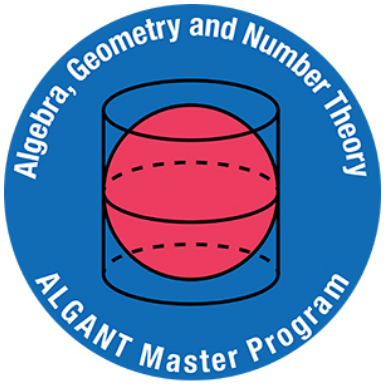


Modular forms modulo p

FRANCESC GISPERT

under the direction of
ADRIAN IOVITA and MATTEO LONGO

A thesis submitted to

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

in partial fulfilment of the requirements for the
ALGANT MASTER'S DEGREE IN MATHEMATICS

Defence date: 16th July 2018

Biblatex information:

```
@thesis{gispert2018modformsmodp,
   author={Gispert, Francesc},
   title={Modular forms modulo \(p\)},
   date={2018-07-16},
   institution={Universität Regensburg and Università degli Studi di
Padova},
   type={Master's thesis},
   pagetotal={99}
}
```

Dedicat a en Julian Assange, la Tamara Carrasco, en Roger Español,
l'Adrià Carrasco, l'Anna Gabriel, en Pablo *Hasél*,
en Jordi Perelló, la Clara Ponsatí, en Josep *Valtònyc*
i, en general, a tots els valents que són injustament perseguits
per la seva incansable lluita per a l'alliberament nacional.

Segarem cadenes fins a la vostra llibertat, que és també la nostra.
Donec perficiam.

Abstract

This thesis presents the theory of modular forms in positive characteristic p.

Under certain hypotheses, there are modular curves parametrizing elliptic curves with additional level structures of arithmetic interest. Katz defined modular forms as global sections of a family of line bundles over these modular curves.

Moreover, the modular curves have a special kind of points called the cusps. The germs of a modular form at the cusps correspond to power series in one variable called the q–expansions of the modular form. The q–expansions often hold a lot of information about the modular form.

In particular, over an algebraically closed field of characteristic p, the algebra of modular forms has a very simple description in terms of a modular form known as the Hasse invariant. Specifically, modular forms are determined up to powers of the Hasse invariant by their q–expansions. In addition, there is a differential operator acting on the algebra of modular forms which helps us to understand this structure.

Keywords: modular forms, congruences for modular forms, elliptic curves, modular curves

MSC2010: 11F33, 11G18, 11F11, 11F25

Preface

This thesis explains the main results describing the algebra of modular forms modulo a prime number p. Specifically, the objective of the thesis is to develop all the necessary theory in order to understand Katz's article [9] in detail (including the claims without proof in its introduction).

Modular forms are defined classically as analytic functions in the complex upper half-plane which transform in a certain way under the action of a group of matrices. Therefore, modular forms satisfy many symmetries which endow them with a very rich structure. In particular, modular forms have Fourier series expansions called q–expansions.

The Fourier coefficients of certain modular forms carry a large amount of arithmetic information. For instance, modular forms occur as generating functions of numbers of representations of integers by positive definite quadratic forms, special values of L–functions or invariants in algebraic number theory such as class numbers. Not to mention the deep connections of modular forms with elliptic curves and Galois representations. In other words, modular forms (and their generalizations) are virtually ubiquitous in number theory. This justifies why one might be interested in understanding well the theory of modular forms.

One can study classical modular forms just focusing on their q–expansions, which are power series. In this context, there is a very simple notion of modular forms modulo p, namely, the power series obtained by reducing the coefficients of q–expansions modulo p whenever this is possible (i.e., when all the coefficients are rational numbers with no powers of p in the denominators). This notion was developed by Serre and Swinnerton-Dyer.

