{\mathbf{W}_p}).$$

This implies the upper bound

$$(\Psi\,.\,D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}} \leq g\max\{1, 2g-2\}\sum_{\mathfrak{p}} 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\log\#k(\mathfrak{p}),$$

where $\mathfrak{p}$ runs over the closed points of $\operatorname{Spec}\mathbf{Z}_K$. An entirely analogous argument gives the lower bound

$$(\Psi\,.\,D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}} \geq -g\max\{1, 2g-2\}\sum_{\mathfrak{p}} 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\log\#k(\mathfrak{p}).$$

We conclude that

$$\left|(\Psi\,.\,D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}}\right| \leq g\max\{1, 2g-2\}\sum_{\mathfrak{p}} 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})\log\#k(\mathfrak{p}).$$

Grouping the $\mathfrak{p}$ by their residue characteristics and using (1.1), we get

$$\left|(\Psi\,.\,D'_x + (g-2)O - \Omega_\pi)_{\mathcal{X}}\right| \leq 2[K:\mathbf{Q}]g\max\{1, 2g-2\}\gamma(X). \tag{6.28}$$

We now combine (6.22), Lemmata 6.7 and 6.9, (6.25), (6.26), (6.27) and (6.28) to get the bound

$$\begin{aligned}\frac{\log\#\mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x - O))}{[K:\mathbf{Q}]} \leq{}& (2\pi + 3\pi(g-1))g\sup_{\mathfrak{X}\times\mathfrak{X}}\mathrm{gr}^{\mathrm{can}}_{\mathfrak{X}}\\ &+ g\sup_{\langle\alpha,\alpha\rangle=1}\int_{\mathfrak{X}}\log|\alpha|_{\Omega^1_{\mathfrak{X}}}\mu^{\mathrm{can}}_{\mathfrak{X}}\\ &+ \frac{(g-d_x)^2}{4g-4}\Omega^2_{X/\mathbf{Q},\mathrm{a}} + \epsilon\\ &+ \frac{g(g-1)}{2}\deg O^*\Omega_{\mathrm{X}_1(n')/\mathbf{Z}}\\ &+ g\max\{1, 2g-2\}\gamma(X) + g\log(\pi\sqrt{2}).\end{aligned} \tag{6.29}$$

A similar computation using (6.23), Lemma 6.9, (6.25), (6.26), (6.27) and (6.28) gives

$$
\begin{aligned}
\frac{(D'_x \,.\, O)_{\mathcal{X}} + \log \#\mathrm{H}^1(\mathcal{X}, \mathcal{O}_{\mathcal{X}}(D'_x))}{[K:\mathbf{Q}]} \le {}& 3\pi g(g-1) \sup_{\mathfrak{X}\times\mathfrak{X}} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}} \\
&+ g \sup_{\langle\alpha,\alpha\rangle=1} \int_{\mathfrak{X}} \log|\alpha|_{\Omega^1_{\mathfrak{X}}} \mu_{\mathfrak{X}}^{\mathrm{can}} \\
&+ \frac{(g-d_x)^2}{4g-4} \Omega^2_{X/\mathbf{Q},\mathrm{a}} + \epsilon \qquad (6.30) \\
&+ \frac{g(g-1)}{2} \deg O^* \Omega_{\mathrm{X}_1(n')/\mathbf{Z}} \\
&+ g \max\{1, 2g-2\}\gamma(X) \\
&+ g \log(\pi\sqrt{2}).
\end{aligned}
$$

In both of the above inequalities, the term involving $\Omega^2_{X/\mathbf{Q},\mathrm{a}}$ is to be interpreted as zero if $g = 1$.

6.7. Height bounds: conclusion

In the bound (6.16) for the height of the polynomial P_ι, the term that remains to be bounded from above is the real number M defined by

$$
M = \frac{1}{[K:\mathbf{Q}]} \sum_{D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} \Bigl(D \,.\, \frac{w}{12} j_{\mathcal{X}}^* \infty - \mathrm{cusps}\Bigr)_{\mathcal{X}}.
$$

Here $K \subset \overline{\mathbf{Q}}$ is any number field such that $D_{\mathfrak{m}}$ (or, equivalently, $J[\mathfrak{m}]$) splits over K and such that $X \times \operatorname{Spec} K$ has a regular and semi-stable model

$$
\pi\colon \mathcal{X} \to \operatorname{Spec} \mathbf{Z}_K;
$$

we view the D as horizontal divisors on $\mathcal{X}$.

For each $x \in J[\mathfrak{m}](\overline{\mathbf{Q}})$, we define a positive integer d_x as in § 3.2 and an effective divisor D_x^0 of degree d_x as in § 6.5, and we write

$$
D_x = D_x^0 + (g - d_x)O.
$$

With this notation, we get

$$
M = \frac{1}{[K:\mathbf{Q}]} \sum_{x \in J[\mathfrak{m}](\overline{\mathbf{Q}})} \Bigl(D_x \,.\, \frac{w}{12} j_{\mathcal{X}}^* \infty - \mathrm{cusps}\Bigr)_{\mathcal{X}}.
$$

Let ϵ be a positive real number. After extending K if needed, we choose auxiliary divisors R_x of degree $g - d_x$ satisfying the bound (6.17) on the Néron–Tate height of $[R_x - (g - d_x)O]$ and such that the line bundle $\mathcal{O}(D_x^0 + R_x - O)$ on X_K has no non-zero global sections; see § 6.5. We abbreviate

$$
D'_x = D_x^0 + R_x.
$$

We write

$$H = \frac{w}{12} j_{\mathcal{X}}^{*} \infty - \text{cusps};$$

this is a rational linear combination of the cusps (viewed as horizontal divisors on $\mathcal{X}$) of degree h on the fibres, where

$$h = \deg \omega^{\otimes w}(-\text{cusps}).$$

We fix a vertical divisor Ψ_H with rational coefficients such that

$$(\Psi_H \,.\, V)_{\mathcal{X}} = (H - hO \,.\, V)_{\mathcal{X}} \quad \text{for every vertical divisor } V.$$

For each $x \in J[\mathfrak{m}](\overline{\mathbf{Q}})$, we rewrite $(D_x \,.\, H)_{\mathcal{X}}$ as

$$(D_x \,.\, H)_{\mathcal{X}} = (D_x - gO \,.\, H - hO)_{\mathcal{X}} + g(H \,.\, O)_{\mathcal{X}} + h(D_x \,.\, O)_{\mathcal{X}} - gh(O \,.\, O)_{\mathcal{X}}. \tag{6.31}$$

We consider each term separately. First, it follows from the Faltings–Hriljac formula and the Manin–Drinfeld theorem that

$$\begin{aligned} (D_x - gO \,.\, H - hO)_{\mathcal{X}} &= (D_x^0 - d_x O \,.\, H - hO - \Psi_H)_{\mathcal{X}} + (D_x^0 - d_x O \,.\, \Psi_H)_{\mathcal{X}} \\ &= -[K : \mathbf{Q}]\langle [D_x^0 - d_x O], [H - hO]\rangle_{J/\mathbf{Q}}^{\mathrm{NT}} + (D_x^0 - d_x O \,.\, \Psi_H)_{\mathcal{X}} \\ &= (D_x^0 - d_x O \,.\, \Psi_H)_{\mathcal{X}}. \end{aligned}$$

Second,

$$(D_x \,.\, O)_{\mathcal{X}} = (D'_x \,.\, O)_{\mathcal{X}} - (R_x \,.\, O)_{\mathcal{X}} + (g - d_x)(O \,.\, O)_{\mathcal{X}}.$$

Third,

$$(H \,.\, O)_{\mathcal{X}} = \Bigl(\frac{n'w}{12} - 1\Bigr)(O \,.\, O)_{\mathcal{X}}$$

since the irreducible components of H do not intersect; note that $n'w/12 - 1$ is the multiplicity with which O occurs in H. After a small simplification, we get the equality

$$\begin{aligned} (D_x \,.\, H)_{\mathcal{X}} = h(D'_x \,.\, O)_{\mathcal{X}} - h(R_x \,.\, O)_{\mathcal{X}} + (D_x^0 - d_x O \,.\, \Psi_H)_{\mathcal{X}} \\ + \Bigl(-d_x h + \Bigl(\frac{n'w}{12} - 1\Bigr) g\Bigr)(O \,.\, O)_{\mathcal{X}}. \end{aligned} \tag{6.32}$$

We proceed with bounding the right-hand side. An upper bound for $(D'_x \,.\, O)$ is given by (6.30). The definition of R_x implies that O is not in the support of R_x, so that

$$-(R_x \,.\, O) \le 2\pi(g - d_x) \sup_{\mathfrak{X}\times\mathfrak{X}} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}}.$$

To bound the term $(D_x^0 - d_x O \,.\, \Psi_H)_{\mathcal{X}}$, we write

$$\Psi_H = \sum_{\mathfrak{p}} \sum_{V \in W_{\mathfrak{p}}} a_{\mathfrak{p},V} V,$$

where $\mathfrak{p}$ runs over the closed points of $\operatorname{Spec} \mathbf{Z}_K$ and $W_\mathfrak{p}$ is the set of irreducible components of the fibre $\mathcal{X}_{k(\mathfrak{p})}$. For each $\mathfrak{p}$, the $a_{\mathfrak{p},V}$ satisfy the equations

$$\sum_{V' \in W_\mathfrak{p}} a_{\mathfrak{p},V'} (V \,.\, V')_\mathcal{X} = b_V \quad \text{for all } V \in W_\mathfrak{p},$$

where

$$b_V = (H - hO \,.\, V)_\mathcal{X}.$$

The fact that $\omega^{\otimes w}(-\text{cusps})$ has non-negative degree on each irreducible component of each fibre of $\mathcal{X}$ implies that

$$\sum_{V : b_V > 0} b_V \le h.$$

Lemma III.4.1 therefore gives

$$\max_{V \in W_\mathfrak{p}} a_{\mathfrak{p},V} - \min_{V \in W_\mathfrak{p}} a_{\mathfrak{p},V} \le 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p})h,$$

where p is the residue characteristic of $\mathfrak{p}$. This implies

$$\begin{aligned}
(D_x^0 - d_x O \,.\, \Psi_H)_\mathcal{X} &\le d_x \sum_\mathfrak{p} \Big(\max_{V \in W_\mathfrak{p}} a_{\mathfrak{p},V} - \min_{V \in W_\mathfrak{p}} a_{\mathfrak{p},V} \Big) \log \# k(\mathfrak{p}) \\
&\le d_x \sum_\mathfrak{p} 2e(\mathfrak{p})\gamma(X_{\mathbf{W}_p}) h \log \# k(\mathfrak{p}) \\
&= 2 d_x [K : \mathbf{Q}] \gamma(X) h,
\end{aligned}$$

where in the last step we have used the definition (1.1) of $\gamma(X)$. As for the last term in (6.32), it follows from

$$0 \le [K : \mathbf{Q}] \deg O^* \Omega_{\mathrm{X}_1(n')/\mathbf{Z}} = -(O \,.\, O)_\mathcal{X}$$

and the inequalities $n' \ge 5$, and $w \ge 3$ that

$$\begin{aligned}
\Big(-d_x h + \Big(\frac{n'w}{12} - 1\Big) g\Big)(O \,.\, O)_\mathcal{X} &\le [K : \mathbf{Q}] \max\Big\{0, d_x h - \Big(\frac{n'w}{12} - 1\Big) g\Big\} \\
&\qquad \cdot \deg O^* \Omega_{\mathrm{X}_1(n')/\mathbf{Z}} \\
&\le [K : \mathbf{Q}] d_x h \deg O^* \Omega_{\mathrm{X}_1(n')/\mathbf{Z}}.
\end{aligned}$$

Inserting the above bounds into (6.32), using the inequality $0 \le d_x \le g$ for all x and the fact that $\epsilon > 0$ can be chosen arbitrarily small, and rearranging, we get

$$\begin{aligned}
M \le gh \deg J[\mathfrak{m}] \bigg(&(2\pi + 3\pi(g-1)) \sup_{\mathfrak{X} \times \mathfrak{X}} \mathrm{gr}_\mathfrak{X}^{\mathrm{can}} + \sup_{\langle \alpha, \alpha \rangle = 1} \int_\mathfrak{X} \log |\alpha|_{\Omega^1_\mathfrak{X}} \mu_\mathfrak{X}^{\mathrm{can}} \\
&+ \frac{g}{4g-4} \Omega^2_{X/\mathbf{Q},\mathrm{a}} + \frac{g+1}{2} \deg O^* \Omega_{\mathrm{X}_1(n')/\mathbf{Z}} \\
&+ \max\{3, 2g\} \gamma(X) + \log(\pi\sqrt{2}) \bigg).
\end{aligned}$$

It is possible to write down an upper bound in terms of n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$, using the results of §§ III.3.5, III.3.7, III.5.1 and III.5.2, although for the sake of brevity we will not give such a bound explicitly. Finally, combining this with (6.16) and the results of § 6.4 gives the desired bound on $h(P_\iota)$.

6.8. Bounds on $\mathfrak{m}$-bad prime numbers: conclusion

From (6.18), (6.19), (6.29), the fact that $d_x \geq 0$ for all x, and the fact that $\epsilon > 0$ can be chosen arbitrarily small, we conclude that

$$\begin{aligned}\log B_{\mathfrak{m}} \leq \log nl + g \deg J[\mathfrak{m}] \biggl(&(2\pi + 3\pi(g-1)) \sup_{\mathfrak{X}\times\mathfrak{X}} \mathrm{gr}_{\mathfrak{X}}^{\mathrm{can}} + \sup_{\langle\alpha,\alpha\rangle=1} \int_{\mathfrak{X}} \log|\alpha|_{\Omega^1_{\mathfrak{X}}} \mu_{\mathfrak{X}}^{\mathrm{can}} \\ &+ \frac{g}{4g-4}\Omega^2_{X/\mathbf{Q},\mathrm{a}} + \frac{g-1}{2}\deg O^*\Omega_{\mathrm{X}_1(n')/\mathbf{Z}} \\ &+ \max\{1, 2g-2\}\gamma(X) + \log(\pi\sqrt{2})\biggr).\end{aligned}$$

As in § 6.7, we can bound the right-hand side from above in terms of n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$, using the results of §§ III.3.5, III.3.7, III.5.1 and III.5.2. We do not give such a bound explicitly.

6.9. Bounds on $(\mathfrak{m}, \psi)$-bad prime numbers

As in § 6.3, we write H_α for the horizontal part of the divisor of the form α as in (6.13), and similarly for β. By construction, H_α and H_β do not have any irreducible components in common.

Let p be a prime number not dividing nl. Then p is $(\mathfrak{m}, \psi)$-good if and only if p is $\mathfrak{m}$-good and H_α and H_β do not intersect on the fibre $X_{\mathbf{F}_p}$. Let K be any number field such that $X \times \operatorname{Spec} K$ has a regular and semi-stable model $\mathcal{X}$ over $\operatorname{Spec} \mathbf{Z}_K$. Viewing H_α and H_β as divisors on $\mathcal{X}$, we see that to find a bound for $\log B_{\mathfrak{m},\psi}$ that is polynomial in n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$, it suffices to find such an upper bound for the Arakelov intersection number $\frac{1}{[K:\mathbf{Q}]}(H_\alpha \,.\, H_\beta)_{\mathcal{X}}$. (There is a difference at the infinite places, given by Green functions, but we do not worry about these since we already have bounds for them.) The upper bound on $\log B_{\mathfrak{m},\psi}$ is only needed to prove the desired bound on the expected running time. We therefore do not need to know an actual upper bound on this intersection number, and we permit ourselves to omit some details. We justify our brevity by noting that all the ideas involved have already been explained.

We start by writing

$$(H_\alpha \,.\, H_\beta)_{\mathcal{X}} = (H_\alpha - hO \,.\, H_\beta - hO)_{\mathcal{X}} + h(H_\alpha \,.\, O)_{\mathcal{X}} + h(H_\beta \,.\, O)_{\mathcal{X}} - h^2(O \,.\, O)_{\mathcal{X}}.$$

To bound the first term, we take a vertical divisor Ψ with rational coefficients such that $(H_\alpha - hO - \Psi \,.\, V)_{\mathcal{X}} = 0$ for all vertical divisors V; compare § 6.6. We then apply the Faltings–Hriljac formula and the fact that $\omega^{\otimes w}(-\mathrm{cusps} - hO)$ has Néron–Tate height zero. For the second term, we put any admissible metric $|\ \ |_\omega$ on the line bundle ω on $\mathfrak{X}$ as in § 6.3, and we use that multiplication by α gives an isomorphism

$$\mathcal{O}_{\mathcal{X}}\biggl(H_\alpha + \sum_{\mathfrak{p}} \sum_{V\in W_{\mathfrak{p}}} n_{\mathfrak{p},V} V + \sum_{v\in K_{\mathrm{inf}}} a_\alpha \mathfrak{X}_v\biggr) \xrightarrow{\sim} \omega^{\otimes w}(-\mathrm{cusps}).$$

Here $\mathfrak{p}$ runs over the closed points of $\operatorname{Spec} \mathbf{Z}_K$, $W_{\mathfrak{p}}$ is the set of irreducible components of the fibre $\mathcal{X}_{\mathfrak{p}}$, the $n_{\mathfrak{p},V}$ are the integers defined in (6.13), and

$$a_\alpha = -\int_{\mathfrak{X}} \log|\alpha|_{\omega^{\otimes w}} \mu_{\mathfrak{X}}^{\mathrm{can}}.$$

This leads to

$$
\begin{aligned}
(H_\alpha \,.\, O)_{\mathcal{X}} &= (\omega^{\otimes w}(-\text{cusps}) \,.\, O)_{\mathcal{X}} - \sum_{\mathfrak{p}} \sum_{V\in W_{\mathfrak{p}}} n_{\mathfrak{p},V}(V \,.\, O)_{\mathcal{X}} - [K:\mathbf{Q}]a_\alpha \\
&\le w(\omega \,.\, O)_{\mathcal{X}} - (O \,.\, O)_{\mathcal{X}} - \sum_{\mathfrak{p}} \min_{V\in W_{\mathfrak{p}}} n_{\mathfrak{p},V} \log \# k(\mathfrak{p}) - [K:\mathbf{Q}]a_\alpha \\
&\quad + 2\pi \sum_{v\in K_{\mathrm{inf}}} \sum_{Q_v\ne O_v} \mathrm{gr}_{\mathfrak{X}_v}(Q_v, O_v),
\end{aligned}
$$

where Q_v runs over the cusps of $\mathfrak{X}_v$ other than O_v. A lower bound for the $n_{\mathfrak{p},V}$ was found in (6.14). We also know upper bounds for the Green functions $\mathrm{gr}_{\mathfrak{X}_v}$. Next we note as in § 6.3 that

$$
(O \,.\, \omega)_{\mathcal{X}} = \frac{1}{12}(O \,.\, j_{\mathcal{X}}^*\infty)_{\mathcal{X}} + \frac{[K:\mathbf{Q}]}{12}a_\Delta,
$$

where a_Δ is defined by (6.15). In the same way as in § 6.4, the resulting expression

$$
\int_{\mathfrak{X}} \log|\alpha|_{\omega^{\otimes w}}\mu_{\mathfrak{X}}^{\mathrm{can}} - \frac{w}{12}\int_{\mathfrak{X}} \log|\Delta|_{\omega^{\otimes 12}}\mu_{\mathfrak{X}}^{\mathrm{can}}
$$

can be bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\log B_{\mathfrak{m}}$. Finally, the adjunction formula gives

$$
(O \,.\, O)_{\mathcal{X}} = -[K:\mathbf{Q}] \deg O^*\Omega_{\mathrm{X}_1(n')/\mathbf{Z}},
$$

and an upper bound for the right-hand side was found in § III.5.1.