In this way, one obtains a subalgebra of $\mathbb{F}_p[\![q]\!]$ which can be described explicitly in terms of a particular modular form, the Eisenstein series of weight $p-1$ (this is true for $p \geq 5$; for $p = 2$ or $p = 3$ the result changes a bit), as explained in Swinnerton-Dyer's article [21] and in Serre's article [17]. Indeed, the q–expansion of the Eisenstein series of weight $p-1$ reduces to 1 modulo p and this determines completely the relations between the q–expansions modulo p of classical modular forms (see theorem 1.41). In addition, divisibility by the Eisenstein series of weight $p-1$ induces a filtration and there is a derivation acting on q–expansions as $q\frac{d}{dq}$ which *behaves well* with respect to this filtration

(see corollary 1.37 and proposition 1.44).

The initial definition of classical modular forms may lead to the impression that these objects are analytic in nature and so better understood by means of complex analysis and differential geometry. Nevertheless, there is an alternative interpretation in terms of moduli spaces (whence the name modular forms) which allows us to study modular forms using the tools of modern algebraic geometry.

More precisely, modular forms can be regarded as *rules* which assign values to elliptic curves with certain level structures and there exist modular schemes whose points *parametrize* elliptic curves with such structures (i.e., they represent the corresponding moduli problems). One can then view modular forms as global sections of certain line bundles on a modular scheme.

These modular schemes and the line bundles giving modular forms turn out to be defined over very general rings. In this way, we can define modular forms over rings other than $\mathbb{C}$ and use techniques from algebraic geometry to study them. This was the approach taken by Katz in his article [8], where he also introduced his notion of p–adic modular forms. Later generalizations are also based on the ideas of Katz.

Using Katz's definition of modular forms, the notion of modular forms modulo p becomes simply that of modular forms defined over a field of characteristic p. Following Katz's article [9], we focus especially on the case of an algebraically closed field K of characteristic p. The results of Serre and Swinnerton-Dyer in the classical case hold in this context after some slight modifications.

More concretely, there is a modular form called the Hasse invariant which plays the same role as the Eisenstein series of weight $p-1$ in the classical case. Indeed, the q–expansions of the Hasse invariant over K are equal to 1 and this determines completely the relations between the q–expansions of modular forms defined over K (see theorem 4.8). Again, divisibility by the Hasse invariant induces a filtration and there is a derivation acting on q–expansions as $q\frac{d}{dq}$ which *behaves well* with respect to this filtration (see theorem 4.12).

This thesis can be divided in two parts. The first part, corresponding to the first chapter, explains the simplest case of classical modular forms defined analytically with the goal of illustrating the results which one can expect in more general situations. The proofs in this part use quite simple techniques of analysis in one complex variable and manipulations of power series. In turn, the second part, corresponding to the other three chapters, develops Katz's theory

of modular forms and, in this context, exhibits results analogous to those of the first part. However, Katz's theory relies on much more involved concepts of algebraic geometry which are mostly summarized without proofs.

Chapter 1 introduces classical modular forms (for $\mathrm{SL}_2(\mathbb{Z})$) and describes the main examples, with an emphasis on their q–expansions, and the structure of the algebra of such modular forms. There is also a brief introduction to the theory of Hecke operators. After that, the structure of the algebra of modular forms modulo p is studied in detail.

Chapter 2 defines Katz's modular forms. At the beginning of the chapter, we give a quite *naive* definition of modular forms as *rules* assigning a scalar value to each triple consisting of an elliptic curve together with a global differential and a level structure. The q–expansions of a modular form are defined using this interpretation by means of the Tate curve. After that, there is a summary of results (without proofs) about the representability of moduli problems which allow us to redefine modular forms as global sections of certain line bundles on modular schemes (in fact, curves). Using this last interpretation, we prove the q–expansion principle, which roughly states that a modular form is uniquely determined by its q–expansions.

Chapter 3 serves as a preparation for chapter 4. That is to say, chapter 3 presents some geometric tools which could be defined more generally but are later used only in special cases in order to study Katz's modular forms over a field of characteristic p. In particular, we recall the definition of the Frobenius morphisms, some facts about de Rham cohomology and the constructions of the Gauss–Manin connection and the Kodaira–Spencer morphism.