6.10. Bounds on $(\mathfrak{m},\psi,\lambda)$-bad prime numbers

To bound the $(\mathfrak{m},\psi,\lambda)$-bad prime numbers, we write

$$
B_{\mathfrak{m},\psi,\lambda} = B_{\mathfrak{m},\psi}P,
$$

where P is the product of all $(\mathfrak{m},\psi)$-good prime numbers such that the rational map

$$
\lambda_{\mathbf{F}_p}\colon \mathbf{P}^g_{\mathbf{F}_p} \dashrightarrow \mathbf{A}^1_{\mathbf{F}_p}
$$

is not defined on the image of the map

$$
D_{\mathfrak{m}}^{\mathbf{F}_p} \rightarrowtail \mathrm{Sym}^g X_{\mathbf{F}_p} \xrightarrow{\psi_*} \mathrm{Sym}^g \mathbf{P}^1_{\mathbf{F}_p} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{F}_p}.
$$

As in Algorithm 5.3, we let Λ denote a $2\times(g+1)$-matrix with coprime integral coefficients defining λ. We write Λ_1 for the bottom row of Λ; we view Λ_1 as a linear map from $\overline{\mathbf{Q}}^{g+1}$ to $\overline{\mathbf{Q}}$. For every $D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})$, let y_D denote the image of D in $\mathbf{P}^g(\overline{\mathbf{Q}})$. By (6.1) and the estimates in this chapter, the height $h(y_D)$ is bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$. We write K_D for the field of definition of D, and we choose a representative

$$
z_D = (z_{D,0},\ldots,z_{D,g}) \in K_D^{g+1}
$$

of y_D such that one of the $z_{D,i}$ equals 1. Then the $h(z_{D,i})$ are bounded by $h(y_D)$. Furthermore, $\Lambda_1(z_D)$ is a non-zero element of K_D by the construction of λ, and $h(\Lambda_1(z_D))$ is bounded linearly in $h(\lambda)$ and the $h(z_{D,i})$.

Let p be an $(\mathfrak{m},\psi)$-good prime number. If for all $D \in D_{\mathfrak{m}}(\overline{\mathbf{Q}})$ and all finite places v of K_D with residue characteristic p we have

$$|z_{D,i}|_v \le 1 \text{ for all } i \quad \text{and} \quad |\Lambda_1(z_D)|_v = 1,$$

then p is $(\mathfrak{m},\psi,\lambda)$-good. It is not hard to check that this implies

$$\log P \le \sum_{D\in D_{\mathfrak{m}}(\overline{\mathbf{Q}})} \biggl(\sum_{i=0}^{g} h(z_{D,i}) + h(\Lambda_1(z_D))\biggr).$$

We conclude that $\log P$, and hence $\log B_{\mathfrak{m},\psi,\lambda}$, is bounded by a polynomial in n, $\deg J[\mathfrak{m}]$ and $\gamma(X)$.

7. Computing modular Galois representations

In this final section we explain how to compute the Galois representation ρ_f associated to an eigenform f over a finite field $\mathbf{F}$. Let us briefly recall the strategy, which was described in more detail at the beginning of this chapter. We start by checking whether ρ_f is absolutely irreducible. If this is not the case, then ρ_f is Abelian and relatively easy to compute. If ρ_f is absolutely irreducible, we reduce to the problem of computing a certain $\mathbf{F}$-vector space scheme $J[\mathfrak{m}]$. Here we apply Couveignes's idea of approximating $J[\mathfrak{m}]$. We first compute the reductions of $J[\mathfrak{m}]$ modulo sufficiently many prime numbers. Then we reconstruct $J[\mathfrak{m}]$ over $\mathbf{Q}$ from these reductions. Finally, we compute the representation ρ_f from $J[\mathfrak{m}]$.

We start by describing how to compute $J[\mathfrak{m}]$ modulo prime numbers. We place ourselves in the setting of Section 3. We assume furthermore that a closed immersion ι (given by maps ψ and λ) has been chosen as in Section 5. Using the following algorithm, we can compute the reduction modulo p of the closed subscheme $\iota(J[\mathfrak{m}])$ of $\mathbf{P}^1_{\mathbf{Q}}$, as a finite $\mathbf{F}$-vector space scheme over $\mathbf{F}_p$, for prime numbers p that are $(\mathfrak{m},\psi,\lambda)$-good.

Algorithm 7.1 *(Compute $J[\mathfrak{m}]$ modulo p).* Given positive integers n and k, a finite field $\mathbf{F}$ of characteristic l, a surjective ring homomorphism $e\colon \mathbf{T}_1(n) \to \mathbf{F}$, maps ψ and λ as computed by Algorithms 5.2 and 5.3, and a prime number p, this algorithm checks whether p is $(\mathfrak{m},\psi,\lambda)$-good. If this is the case, the algorithm computes the image under ι of the reduction $J[\mathfrak{m}]_{\mathbf{F}_p}$ of $J[\mathfrak{m}]$ as a finite $\mathbf{F}$-vector space scheme over $\mathbf{F}_p$, represented as in § IV.5.3.

1. Using Algorithm 4.1, compute $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$, a splitting field k_p for $J[\mathfrak{m}]_{\mathbf{F}_p}$ and the $\mathbf{F}$-vector space $D^{\mathbf{F}_p}_{\mathfrak{m}}(k_p)$ given by the positive integer $d = \dim_{\mathbf{F}} J[\mathfrak{m}](k_p)$ and a list of pairs $(v, \Gamma(X_{k_p}, \mathcal{L}^{\otimes 2}(-D_{x(v)})))$.
2. If $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{F}_p})$ is not equal to $\operatorname{strat}(J[\mathfrak{m}]_{\mathbf{Q}})$, output "$p$ is $\mathfrak{m}$-bad" and stop.

3. Compute the image of the map

$$\mathbf{F}^d \xrightarrow{\sim} D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p) \xrightarrow{\psi_*} (\operatorname{Sym}^g \mathbf{P}^1)(k_p) \xrightarrow{\sim} \mathbf{P}^g(k_p)$$

as a list of elements of $\mathbf{P}^g(k_p)$ indexed by $\mathbf{F}^d$.

4. If the points of $\mathbf{P}^g(k_p)$ computed in the previous step are not pairwise distinct, output "p is $(\mathfrak{m}, \psi)$-bad" and stop. If λ is undefined at one of these points, output "p is $(\mathfrak{m}, \psi, \lambda)$-bad" and stop.

For each $v \in \mathbf{F}^d$, we now have the element $\zeta(v) \in \mathbf{A}^1(k_p) = k_p$ that is the image of $D_{x(v)}$ under the map

$$\mathbf{F}^d \xrightarrow{\sim} D_{\mathfrak{m}}^{\mathbf{F}_p}(k_p) \xrightarrow{\psi_*} (\operatorname{Sym}^g \mathbf{P}^1)(k_p) \xrightarrow{\sim} \mathbf{P}^g(k_p) \overset{\lambda}{\dashrightarrow} \mathbf{A}^1(k_p).$$

5. Compute the polynomial $P = \prod_{v\in\mathbf{F}^d}(x - \zeta(v))$; this lies in $\mathbf{F}_p[x]$.

6. Find the unique element $S \in \mathbf{F}_p[x_1, x_2]/(P(x_1), P(x_2))$ satisfying $S(\zeta(v), \zeta(w)) = \zeta(v + w)$ for all $v, w \in V$. (This can be done using Lagrange interpolation.)

7. For each $a \in \mathbf{F}$, find the unique element $M_a \in \mathbf{F}_p[x]/(P)$ satisfying $M_a(\zeta(v)) = \zeta(av)$ for all $v \in V$.

8. Output P, S and the M_a for $a \in \mathbf{F}$.

Analysis. It is straightforward to check that the algorithm is correct and that its expected running time is polynomial in n, $\deg J[\mathfrak{m}]$ and p. $\diamond$

We next give the algorithm for computing the vector space scheme $J[\mathfrak{m}]$ over $\mathbf{Q}$.

Algorithm 7.2 *(Compute the vector space scheme $J[\mathfrak{m}]$).* Given positive integers n and k, a finite field $\mathbf{F}$ of characteristic l, a surjective ring homomorphism $e\colon \mathbf{T}_1(n) \to \mathbf{F}$ and maps ψ and λ as computed by Algorithms 5.2 and 5.3, this algorithm computes the image under ι of $J[\mathfrak{m}]$ as a finite $\mathbf{F}$-vector space scheme over $\mathbf{Q}$ represented as in § IV.5.3.