Finally, chapter 4 introduces the Hasse invariant and uses it to describe the structure of the algebra of modular forms over an algebraically closed field of characteristic p. The results presented in chapter 4 are analogous to those of the last section of chapter 1.

Each chapter contains a little summary and some references at the beginning.

Acknowledgements

En primer lloc, vull donar les gràcies a en Jordi Quer per introduir-me al món de la teoria de nombres. Tot i que al principi va mirar de prevenir-me'n —«Et fiques en un bon merder, ho saps? Ja ho has pensat prou bé, això?»—, el tema que va triar per al meu treball de fi de grau i el seu guiatge em van motivar encara més

a dedicar-me a la recerca en teoria de nombres. A més a més, en Jordi sempre ha estat encantat d'explicar-me les idees principals darrere de tota mena de temes, amb la seva meravellosa capacitat de fer semblar simples les coses que no ho són tant.

He d'agrair a en Francesc Fité que em recomanés el màster d'Algant quan cercava què fer després del grau. Si no fos per ell, probablement no hauria descobert l'existència d'aquest programa i qui sap on seria ara. També m'ha ajudat molt en Víctor Rotger, que aparentment coneix tothom que es dedica a la teoria de nombres i té una molt bona idea de què fa cadascú. A en Víctor, li he de donar les gràcies pels seus encertats consells sobre la gent de qui val la pena aprendre.

Ich muss Jonathan Glöckle, Kathrin Lechl, Benedikt Preis und Julian Seipel für ihre unendliche Geduld, mein gebrochenes Deutsch zu verbessern, danken. Sie haben mir immer so viel beigebracht und ich fühlte mich in Regensburg zu Hause. Es bedeutet mir viel.

Ich bedanke mich bei Ulrich Görtz für seine immer hilfreiche E-Mails, als ich zum ersten Mal Schematheorie studierte. Er hat mir viele Fragen davon beantwortet.

I am also thankful to Victor Lisinski and to Benedikt Preis for many amusing discussions of mathematical problems, which sometimes degenerated into more banal but equally fun conversations.

Devo ringraziare Giacomo Graziani e Alberto Prandini per avermi aiutato tanto a imparare un po' d'italiano, insegnandomi un sacco di parole ed espressioni interesanti. A parte questo, sono grato a Giacomo per le sue spiegazioni sempre giuste e facili di capire su tante idee matematiche. La mia comprensione di diversi concetti che apparono in questa tesi è molto migliorata grazie ai suoi commenti.

I am also indebted to my advisors Matteo Longo and Adrian Iovita. On the one hand, I want to thank Matteo for always being very approachable and helping me understand several results which I did not know, even though many of the things I learnt from him did not make it into the thesis. On the other hand, I am grateful to Adrian for suggesting such an interesting topic to study and for giving me the right references at every moment, especially taking into account my often imprecise questions. In addition, I must thank him for spotting an important mistake in a proof that I wrote. I am looking forward to learning much more from Adrian in the years to come.

Last but not least, I would like to express my gratitude to *la Caixa* Foundation, for their fellowship programme for postgraduate studies in Europe has allowed me to study for the Algant master's degree during the last two years.

FRANCESC GISPERT

Padua, Italy
June 2018

Contents

Chapter 1

The classical theory

This chapter introduces the classical theory of modular forms of level 1 in order to illustrate in the simplest case the kind of objects and results which are studied later.

Modular forms are a class of holomorphic functions with some *symmetry* properties which allow us to express them in terms of certain power series called q–expansions. One can study these power series directly, but then it is convenient to find additional structure on the space of modular forms. One way to do so is by means of Hecke operators, which have particularly simple expressions in terms of the aforementioned q–expansions. Alternatively, one can focus on those power series with integer coefficients and reduce them modulo a prime number.

The presentation in the first two sections is based mainly on chapter VII of Serre's book [16] and on some parts of chapter III of Koblitz's book [15]. Also, the first two chapters of Stein's book [20] treat the same topics but from a more computational point of view. The last section of this chapter follows closely section 3 of Swinnerton-Dyer's article [21] and section 1 of Serre's article [17].