1. Determine the generic stratification type of $J[\mathfrak{m}]$ using Algorithm 4.2.

2. Using Algorithm 5.2, choose a rational function ψ on $X_{\mathbf{Q}}$ (defined over $\mathbf{Q}$) such that

$$\operatorname{Sym}^g \psi\colon \operatorname{Sym}^g X_{\mathbf{Q}} \to \operatorname{Sym}^g \mathbf{P}^1_{\mathbf{Q}} \xrightarrow{\sim} \mathbf{P}^g_{\mathbf{Q}}$$

is a closed immersion on the closed subscheme $D_{\mathfrak{m}}$ of $\operatorname{Sym}^g X_{\mathbf{Q}}$.

3. Using Algorithm 5.3, choose a rational map $\lambda\colon \mathbf{P}^g_{\mathbf{Q}} \dashrightarrow \mathbf{A}^1_{\mathbf{Q}}$, given as a quotient of two linear forms, such that λ is defined and injective on the image of $D_{\mathfrak{m}}$ in $\mathbf{P}^g_{\mathbf{Q}}$.

4. Compute a bound h on the height of $J[\mathfrak{m}]$, as defined in § 3.3; see § 6.1.

5. Put $Q = \emptyset$. For all prime numbers p not dividing nl, in increasing order:

6. Using Algorithm 7.1, check whether p is $(\mathfrak{m}, \psi, \lambda)$-good, and in this case compute $\iota(J[\mathfrak{m}]_{\mathbf{F}_p})$ and replace Q by $Q \sqcup \{p\}$.

7. If $\prod_{p\in Q} p > 2\exp(h)$, go to step 8.

8. Compute $\iota(J[\mathfrak{m}])$ from the $\iota(J[\mathfrak{m}]_{\mathbf{F}_p})$ with $p \in Q$ by lifting the polynomials P, S and M_a for $a \in \mathbf{F}$, which are known modulo p for all the $p \in Q$, to polynomials over $\mathbf{Q}$ with coefficients of minimal height.

Analysis. By the prime number theorem, the prime numbers occurring in Q are at most $O(\log B_{\mathfrak{m},\psi\lambda})$; recall that $B_{\mathfrak{m},\psi,\lambda}$ is the product of all $(\mathfrak{m},\psi,\lambda)$-bad primes. This implies that the expected running time of the algorithm is polynomial in n, $\deg J[\mathfrak{m}]$, $\log B_{\mathfrak{m},\psi,\lambda}$ and h. The condition in step 7 ensures that $\iota(J[\mathfrak{m}])$ can indeed be reconstructed from its reductions modulo the prime numbers in $\mathbf{Q}$. ◇

Finally, we present the end product of this thesis.

Algorithm 7.3 *(Compute the Galois representation associated to an eigenform over a finite field).* Let n and k be positive integers, let l be a prime number, and let f be an eigenform of weight k for $\Gamma_1(n)$ over a finite field $\mathbf{F}$ of characteristic l. Let K_f denote the finite Galois extension of $\mathbf{Q}$ such that the semi-simple representation ρ_f associated to f factors as

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K_f/\mathbf{Q}) \rightarrowtail \mathrm{Aut}_{\mathbf{F}} W_f.$$

Given n, k, $\mathbf{F}$, the eigenvalues $\epsilon(d) \in \mathbf{F}^\times$ of the operators $\langle d\rangle$ for $d \in (\mathbf{Z}/n\mathbf{Z})^\times$ and the eigenvalues $a_p \in \mathbf{F}$ of the Hecke operators T_p for $p \le \frac{k}{12}[\mathrm{SL}_2(\mathbf{Z}) : \{\pm 1\}\Gamma_1(n)]$ prime, this algorithm outputs the representation ρ_f in the form of the following data:

(1) the multiplication table of K_f with respect to some $\mathbf{Q}$-basis $(b_1, \dots, b_r)$ of K_f;

(2) for every $\sigma \in \mathrm{Gal}(K_f/\mathbf{Q})$, the matrix of σ with respect to the basis $(b_1, \dots, b_r)$ and the matrix of $\rho_f(\sigma)$ with respect to a fixed $\mathbf{F}$-basis of W_f.

1. Using the method of § 2.1, check whether ρ_f is absolutely irreducible, and if it is not, output ρ_f and stop.

We now know f is a cusp form; define j, $\tilde{k}$, n', $\tilde{f}$, $e_{\tilde{f}}$, $\mathbf{F}_{\tilde{f}}$, X, J and $\mathfrak{m}_{\tilde{f}}$ as in § 2.2.

2. Compute the Hecke algebra $\mathbf{T}_1(n')$ and the surjective ring homomorphism

$$e_{\tilde{f}}\colon \mathbf{T}_1(n') \to \mathbf{F}_{\tilde{f}}.$$

3. Compute the $\mathbf{F}_{\tilde{f}}$-vector space scheme $J[\mathfrak{m}_{\tilde{f}}]$ using Algorithm 7.2.

4. Using Algorithm IV.5.1, compute a minimal non-trivial $\mathbf{F}_{\tilde{f}}$-vector space scheme V contained in $J[\mathfrak{m}_{\tilde{f}}]$. (This is trivial if $\deg J[\mathfrak{m}_{\tilde{f}}] = \#\mathbf{F}_{\tilde{f}}^2$.)

5. Compute the representation ρ_V from V using Algorithm IV.5.2.

6. Compute the representation ρ_f as

$$\rho_f = \rho_V \otimes_{\mathbf{F}_l} \chi_l^{\otimes -j}$$

using Algorithm IV.5.3.

Analysis. It is straightforward to check that the algorithm is correct, and that its expected running time is bounded by a polynomial in n, k and $\#\mathbf{F}$ in the case where ρ_f is not absolutely irreducible, and by a polynomial in n, k, $\deg J[\mathfrak{m}_{\tilde{f}}]$ and $\gamma(\mathrm{X}_1(n'))$ in the case where ρ_f is absolutely irreducible. ◇

Bibliography

The numbers in italics following each entry refer to the pages of this thesis on which the work is cited.

[1] L. M. Adleman and H. W. Lenstra, Jr., Finding irreducible polynomials over finite fields. In: *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing (Berkeley, CA, 1986)*, 350–355. Association for Computing Machinery, New York, 1986. *103*

[2] С. Ю. Аракелов, Теория пересечений дивизоров на арифметической поверхности. *Известия Академии Наук СССР, серия математическая* **38** (1974), № 6, 1179–1192.
S. Yu. Arakelov, Intersection theory of divisors on an arithmetic surface. *Mathematics of the USSR Izvestiya* **8** (1974), 1167–1180. (English translation.) *73, 74, 78*

[3] A. O. L. Atkin and J. Lehner, Hecke operators on $\Gamma_0(m)$. *Mathematische Annalen* **185** (1970), 134–160. *38*

[4] P. Autissier, Hauteur de Faltings et hauteur de Néron–Tate du diviseur thêta. *Compositio Mathematica* **142** (2006), no. 6, 1451–1458. *200*

[5] A. F. Beardon, *The Geometry of Discrete Groups*. Springer-Verlag, New York, 1983. *31*

[6] K. Belabas, A relative van Hoeij algorithm over number fields. *Journal of Symbolic Computation* **37** (2004), no. 5, 641–668. *156*

[7] K. Belabas, M. van Hoeij, J. Klüners, and A. Steel, Factoring polynomials over global fields. *Journal de Théorie des Nombres de Bordeaux* **21** (2009), no. 1, 15–39. *156*

[8] S. Bosch, W. Lütkebohmert and M. Raynaud, *Néron Models*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **21**. Springer-Verlag, Berlin, 1990. *10*

[9] J. G. Bosman, *Explicit computations with modular Galois representations*. Proefschrift, Universiteit Leiden, 2008. *6*

[10] N. Boston, H. W. Lenstra and K. A. Ribet, Quotients of group rings arising from two-dimensional representations. *C. R. Acad. Sci. Paris* (I) **312** (1991), 323–328. *29*

[11] J. A. Buchmann and H. W. Lenstra, Jr., Approximating rings of integers in number fields. *Journal de Théorie de Nombres de Bordeaux* **6** (1994), 221–260. *162*

[12] K. Buzzard, On level-lowering for mod 2 representations. *Mathematical Research Letters* **7** (2000), no. 1, 95–110. *21*

[13] G. Castelnuovo, Sui multipli di una serie lineare di gruppi di punti appartenente ad una curva algebrica. *Rendiconti del Circolo Matematico di Palermo* **7** (1893), 89–110. (= *Memorie scelte*, 95–113. Zanichelli, Bologna, 1937.) *106*

[14] B. Conrad, Arithmetic moduli of generalized elliptic curves. *Journal de l'Institut de Mathématiques de Jussieu* **6** (2007), no. 2, 209–278. *7, 8, 9, 13, 14, 16*

[15] B. Conrad, *Modular Forms and the Ramanujan Conjecture*. Cambridge University Press, to appear. *19*

[16] J.-M. Couveignes, Linearizing torsion classes in the Picard group of algebraic curves over finite fields. *Journal of Algebra* **321** (2009), 2085–2118. *3, 101, 129*

[17] J.-M. Couveignes and S. J. Edixhoven (with J. G. Bosman, R. S. de Jong and F. Merkl), *Computational aspects of modular forms and Galois representations*. Princeton University Press, to appear. *1, 2, 3, 4, 6, 19, 21, 160, 170, 179, 182, 183*

[18] J. E. Cremona, *Algorithms for modular elliptic curves*. Cambridge University Press, 1997. *152*

[19] A. J. de Jong, Families of curves and alterations. *Annales de l'Institut Fourier* **47** (1997), no. 2, 599–621. *99*

[20] P. Deligne, Formes modulaires et représentations l-adiques. Séminaire Bourbaki, 21e année (1968/1969), exposé 355. Lecture Notes in Mathematics **179**, 139–172. Springer-Verlag, Berlin/Heidelberg/New York, 1971. *19, 39*

[21] P. Deligne, La conjecture de Weil. I. *Publications mathématiques de l'I.H.É.S.* **43** (1973), 273–307. *39*

[22] P. Deligne and D. Mumford, The irreducibility of the space of curves of given genus. *Publications mathématiques de l'I.H.É.S.* **36** (1969), 75–110. *94*

[23] P. Deligne et M. Rapoport, Les schémas de modules de courbes elliptiques. In: P. Deligne and W. Kuyk (editors), *Modular Functions of One Variable II* (Proceedings of the International Summer School, University of Antwerp, 1972). Lecture Notes in Mathematics **349**, 143–316. Springer-Verlag, Berlin/Heidelberg, 1973. *7, 16, 36, 37*

[24] P. Deligne et J.-P. Serre, Formes modulaires de poids 1. *Annales scientifiques de l'É.N.S. (4e série)* **7** (1974), no. 4, 507–530. *19, 21*

[25] F. Diamond and J. Im, Modular forms and modular curves. In: V. Kumar Murty (editor), *Seminar on Fermat's Last Theorem* (Fields Institute for Research in Mathematical Sciences, Toronto, ON, 1993–1994), 39–133. CMS Con-

ference Proceedings **17**. American Mathematical Society, Providence, RI, 1995. *7, 12, 16*

[26] F. Diamond and J. Shurman, *A First Course in Modular Forms*. Springer-Verlag, Berlin/Heidelberg/New York, 2005. *7*

[27] C. Diem, *On arithmetic and the discrete logarithm problem in class groups of curves*. Habilitationsschrift, Universität Leipzig, 2008. *101, 134*

[28] C. Diem, On the discrete logarithm problem in class groups of curves. To appear in *Mathematics of Computation*. *134*

[29] В. Г. Дринфельд, Две теоремы о модулярных кривых. *Функциональный анализ и его приложения* **7** (1973), № 2, 83–84.
V. G. Drinfeld, Two theorems on modular curves. *Functional Analysis and Applications* **7** (1973), 155–156. (English translation.) *192*

[30] W. Eberly and M. Giesbrecht, Efficient decomposition of associative algebras over finite fields. *Journal of Symbolic Computation* **29** (2000), 441–458. *102*

[31] S. J. Edixhoven, The weight in Serre's conjectures on modular forms. *Inventiones mathematicae* **109** (1992), 563–594. *15, 16, 19, 26, 27, 29, 167*

[32] S. J. Edixhoven, Serre's conjecture. In: G. Cornell, J. H. Silverman and G. Stevens (editors), *Modular Forms and Fermat's Last Theorem* (Boston, MA, August 9–18, 1995), 209–242. Springer-Verlag, Berlin/Heidelberg/New York, 1997. *27*

[33] M. Eichler, Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunction. *Archiv der Mathematik* **5** (1954), 355–366. *19*

[34] A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi, *Higher Transcendental Functions*, Volume I. Bateman Manuscript Project, California Institute of Technology. McGraw-Hill, New York/Toronto/London, 1953. *33, 45, 46, 49, 62, 65, 66, 67*

[35] A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi, *Tables of Integral Transforms*, Volume I. Bateman Manuscript Project, California Institute of Technology. McGraw-Hill, New York/Toronto/London, 1954. *66*

[36] Л. Д. Фаддеев, Разложение по собственным функциям оператора Лапласа на фундаментальной области дискретой группы на плоскости Лобачевского. *Труды Московского Математического Общества* **17** (1967), 323–350.
L. D. Faddeev, Expansion in eigenfunctions of the Laplace operator on the fundamental domain of a discrete group on the Lobačevskiĭ plane. *Transactions of the Moscow Mathematical Society* **17** (1967), 357–386. (English translation.) *43, 44, 55, 62*

[37] G. Faltings, Calculus on arithmetic surfaces. *Annals of Mathematics* (2) **119** (1984), 387–424. *73, 74, 75, 78, 81, 97*

[38] J. D. Fay, Fourier coefficients of the resolvent for a Fuchsian group. *Journal für die reine und angewandte Mathematik* **293/294** (1977), 143–203. *55*

[39] G. Frey and H.-G. Rück, A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves. *Mathematics of Computation* **62** (1994), 865–874. *138*

[40] B. H. Gross, Local Heights on Curves. In: G. Cornell and J. H. Silverman (editors), *Arithmetic Geometry* (Instructional conference, University of Connecticut in Storrs, 1984), Chapter XIV. Springer-Verlag, New York, 1986. *51*

[41] B. H. Gross, A tameness criterion for Galois representations associated to modular forms (mod p). *Duke Mathematical Journal* **61** (1990), no. 2, 445–517. *12, 19, 21, 29*

[42] R. Hartshorne, *Residues and Duality*. Lecture Notes in Mathematics **20**. Springer-Verlag, Berlin/Heidelberg, 1966. *80*

[43] R. Hartshorne, *Algebraic Geometry*. Springer-Verlag, New York, 1977. *105, 106, 107, 113, 114*

[44] E. Hecke, Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung. II. *Mathematische Annalen* **114** (1937), 316–351. (= *Mathematische Werke*, 672–707. Vandenhoeck und Ruprecht, Göttingen, 1959.) *38*

[45] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,\mathbf{R})$, Volume 1. Lecture Notes in Mathematics **548**. Springer-Verlag, Berlin/Heidelberg, 1976. *31*

[46] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,\mathbf{R})$, Volume 2. Lecture Notes in Mathematics **1001**. Springer-Verlag, Berlin/Heidelberg, 1983. *31, 43, 54, 55, 56*

[47] J. Igusa, Kroneckerian model of fields of elliptic modular functions. *American Journal of Mathematics* **81** (1959), 561–577. *19*

[48] H. Iwaniec, Small eigenvalues of Laplacian for $\Gamma_0(N)$. *Acta Arithmetica* **56** (1990), no. 1, 65–82. *39*

[49] H. Iwaniec, *Introduction to the Spectral Theory of Automorphic Forms*. Revista Matemática Iberoamericana, Madrid, 1995. *31, 32, 33, 34, 43, 44, 46, 47, 49, 50*

[50] J. Jorgenson and J. Kramer, Bounding the sup-norm of automorphic forms. *Geometric and Functional Analysis* **14** (2004), no. 6, 1267–1277. *56, 57*

[51] J. Jorgenson and J. Kramer, Bounds on canonical Green's functions. *Compositio Mathematica* **142** (2006), no. 3, 679–700. *83*

[52] N. M. Katz, A result on modular forms in characteristic p. In: J.-P. Serre and D. B. Zagier (editors), *Modular Functions of One Variable V* (Proceedings of the International Conference, University of Bonn, 1976). Lecture Notes in Mathematics **601**, 53–61. Springer-Verlag, Berlin/Heidelberg, 1977. *22*

[53] N. M. Katz and B. Mazur, *Arithmetic Moduli of Elliptic Curves*. Princeton University Press, Princeton, NJ, 1985. *7, 8, 98, 99*

[54] C. Khare and J-P. Wintenberger, Serre's modularity conjecture (I). *Inventiones mathematicae* **178** (2009), no. 3, 485–504. *6, 21, 28*

[55] C. Khare and J-P. Wintenberger, Serre's modularity conjecture (II). *Inventiones mathematicae* **178** (2009), no. 3, 505–586. *6, 28*

[56] K. Khuri-Makdisi, Linear algebra algorithms for divisors on an algebraic curve. *Mathematics of Computation* **73** (2004), no. 245, 333–357.
Available online: `http://arxiv.org/abs/math.NT/0105182`.
3, 105, 106, 107, 108, 121

[57] K. Khuri-Makdisi, Asymptotically fast group operations on Jacobians of general curves. *Mathematics of Computation* **76** (2007), no. 260, 2213–2239.
Available online: `http://arxiv.org/abs/math.NT/0409209`.
Draft version 2: `http://arxiv.org/abs/math.NT/0409209v2`.
3, 102, 105, 106, 108, 109, 122, 178

[58] L. J. P. Kilford and G. Wiese, On the failure of the Gorenstein property for Hecke algebras of prime weight. *Experimental Mathematics* **17** (2008), no. 1, 37–52. *166*

[59] H. H. Kim, Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2. With appendix 1 by D. Ramakrishnan and appendix 2 by Kim and P. Sarnak. *Journal of the A.M.S.* **16** (2002), no. 1, 139–183. *70*

[60] M. Kisin, Modularity of 2-dimensional Galois representations. In: D. Jerison, B. Mazur, T. Mrowka, W. Schmid, R. Stanley and S.-T. Yau (editors), *Current Developments in Mathematics, 2005*, 191–230. International Press, Somerville, MA, 2007. *28*

[61] M. Kisin, Modularity of 2-adic Barsotti–Tate representations. *Inventiones mathematicae* **178** (2009), no. 3, 587–634. *6, 28*

[62] S. Lang, $SL_2(\mathbf{R})$. Addison-Wesley, Reading, MA/London/Amsterdam, 1975. *62*

[63] G. Laumon et L. Moret-Bailly, Champs algébriques. Springer-Verlag, Berlin/Heidelberg, 2000. *8*

[64] R. Lazarsfeld, A sampling of vector bundle techniques in the study of linear series. In: M. Cornalba, X. Gomez-Mont and A. Verjovsky (editors), *Lectures on Riemann Surfaces (Trieste, 1987)*, 500–559. World Scientific Publishing, Teaneck, NJ, 1989. *106*