1.1 Modular forms

To begin with, we present some definitions and results in the simplest case. Unless otherwise stated, all results appearing in this section and the next are proved in sections VII.2 to VII.5 of Serre's book [16].

Consider the complex upper half-plane $\mathbb{H} = \{ z \in \mathbb{C} : \mathrm{Im}(z) > 0 \}$, which is then extended by adding the cusps $\mathbb{P}^1_{\mathbb{Q}} = \mathbb{Q} \sqcup \{ \infty \}$. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on both $\mathbb{H}$ and $\mathbb{P}^1_{\mathbb{Q}}$ via the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}$$

with the usual conventions that

$$\frac{w}{0} = \infty \quad \text{and} \quad \frac{a\infty + b}{c\infty + d} = \frac{a}{c}\,.$$

In fact, one checks that the action on $\mathbb{P}^1_{\mathbb{Q}}$ is transitive. Modular forms are a class of analytic functions which are *almost* invariant under this action.

Definition 1.1. Let $k \in \mathbb{Z}$. A *weakly modular form* for $\mathrm{SL}_2(\mathbb{Z})$ of weight k is a meromorphic function $f\colon \mathbb{H} \to \mathbb{C}$ with the property that

$$f(z) = (cz+d)^{-k} f\Big(\frac{az+b}{cz+d}\Big) \quad \text{for all } z \in \mathbb{H} \text{ and all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})\,.$$

Remarks.

(1) There are no non-zero weakly modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of odd weight because
$$f(z) = f\Big(\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot z\Big) = (-1)^k f(z)\,.$$

(2) One can check that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the two matrices
$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$
so the condition in the definition is equivalent to
$$f\Big(\frac{1}{z}\Big) = z^k f(z) \quad \text{and} \quad f(z+1) = f(z) \quad \text{for all } z \in \mathbb{H}\,.$$

(3) Since $f(z+1) = f(z)$, f admits a Fourier series expansion of the form
$$f(z) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n z}$$
which is absolutely convergent on $\mathbb{H}$ minus the poles of f. The map $z \mapsto q = e^{2\pi i z}$ defines an analytic isomorphism from $\langle T \rangle \backslash \mathbb{H}$ to the open unit disk with the origin removed and this isomorphism can be extended to send ∞ to the origin. We say that f is *holomorphic* (resp. is *meromorphic* or *vanishes*) *at* ∞ if $a_n = 0$ for all $n < 0$ (resp. $n < n_0$ for some $n_0 \in \mathbb{Z}$ or $n \leq 0$). That is, f is holomorphic at ∞ if its Fourier series expansion is holomorphic at $q = 0$. In this case we write $f(\infty) = a_0$.

Definition 1.2. Let $k \in \mathbb{Z}$. A *modular form* for $\mathrm{SL}_2(\mathbb{Z})$ of weight k is a function $f\colon \mathbb{H} \to \mathbb{C}$ satisfying the following conditions:

(i) f is holomorphic on $\mathbb{H}$;

(ii) $f(z) = (cz+d)^{-k} f\Big(\frac{az+b}{cz+d}\Big)$ for all $z \in \mathbb{H}$ and all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$, and

(iii) f is holomorphic at ∞ in the sense that it has a Fourier series expansion of

the form

$$f(z) = \sum_{n=0}^{\infty} a_n q^n\,, \quad \text{where } q = e^{2\pi i z}\,,$$

which is absolutely convergent for each $z \in \mathbb{H}$, called the *q–expansion* of f. If, in addition, f vanishes at ∞ (i.e., $a_0 = 0$), we say that f is a *cusp form*. We refer to a function which satisfies (i), (ii) and (iii) with "holomorphic" replaced by "meromorphic" as a *meromorphic modular form* for $\mathrm{SL}_2(\mathbb{Z})$ of weight k.

Remarks.

(1) As $\mathrm{SL}_2(\mathbb{Z})$ acts transitively on the set of cusps, condition (ii) implies that f has a power series expansion at each cusp with the same order of vanishing as the q–expansion at ∞.

(2) This definition can be generalized to certain subgroups Γ of $\mathrm{SL}_2(\mathbb{Z})$. One must only take into account that there might be several equivalence classes of cusps and require holomorphicity at all of them.

(3) The set $M_k(\mathrm{SL}_2(\mathbb{Z}))$ of modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weight k is a $\mathbb{C}$–vector space and the subset $S_k(\mathrm{SL}_2(\mathbb{Z}))$ consisting of cusp forms is a subspace of codimension ≤ 1 because it is the kernel of the linear map $f \mapsto f(\infty)$. Furthermore,

$$M(\mathrm{SL}_2(\mathbb{Z})) = \bigoplus_{k \in \mathbb{Z}} M_k(\mathrm{SL}_2(\mathbb{Z}))$$

is a graded $\mathbb{C}$–algebra with the usual multiplication.

Often in number theory one can attach a power series to arithmetic objects. When this happens to be the q–expansion of a modular form, we can use the additional structure of the spaces of modular forms to obtain relations between the coefficients forming the series, which might lead to interesting arithmetic results.

In the first two sections of this chapter we study the structure of the algebra of modular forms, while in the last one we focus on the q–expansions reduced modulo a prime number p (whenever this makes sense). But, before that, we show the first examples of modular forms and compute their q–expansions.

Example 1.3. Let $k > 1$ be an integer. We define the *Eisenstein series* of weight $2k$ as

$$G_{2k}(z) = \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(mz+n)^{2k}}$$

for $z \in \mathbb{H}$. (The symbol $\sum'$ means that the summation runs over all values for which the corresponding addends *make sense*; in this case, over all pairs (m, n)

distinct from $(0,0)$.) This series converges to a modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight $2k$.

Indeed, since the exponents of the denominators are at least 4, one can prove that the series converges absolutely and uniformly on compact subsets of $\mathbb{H}$, thus defining a holomorphic function on $\mathbb{H}$. Also, we can compute

$$G_{2k}(z+1) = \sum_{m,n\in\mathbb{Z}}{}' (mz+m+n)^{-2k} = \sum_{m,n\in\mathbb{Z}}{}' (mz+n)^{-2k} = G_{2k}(z)\,,$$
$$G_{2k}\Big(\frac{-1}{z}\Big) = \sum_{m,n\in\mathbb{Z}}{}' \Big(\frac{m}{z}+n\Big)^{-2k} = z^{2k}\sum_{m,n\in\mathbb{Z}}{}' (m+nz)^{-2k} = z^{2k}G_{2k}(z)\,.$$

It remains to see that G_{2k} is holomorphic at infinity. To this aim, we need to prove that G_{2k} has a limit for $\mathrm{Im}(z)\to\infty$. But, by the good convergence properties of G_{2k}, we can make the passage to the limit term by term:

$$\lim_{z\to i\infty} G_{2k}(z) = \sum_{m,n\in\mathbb{Z}}{}' \lim_{z\to i\infty} \frac{1}{(mz+n)^{2k}} = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^{2k}} = 2\sum_{n=1}^{\infty}\frac{1}{n^{2k}} = 2\zeta(2k)\,,$$

where ζ denotes the Riemann zeta function.

We can even compute the q–expansions of the Eisenstein series explicitly.

Proposition 1.4. *For every integer $k>1$, the q–expansion of the Eisenstein series G_{2k} is*

$$G_{2k}(z) = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)\cdot q^n\,,$$

where

$$\sigma_j(n) = \sum_{d\,|\,n} d^j$$

is the sum of j–th powers of positive divisors of n and $q = e^{2\pi i z}$.