[65] A. K. Lenstra, Factoring polynomials over algebraic number fields. In: J. A. van Hulzen (editor), *Computer algebra* (London, 1983), 245–254. Lecture Notes in Computer Science **162**. Springer-Verlag, Berlin/Heidelberg, 1983. *156*

[66] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, Factoring polynomials with rational coefficients. *Mathematische Annalen* **261** (1982), no. 4, 515–534. *152, 156*

[67] H. W. Lenstra, Jr., Algorithms in algebraic number theory. *Bulletin of the American Mathematical Society* **26** (1992), no. 2, 211–244. *156*

[68] H. W. Lenstra, Jr., Lattices. In: J. Buhler and P. Stevenhagen (editors), *Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography*. Cambridge University Press, 2008. *152*

[69] W.-C. W. Li, Newforms and functional equations. *Mathematische Annalen* **212** (1975), 285–315. *38*

[70] W.-C. W. Li, L-series of Rankin type and their functional equations. *Mathematische Annalen* **244** (1979), no. 2, 135–166. *39, 40*

[71] R. Livné, On the conductors of mod l Galois representations coming from modular forms. *Journal of Number Theory* **31** (1989), no. 2, 133-141. *27*

[72] H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. *Mathematische Annalen* **121** (1949), 141–183. *42*

[73] Ю. И. Манин, Параболические точки и дзета-функции модулярных кривых. *Известия Академии Наук СССР, серия математическая* **36** (1972), 19–66. Yu. I. Manin, Parabolic points and zeta functions of modular curves. *Mathematics of the USSR Izvestija* **6** (1964), no. 1, 19–64. *152*

[74] A. I. Markushevich, *Theory of Functions of a Complex Variable*. Revised English edition. Chelsea, New York, 1977. *41*

[75] A. Mattuck, Symmetric products and Jacobians. *American Journal of Mathematics* **83** (1961), no. 1, 189–206. *106*

[76] B. Mazur, Modular curves an the Eisenstein ideal. *Publications mathématiques de l'I.H.É.S.* **47** (1977), 33–186. *29*

[77] B. Mazur and K. A. Ribet, Two-dimensional representations in the arithmetic of modular curves. In: *Courbes modulaires et courbes de Shimura*, 215–255. Astérisque **196–197**. Société mathématique de France, Paris, 1992. *29*

[78] F. G. Mehler, Über eine mit den Kugel- und Cylinderfunctionen verwandte Function und ihre Anwendung in der Theorie der Electricitätsvertheilung. *Mathematische Annalen* **18** (1881), 161–194. *33*

[79] L. Merel, Universal Fourier expansions of modular forms. In: G. Frey (editor), *On Artin's conjecture for odd 2-dimensional representations*, 59–94. Lecture Notes in Mathematics **1585**. Springer-Verlag, Berlin/Heidelberg, 1994. *152*

[80] T. Miyake, *Modular Forms*. Springer-Verlag, Berlin/Heidelberg, 1989. *10, 38*

[81] D. Mumford, Varieties defined by quadratic equations. With an appendix by G. Kempf. In: *Questions on Algebraic Varieties* (Centro Internationale Matematico Estivo, 3° ciclo, Varenna, 1969), 29–100. Edizioni Cremonese, Roma, 1970. *106*

[82] S. J. PATTERSON, A lattice problem in hyperbolic space. *Mathematika* **22** (1975), 81–88. *49*

[83] M. O. RABIN, Probabilistic algorithms in finite fields. *SIAM Journal on Computing* **9** (1980), no. 2, 273–280. *103*

[84] R. A. RANKIN, Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. II. *Proceedings of the Cambridge Philosophical Society* **35** (1939), 357–372. *40*

[85] R. A. RANKIN, *Modular forms and functions*. Cambridge University Press, 1977. *51*

[86] K. A. RIBET, Galois representations attached to modular forms with Nebentypus. In: J.-P. SERRE and D. B. ZAGIER (editors), *Modular Functions of One Variable V* (Proceedings of the International Conference, University of Bonn, 1976). Lecture Notes in Mathematics **601**, 17–51. Springer-Verlag, Berlin/Heidelberg, 1977. *20*

[87] K. A. RIBET, Mod p Hecke operators and congruences between modular forms. *Inventiones mathematicae* **71** (1983), 193–205. *18*

[88] K. A. RIBET, On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms. *Inventiones mathematicae* **100** (1990), 431–476. *10*

[89] K. A. RIBET, Report on mod l representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. In: *Motives* (Seattle, WA, 1991), 639–676. Proceedings of Symposia in Pure Mathematics **55**, Part 2. American Mathematical Society, Providence, RI, 1994. *21*

[90] K. A. RIBET and W. A. STEIN, Lectures on Serre's conjectures. With an appendix by B. CONRAD and an appendix by K. BUZZARD. In: B. CONRAD and K. RUBIN (editors), *Arithmetic Algebraic Geometry* (Graduate Summer School of the Institute for Advanced Study/Park City Mathematics Institute held in Park City, UT, 1999), 143–232. IAS/Park City Mathematics Series **9**. American Mathematical Society, Providence, RI, and Institute for Advanced Study, Princeton, NJ, 2001. *27, 29*

[91] W. ROELCKE, Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, I. *Mathematische Annalen* **167** (1966), 292–337. *52, 54*

[92] W. ROELCKE, Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, II. *Mathematische Annalen* **168** (1967), 261–324. *54*

[93] E. F. SCHAEFER, A new proof for the non-degeneracy of the Frey–Rück pairing and a connection to isogenies over the base field. In: T. SHASKA (editor), *Computational Aspects of Algebraic Curves* (Conference held at the University of Idaho, 2005), 1–12. Lecture Notes Series in Computing **13**. World Scientific Publishing, Hackensack, NJ, 2005. *138*

[94] A. SELBERG, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *Journal of the Indian Mathematical Society (N.S.)* **20** (1956), 47–87. (= *Collected Papers*, Volume 1, 423–463. Springer-Verlag, Berlin, 1989.) *31, 32, 33*

[95] A. Selberg, Discontinuous groups and harmonic analysis. In: *Proceedings of the International Congress of Mathematicians (Stockholm, 15–22 August 1962)*, 177-189. Institut Mittag-Leffler, Djursholm, 1963. (= *Collected Papers*, Volume 1, 493–505. Springer-Verlag, Berlin, 1989.) *43*

[96] A. Selberg, On the estimation of Fourier coefficients of modular forms. In: A. L. Whiteman (editor), *Theory of Numbers*, 1–15. Proceedings of Symposia in Pure Mathematics, VIII. American Mathematical Society, Providence, RI, 1965. (= *Collected Papers*, Volume 1, 506–520. Springer-Verlag, Berlin, 1989.) *70*

[97] J-P. Serre, *Abelian l-adic representations and elliptic curves*. Benjamin, New York, 1968. *19*

[98] J-P. Serre, Une interprétation des congruences relatives à la fonction τ de Ramanujan. Séminaire Delange–Pisot–Poitou 1967/68, n° **14**. *19*

[99] J-P. Serre, Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Mathematical Journal* **54** (1987), 179–230. *15, 19, 21, 27*

[100] *Théorie des topos et cohomologie étale des schémas* (SGA 4). Tome 3 (exposés IX à XIX). Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964, dirigé par M. Artin, A. Grothendieck et J.-L. Verdier, avec la collaboration de P. Deligne et B. Saint-Donat. Lecture Notes in Mathematics **305**. Springer-Verlag, Berlin/Heidelberg/New York, 1973. *112, 113, 139*

[101] *Théorie des intersections et théorème de Riemann–Roch* (SGA 6). Séminaire de Géometrie Algébrique du Bois-Marie 1966–1967, dirigé par P. Berthelot, A. Grothendieck et L. Illusie, avec la collaboration de D. Ferrand, J. P. Jouanolou, O. Jussila, S. Kleiman, M. Raynaud et J.-P. Serre. Lecture Notes in Mathematics **225**. Springer-Verlag, Berlin/Heidelberg/New York, 1971. *193*

[102] G. Shimura, Correspondances modulaires et les fonctions ζ de courbes algébriques. *Journal of the Mathematical Society of Japan* **10** (1958), no. 1, 1–28. *19*

[103] В. В. Шокуров, Интегралы Шимуры параболических форм. *Известия Академии Наук СССР, серия математическая* **44** (1980), № 3, 670–718.
V. V. Shokurov, Shimura integrals of cusp forms. *Mathematics of the USSR Izvestiya* **16** (1981), no. 3, 603–646. (English translation.) *152*

[104] W. A. Stein, *Modular Forms, a Computational Approach*. With an appendix by P. E. Gunnells. American Mathematical Society, Providence, RI, 2007. *38, 107, 152*

[105] J. Sturm, On the congruence of modular forms. In: D. V. Chudnovsky, G. V. Chudnovsky, H. Cohn and M. B. Nathanson (editors), *Number Theory (New York, 1984–1985)*. Lecture Notes in Mathematics **1240**, 275–280. Springer-Verlag, Berlin/Heidelberg, 1987. *17*

[106] L. Szpiro, *Séminaire sur les pinceaux arithmétiques : la conjecture de Mordell*. Astérisque **127**, Société Mathematique de France, 1985. *74, 78, 79, 80, 81*

[107] R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras. *Annals of Mathematics* (2) **141** (1995), no. 3, 553–572. *28*

[108] A. A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications, I.* Springer-Verlag, New York, 1985. *31*

[109] M. van Hoeij, Factoring polynomials and the knapsack problem. *Journal of Number Theory* **95** (2002), no. 2, 167–189. *156*

[110] I. Varma, *Sums of Squares, Modular Forms, and Hecke Characters.* Master's thesis, Universiteit Leiden, 2010. *5*

[111] P. J. Weinberger, Finding the number of factors of a polynomial. *Journal of Algorithms* **5** (1984), no. 2, 180–186. *4*

[112] G. Wiese, Multiplicities of Galois representations of weight one. With an appendix by N. Naumann. *Algebra and Number Theory* **1** (2007), no. 1, 67–85. *29*

[113] G. Wiese, On the faithfulness of parabolic cohomology as a Hecke module over a finite field. *Journal für die reine und angewandte Mathematik* **606** (2007), 79–103. *29*

[114] G. Wiese, Dihedral Galois representations and Katz modular forms. *Documenta Mathematica* **9** (2009), 123–133. *21*

[115] A. Wiles, Modular elliptic curves and Fermat's Last Theorem. *Annals of Mathematics* (2) **141** (1995), no. 3, 443–551. *28*

[116] S. Zhang, Admissible pairing on a curve. *Inventiones mathematicae* **112** (1993), no. 1, 171–193. *81, 82, 92, 93, 95*

List of notation

Index

Samenvatting

In deze samenvatting wil ik in het kort een beeld geven van de inhoud van dit proefschrift, te beginnen met enkele woorden over het deelgebied van de wiskunde waarin het hierin beschreven onderzoek zich afspeelt. Dit deelgebied, dat bekend staat als de *aritmetische meetkunde*, is ontstaan door de wisselwerking van twee veel oudere gebieden, de *getaltheorie* en de *algebraïsche meetkunde*.