Proof. We start with the well-known formula

$$\frac{1}{z} + \sum_{n=1}^{\infty}\Big(\frac{1}{z+n}+\frac{1}{z-n}\Big) = \pi\frac{\cos(\pi z)}{\sin(\pi z)} = \pi i - \frac{2\pi i}{1-q} = \pi i - 2\pi i\sum_{r=0}^{\infty} q^r\,,$$

which can be obtained taking the logarithmic derivative of the expression of $\sin(\pi z)$ as an infinite product. By successive differentiations, we obtain the formula

$$\sum_{n\in\mathbb{Z}}\frac{1}{(z+n)^j} = \frac{1}{(j-1)!}(-2\pi i)^j\sum_{r=1}^{\infty} r^{j-1}q^r$$

for $j \geq 2$. After replacing z with mz, this becomes

$$\sum_{n\in\mathbb{Z}} \frac{1}{(mz+n)^j} = \frac{1}{(j-1)!}(-2\pi i)^j \sum_{r=1}^{\infty} r^{j-1} q^{mr} .$$

Finally, we use this to expand

$$\begin{aligned} G_{2k}(z) &= \sum_{m,n\in\mathbb{Z}}{}' \frac{1}{(mz+n)^{2k}} = 2\zeta(2k) + 2\sum_{m=1}^{\infty}\sum_{n\in\mathbb{Z}} \frac{1}{(mz+n)^{2k}} \\ &= 2\zeta(2k) + \frac{2(-2\pi i)^{2k}}{(2k-1)!}\sum_{m=1}^{\infty}\sum_{r=1}^{\infty} r^{2k-1}q^{mr} \\ &= 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)\cdot q^n \end{aligned}$$

as required. □

Example 1.5. It is well-known that

$$\zeta(2k) = \frac{(-1)^{k+1}B_{2k}(2\pi)^{2k}}{2(2k)!} ,$$

where B_j are the *Bernoulli numbers* defined by the equality (in $\mathbb{Q}[\![T]\!]$)

$$\frac{T}{e^T-1} = \sum_{j=0}^{\infty} B_j \frac{T^j}{j!} .$$

(See proposition 7 of chapter VII of Serre's book [16] for a proof of this identity.) The Bernoulli numbers are all rational and, moreover, there are some well-known congruence relations between them which we use later.

For each $k > 1$, we define the *normalized Eisenstein series*

$$E_{2k}(z) = \frac{1}{2\zeta(2k)}G_{2k}(z) = 1 - \frac{4k}{B_{2k}}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)q^n .$$

These normalized Eisenstein series have q–expansions with rational coefficients with very particular denominators. The first few E_{2k} are:

$$\begin{aligned} E_4(z) &= 1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n , \\ E_6(z) &= 1 - 504\sum_{n=1}^{\infty}\sigma_5(n)q^n , \end{aligned}$$

$$E_8(z) = 1 + 480 \sum_{n=1}^{\infty} \sigma_7(n) q^n \,.$$

We can use E_4 and E_6, whose q–expansions have constant term 1, to construct the first example of cusp form. More precisely, we define the *modular discriminant*

$$\Delta(z) = \frac{1}{1728}(E_4(z)^3 - E_6(z)^2)\,,$$

which is a cusp form for $\mathrm{SL}_2(\mathbb{Z})$ of weight 12. We have normalized Δ so that the first non-zero coefficient of its q–expansion is 1, but often in the literature the modular discriminant is defined to be $(2\pi)^{12}\Delta$ instead of Δ.

Let us check that the coefficients of the q–expansion of Δ are all integers. Observe that $1728 = 2^6 \cdot 3^3$, $240 = 2^4 \cdot 3 \cdot 5$ and $540 = 2^3 \cdot 3^2 \cdot 7$. Looking at the powers of 2 and of 3, it is easy to see that

$$E_4(z)^3 = \left(1 + 2^4 \cdot 3 \cdot 5 \sum_{n=1}^{\infty} \sigma_3(n) q^n\right)^3 \equiv 1 + 2^4 \cdot 3^2 \cdot 5 \sum_{n=1}^{\infty} \sigma_3(n) q^n \mod 2^6 \cdot 3^3$$

and

$$E_6(z)^2 = \left(1 - 2^3 \cdot 3^2 \cdot 7 \sum_{n=1}^{\infty} \sigma_5(n) q^n\right)^2 \equiv 1 - 2^4 \cdot 3^2 \cdot 7 \sum_{n=1}^{\infty} \sigma_5(n) q^n \mod 2^6 \cdot 3^3$$

(also, we have only modified mod 1728 the coefficients of q^n for $n \geq 2$, so the coefficients of 1 and q are exactly these ones). One checks that $d^3 \equiv d^5 \bmod 12$ for every $d \in \mathbb{Z}$, and summing over the positive divisors of $n \in \mathbb{N}$ we see that $2^4 \cdot 3^2 \cdot 5 \cdot \sigma_3(n) + 2^4 \cdot 3^2 \cdot 7 \cdot \sigma_5(n) \equiv 0 \bmod 2^6 \cdot 3^3$. This completes the proof.

Example 1.6. We can define E_2 in a similar way:

$$E_2(z) = \frac{1}{2\zeta(2)} \sum_{m \in \mathbb{Z}} {\sum_{n \in \mathbb{Z}}}' \frac{1}{(mz+n)^2} = 1 + \frac{3}{\pi^2} \sum_{m \in \mathbb{Z} \setminus \{0\}} \sum_{n \in \mathbb{Z}} \frac{1}{(mz+n)^2}\,.$$

In this case, however, the series is not absolutely convergent and so we must take into account the order of summation. In particular, the previous proof that $G_{2k}\left(\frac{-1}{z}\right) = z^{2k} G_{2k}(z)$ does not carry over to the case $k = 1$ because we need to reverse the order of summation over m and n. Therefore, E_2 is not a (classical) modular form for $\mathrm{SL}_2(\mathbb{Z})$. Still, it has similar properties which will be useful later (in fact, it is a p–adic modular form of weight 2; see section 2.1 of Serre's article [18]).

The same computation as in proposition 1.4 shows that

$$\frac{3}{\pi^2}\sum_{n\in\mathbb{Z}}\frac{1}{(mz+n)^2} = -12\sum_{r=1}^{\infty} rq^{mr}\,,$$

whence

$$E_2(z) = 1 - 24\sum_{m=1}^{\infty}\sum_{r=1}^{\infty} rq^{mr}\,.$$

Since $|q|<1$, this last double sum is absolutely convergent and we can collect powers of q to get the expression

$$E_2(z) = 1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n\,.$$

Let us see more precisely how E_2 fails to be a modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight 2. The next result is proposition 7 of chapter 3 of Koblitz's book [15].

Proposition 1.7. *The normalized Eisenstein series of weight* 2 *satisfies the relation*

$$z^{-2}E_2\Big(\frac{-1}{z}\Big) = E_2(z) + \frac{12}{2\pi i z}\,.$$

Proof. Set

$$a_{m,n}(z) = \frac{1}{(mz+n)(mz+n-1)} = \frac{1}{mz+n-1} - \frac{1}{mz+n}$$

and consider

$$\widetilde{E}_2(z) = 1 + \frac{3}{\pi^2}\sum_{m\in\mathbb{Z}\setminus\{0\}}\sum_{n\in\mathbb{Z}}\left[\frac{1}{(mz+n)^2} - a_{m,n}(z)\right].$$

On the one hand, the series $\sum_{n\in\mathbb{Z}} a_{m,n}(z)$ telescopes to 0 and so $\widetilde{E}_2(z) = E_2(z)$. On the other hand, the double series with general term

$$\frac{1}{(mz+n)^2} - \frac{1}{(mz+n)(mz+n-1)} = \frac{-1}{(mz+n)^2(mz+n-1)}$$

is absolutely convergent, which implies that we can change the order of the summation in $\widetilde{E}_2(z)$. That is,

$$E_2(z) = \widetilde{E}_2(z) = 1 + \frac{3}{\pi^2}\sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}\setminus\{0\}}\left[\frac{1}{(mz+n)^2} - a_{m,n}(z)\right]$$

$$= z^{-2}E_2\Big(\frac{-1}{z}\Big) - \frac{3}{\pi^2}\sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}\setminus\{0\}} a_{m,n}(z)\,.$$

It remains to compute this last double sum.