In de getaltheorie gaat het over de eigenschappen van getallen. Laat ik eerst preciezer maken naar wat voor soort getallen we kijken. Allereerst zijn er de *gehele getallen* $(\ldots, -2, -1, 0, 1, 2, \ldots)$ en de *rationale getallen*, dat wil zeggen breuken waarvan de teller en noemer gehele getallen zijn (zoals $3/4$). Deze fundamentele getalstelsels kunnen op verschillende manieren worden uitgebreid. De uitbreiding die voor ons vooral interessant is, bestaat uit de *algebraïsche getallen*. Dit zijn alle getallen die voldoen aan een zogeheten polynoomvergelijking met rationale coëfficiënten. Zo is het getal $\sqrt{2} = 1{,}41421\ldots$ algebraïsch, aangezien het een oplossing is van de vergelijking $x^2 = 2$. Er zijn echter ook algebraïsche getallen die niet op de bekende getallenlijn liggen, zoals een getal i dat een oplossing is van de vergelijking $x^2 = -1$.

In de algebraïsche meetkunde wordt de meetkundige structuur van de oplossingsverzamelingen van stelsels polynoomvergelijkingen bestudeerd. Het taalgebruik van dit onderzoeksgebied omvat begrippen als *algebraïsche kromme* en *algebraïsch oppervlak*. De modulaire krommen uit de titel van dit proefschrift zijn belangrijke voorbeelden van algebraïsche krommen.

Een blik in de geschiedenis

Een probleem dat historisch gezien een belangrijke motivatie gevormd heeft voor ontwikkelingen in de aritmetische meetkunde, is het oplossen van *diophantische vergelijkingen*. Deze vergelijkingen zijn genoemd naar Diophantus van Alexandrië, die leefde in de derde eeuw. Het oplossen van een diophantische vergelijking betekent het vinden van gehele of rationale getallen die in een gegeven betrekking tot elkaar staan. Laten we een beroemd voorbeeld van een diophantische vergelijking bekijken.

De Fransman Pierre de Fermat (begin 17$^{\mathrm{e}}$ eeuw–1665) was jurist van beroep. Hij is echter in de eerste plaats beroemd geworden door zijn bijdragen aan de wiskunde. Eén van zijn belangrijkste nalatenschappen was de *laatste stelling van Fermat*, lange tijd het grootste onopgeloste wiskundige probleem. Deze stelling zegt dat als n een geheel getal is met $n \geq 3$, de diophantische vergelijking

$$a^n + b^n = c^n$$

geen oplossingen heeft in positieve gehele getallen a, b, en c. In 1637 schreef Fermat in zijn exemplaar van de *Arithmetica* van Diophantus dat hij een wondermooi bewijs van deze stelling had gevonden, maar dat hij het uit ruimtegebrek niet in de marge kwijt kon. Helaas heeft Fermat ook nergens anders een bewijs van zijn bewering neergepend, zodat de laatste stelling van Fermat de naam 'stelling' in feite niet waardig was.

De zoektocht naar een bewijs dat de laatste stelling van Fermat tot een echte stelling zou maken, is gedurende meer dan 350 jaar een aaneenschakeling geweest van vergeefse pogingen van talloze amateur- en beroepswiskundigen. Uiteindelijk werd de laatste stelling van Fermat in 1995 bewezen door de Britse wiskundige Andrew Wiles. Voor een laatste, essentieel onderdeel van het bewijs werkte Wiles samen met zijn vroegere promovendus Richard Taylor. Wiles en Taylor maakten gebruik van de modernste technieken uit de getaltheorie en de algebraïsche meetkunde, deels door henzelf ontwikkeld. Het staat wel vast dat deze technieken in de tijd van Fermat nooit ontdekt hadden kunnen worden.

Wiles' bewijs van de laatste stelling van Fermat is geenszins een op zichzelf staande prestatie. Het werd enerzijds mogelijk gemaakt door een lange reeks eerdere ontwikkelingen, en is anderzijds een beslissende stap geweest in de richting van nieuwe ontdekkingen in de getaltheorie. Al deze vorderingen zijn, kort gezegd, pogingen tot het begrijpen van een opmerkelijk verband tussen twee op het eerste gezicht totaal verschillende takken van de wiskunde: de Galoistheorie en de theorie van modulaire vormen. We zullen hieronder kort een idee geven van deze beide gebieden.

De Galoistheorie

Een gereedschap dat tegenwoordig onmisbaar is bij het bestuderen van oplossingen van vergelijkingen, is de *Galoistheorie*, genoemd naar de Franse wiskundige Évariste Galois (1811–1832). Ondanks zijn vroegtijdige dood (op 20-jarige leeftijd bij een pistoolduel met een nooit geheel opgehelderde aanleiding) heeft Galois zijn stempel gedrukt op de wiskunde na hem.

De fundamentele observatie van Galois was dat oplossingen van vergelijkingen allerlei *symmetrieën* kunnen hebben. Laten we bijvoorbeeld eens kijken naar de vergelijking

$$x^4 - 4x^2 + 2 = 0.$$

Deze vergelijking heeft precies vier verschillende oplossingen, namelijk

$$-\sqrt{2+\sqrt{2}}, \quad -\sqrt{2-\sqrt{2}},$$

$$\sqrt{2-\sqrt{2}}, \qquad \sqrt{2+\sqrt{2}}.$$

Er bestaan verschillende algebraïsche relaties tussen deze oplossingen. Die zijn in dit geval zodanig van aard dat er precies vier manieren zijn waarop we de oplossingen onderling kunnen verwisselen zonder deze relaties geweld aan te doen. Deze verwisselingen blijken we te kunnen krijgen door het bovenstaande 'vierkant' met oplossingen een aantal kwartslagen met de klok mee te draaien.

In plaats van per vergelijking de symmetrieën van de verzameling oplossingen te bekijken, kunnen we ook proberen het totaal aan symmetrieën van alle algebraïsche getallen tegelijkertijd te bestuderen. Het object dat daarbij centraal staat, heet de *Galoisgroep van de algebraïsche getallen.* Dit is een nogal gecompliceerd object, wat ertoe heeft geleid dat er allerlei geavanceerde technieken ontwikkeld zijn om het enigszins grijpbaar te maken.

Modulaire vormen

Een ander onderwerp dat in de twintigste eeuw intensief is onderzocht – en dat op het eerste gezicht niet zoveel met de Galoistheorie te maken heeft – is de theorie van *modulaire vormen.* Het is moeilijk om in een paar zinnen uit te leggen wat modulaire vormen precies zijn of om een portret te schetsen dat recht doet aan de vele interessante eigenschappen ervan. Eén van de manieren om een modulaire vorm te beschrijven is door middel van een zogeheten q-reeks, een bepaald soort oneindige som. Om een indruk te geven van hoe een modulaire vorm er in de gedaante van een q-reeks uitziet, geven we hier twee voorbeelden:

$$\begin{aligned} E_4 &= \frac{-B_4}{8} + \sum_{n=1}^{\infty} \sigma_3(n) q^n \\ &= \frac{1}{240} + q + 9q^2 + 28q^3 + 73q^4 + 126q^5 + 252q^6 + 344q^7 + \dots \end{aligned}$$

en

$$\begin{aligned} \Delta &= q \prod_{m=1}^{\infty} (1 - q^m)^{24} \\ &= q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 - 16744q^7 + \dots \end{aligned}$$

Het is ook mogelijk modulaire vormen op verschillende manieren grafisch weer te geven. Voor de doeleinden van dit proefschrift zit de interessante informatie echter in de *coëfficiënten* van modulaire vormen: de getallen die in de voorbeelden hierboven voor de machten van q staan.

Een opmerkelijk verband

De Franse wiskundige Jean-Pierre Serre wierp aan het einde van de jaren 1960 het vermoeden op dat er voor elke modulaire vorm een collectie 'vlakke voorstellingen' van de Galoisgroep van de algebraïsche getallen zou moeten zijn met specifieke eigenschappen die bepaald worden door de modulaire vorm. Het bestaan van dergelijke vlakke voorstellingen zou een bepaald verschijnsel met betrekking tot coëfficiënten van modulaire vormen verklaren. Niet lang daarna liet de Belgische wiskundige Pierre Deligne zien hoe zulke vlakke voorstellingen met behulp van geavanceerde meetkundige technieken inderdaad geconstrueerd kunnen worden.

Eén van de eenvoudigste voorbeelden van een vlakke voorstelling is de manier waarop zojuist door middel van een meetkundige beschrijving – het draaien van het vierkant – de symmetrieën van de verzameling oplossingen van de vergelijking $x^4 - 4x^2 + 2 = 0$ aanschouwelijk gemaakt werden. Eerlijkheidshalve moet vermeld worden dat dit specifieke voorbeeld niet modulair is: doordat er geen complexe getallen in voorkomen, hebben we met een zogenaamde *even* voorstelling te maken, terwijl modulaire exemplaren altijd *oneven* zijn.

Dit proefschrift

Voor 'gewone' modulaire vormen zouden we een oneindige hoeveelheid informatie nodig hebben om de bijbehorende vlakke voorstelling volledig te beschrijven. Voor zogeheten *modulaire vormen over eindige lichamen* wordt de vlakke voorstelling echter door een eindige hoeveelheid gegevens vastgelegd. Omdat 'vlakke voorstelling van de Galoisgroep van de algebraïsche getallen die hoort bij een modulaire vorm over een eindig lichaam' een hele mond vol is, noemen we zo'n object in de rest van deze samenvatting kortweg een *modulaire voorstelling*.

Om verschillende redenen is het een interessante vraag hoe we modulaire voorstellingen concreet uit kunnen rekenen. Eén daarvan is dat de eerdergenoemde coëfficiënten van modulaire vormen via zulke voorstellingen snel berekend kunnen worden. Deze vraag is het onderwerp geweest van eerder onderzoek van mijn promotor en verschillende Leidse en buitenlandse medewerkers. Zij hebben in de afgelopen jaren een strategie ontwikkeld om modulaire voorstellingen te berekenen in het geval dat het zogeheten *niveau*, een maat voor de complexiteit, gelijk is aan 1. In dit proefschrift heb ik voortgebouwd op hun methoden en heb deze uitgebreid naar modulaire voorstellingen van hoger niveau.

Het eindproduct van dit proefschrift is een algoritme (een 'rekenrecept') om gegeven een modulaire vorm een concrete beschrijving van de bijbehorende vlakke voorstelling te vinden. Het belangrijkste ingrediënt van zo'n concrete beschrijving is een algebraïsche vergelijking met de eigenschap dat de symmetrieën van de oplossingen de gezochte vlakke voorstelling leveren. Een probleem is dat het lastig te voorspellen is hoe ingewikkeld die vergelijking eruit ziet. Een maat voor de complexiteit van de vergelijking is de *hoogte*. Deze hoogte is grotendeels bepalend voor de tijd die nodig is om de berekening uit te voeren. Om te bewijzen dat onze algoritme efficiënt is, moeten we dan ook deze hoogte begrenzen. Hiervoor benutten we de *Arakelovtheorie*, een interessant stuk wiskunde waarop we hier helaas niet kunnen ingaan.

Een interessant gevolg van de in dit proefschrift bewezen resultaten gaat over *sommen van kwadraten*, een klassiek onderwerp uit de getaltheorie. Als k en n positieve gehele getallen zijn, schrijven we $r_k(n)$ voor het aantal manieren waarop n geschreven kan worden als een som van k kwadraten van gehele getallen. Om allerlei moeilijkheden te omzeilen, beperken we ons tot de situatie waar k een even getal is. Er bestaan formules voor $r_k(n)$ wanneer k gelijk is aan 2, 4, 6, 8 of 10. Deze formules zijn eind achttiende, begin negentiende eeuw gevonden door Legendre, Gauß, Jacobi en Liouville, klinkende namen uit de wiskundige geschiedenis.

Voor grotere waarden van k zijn er weliswaar allerlei formules bekend, maar die helpen weinig om $r_k(n)$ snel te berekenen. Het is sinds kort zelfs bekend dat er voor geen enkele $k \geq 12$ een formule is die lijkt op de formules voor $k \leq 10$. Met de resultaten van dit proefschrift kan echter het volgende bewezen worden. Neem ten eerste aan dat we de ontbinding van n in priemfactoren kennen. (Er zijn wel algoritmen om die te vinden, maar die zijn niet efficiënt in de betekenis die wij nodig hebben.) Neem ten tweede aan dat de *gegeneraliseerde Riemannhypothese* waar is. Dit is een beroemd onbewezen vermoeden dat samenhangt met de verdeling van priemgetallen. Dan is het mogelijk om $r_k(n)$ efficiënt te berekenen. Onder dezelfde voorwaarden kunnen ook coëfficiënten van algemenere modulaire vormen snel berekend worden.

Dankwoord

Dit proefschrift zou niet tot stand gekomen zijn zonder mijn promotor Bas Edixhoven en mijn copromotor Robin de Jong. Het is me een groot genoegen geweest om de afgelopen jaren met jullie samen te werken. Ik ben jullie veel dank verschuldigd voor jullie inzichten, suggesties en correcties.

Je remercie les membres de la *promotiecommissie* pour leur aide et pour avoir pris le temps de lire cette thèse, particulièrement vu le fait que le bon mot de Pascal est applicable à elle : Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte. (*Les provinciales*, seizième lettre, 4 décembre 1656.)

Enkele van de bij dit proefschrift gevoegde stellingen zijn mede mogelijk gemaakt door nuttige opmerkingen van Klaas Pieter Hart, Hendrik Lenstra en Marco Streng, die ik daarvoor op deze plaats wil bedanken.

I thank Claus Diem for an idea that is used in Algorithm IV.3.5 and that is much simpler than the method I originally had in mind.

I would like to seize this opportunity to express my gratitude to the numerous people whom I have had the pleasure to meet through mathematics, both in the Netherlands and throughout the world, and who have made it a joy for me to be part of the mathematical community. I hope that we will meet many more times in the future.

Ik heb vijf jaar lang met veel plezier in Leiden gestudeerd en heb er vervolgens vier buitengewoon gelukkige jaren beleefd als promovendus. Dat is vooral te danken aan alle collega's op het Mathematisch Instituut, aan wie ik zonder uitzondering goede herinneringen overhoud. Met name noem ik – in onwillekeurige volgorde – Arjen, Birgit, Cecília, Erwin, Franck, Gabriel, Gabriele, Ionica, Jan, Jeanine, Johan, Jos, Lenny, Marco, Oleg, René, Ronald en Willem Jan.

De afgelopen jaren zijn natuurlijk niet uitsluitend met wiskunde gevuld geweest. Voor allerlei zaken die voor mij minstens net zo veel hebben betekend, ben ik mijn ouders, mijn verdere familie en mijn vrienden zeer dankbaar. In plaats van me aan een lang verhaal te wagen, wil ik een indruk geven van wat voor mij betekenisvol geweest is, met het idee dat velen van hen zich daarin zullen herkennen. Dat doe ik – op het risico af als te academisch dan wel te onacademisch opgevat te worden – met een aantal Latijnse frasen: *Nec tamen consumebatur – Concordia – Albanianæ – Præsidium libertatis – Pax et bonum – Veritas – Soli Deo gloria.*

Curriculum vitæ

Pieter Jan Bruin werd geboren op 4 september 1983 te Gouda. Hij groeide op in het boomkwekersdorp Boskoop en doorliep het gymnasium op het Groene Hart Lyceum te Alphen aan den Rijn. In 1997 werd hij de winnaar van de Nationale Spellingwedstrijd voor middelbare scholieren. Samen met vijf andere scholieren vertegenwoordigde hij Nederland bij de Internationale Wiskunde Olympiade in 2000 (Zuid-Korea) en in 2001 (Verenigde Staten).

In 2001 begon Peter sterrenkunde en wiskunde te studeren in Leiden. Aanvankelijk streefde hij een loopbaan in de sterrenkunde na en beschouwde de wiskunde als handig gereedschap voor een toekomstig astronoom. Na ruim drie jaar bleek de wiskunde hem echter genoeg te trekken om hem te doen besluiten zich daarop toe te leggen. In juni 2006 studeerde hij met lof af op een scriptie getiteld *Green functions on Riemann surfaces and an application to Arakelov theory*, onder begeleiding van prof. dr. Bas Edixhoven en dr. Robin de Jong.

Vervolgens bleef hij in Leiden om als promovendus aan de slag te gaan in het kader van Edixhovens NWO-project *Arithmetic geometry, motives*. Het proefschrift dat voor u ligt, is hier het resultaat van. Hij heeft voordrachten over dit onderzoek gegeven in onder andere Amsterdam, Essen, Berlijn, Marseille en Lausanne.

Naast het doen van onderzoek heeft hij zich tijdens zijn studie- en promotietijd ingezet voor het wiskundeonderwijs in Leiden en deels ook op grotere schaal. Dit gebeurde onder meer door het geven van werkcolleges, het begeleiden van een bachelorscriptie en het (mede) organiseren van seminaria voor gevorderde studenten en promovendi. Ook verzorgde hij gastlessen om middelbare scholieren kennis te laten maken met onderwerpen die de schoolwiskunde te buiten gaan.

Behalve wiskundige is Peter van huis uit muziekliefhebber. Zelf speelt hij trompet en slagwerk. Verder is hij in de loop der tijd gefascineerd geraakt door taal en door het grensgebied tussen geloof en wetenschap. Hij is betrokken geweest bij de organisatie van verscheidene fora voor een academisch publiek over onderwerpen in dit grensgebied.

Vanaf het najaar van 2010 zal hij als postdoctoraal onderzoeker verbonden zijn aan de Université Paris-Sud 11.