As in the proof of proposition 1.4, we find that

$$\sum_{m\in\mathbb{Z}\setminus\{0\}}\frac{1}{(mz+n)^2} = \frac{1}{z^2}\sum_{m\in\mathbb{Z}}\frac{1}{\big(m+\frac{n}{z}\big)^2} - \frac{1}{n^2} = -\frac{4\pi^2}{z^2}\sum_{r=1}^{\infty} re^{2\pi irn/z} - \frac{1}{n^2}$$

and the sum over n of this last expression is absolutely convergent. Thus, the outer sum in $\sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}\setminus\{0\}} a_{m,n}(z)$ must also be absolutely convergent (because the sum defining $\widetilde{E}_2(z)$ is). Therefore, we can compute

$$\begin{aligned}\sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}\setminus\{0\}} a_{m,n}(z) &= \lim_{N\to\infty}\sum_{n=-N+1}^{N}\sum_{m\in\mathbb{Z}\setminus\{0\}} a_{m,n}(z)\\ &= \lim_{N\to\infty}\sum_{m\in\mathbb{Z}\setminus\{0\}}\sum_{n=-N+1}^{N} a_{m,n}(z)\,.\end{aligned}$$

Using again the formulae introduced in the proof of proposition 1.4,

$$\begin{aligned}\sum_{m\in\mathbb{Z}\setminus\{0\}}\sum_{n=-N+1}^{N} a_{m,n}(z) &= \sum_{m\in\mathbb{Z}\setminus\{0\}}\Big(\frac{1}{mz-N} - \frac{1}{mz+N}\Big)\\ &= \frac{2}{z}\sum_{m=1}^{\infty}\left(\frac{1}{\frac{-N}{z}-m} + \frac{1}{\frac{-N}{z}+m}\right)\\ &= \frac{2}{z}\left[\pi\frac{\cos\big(\frac{-\pi N}{z}\big)}{\sin\big(\frac{-\pi N}{z}\big)} + \frac{z}{N}\right].\end{aligned}$$

In conclusion, the double sum we wanted to compute is

$$\frac{2\pi}{z}\lim_{N\to\infty}\frac{\cos\big(\frac{-\pi N}{z}\big)}{\sin\big(\frac{-\pi N}{z}\big)} = \frac{2\pi}{z}\lim_{N\to\infty} i\frac{e^{-2\pi iN/z}+1}{e^{-2\pi iN/z}-1} = -\frac{2\pi i}{z}\,,$$

whence

$$E_2(z) = z^{-2}E_2\Big(\frac{-1}{z}\Big) + \frac{6i}{\pi z} = z^{-2}E_2\Big(\frac{-1}{z}\Big) - \frac{12}{2\pi iz}\,.$$ □

After having introduced the main examples of modular forms, we describe explicitly the structure of the $\mathbb{C}$–vector spaces $M_k(\mathrm{SL}_2(\mathbb{Z}))$. Recall that, by the remarks after definition 1.2, we already know that $M_k(\mathrm{SL}_2(\mathbb{Z})) = 0$ when k is odd and that $S_k(\mathrm{SL}_2(\mathbb{Z}))$ has codimension 1 in $M_k(\mathrm{SL}_2(\mathbb{Z}))$ when k is even and

≥ 4 (as the Eisenstein series are not cusp forms). In fact, we can even give an explicit basis of each $M_k(\mathrm{SL}_2(\mathbb{Z}))$, $k \in \mathbb{Z}$.

Theorem 1.8 (valence formula). *Let k be an integer and let f be a meromorphic modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight k, not identically zero. Then,*

$$\mathrm{ord}_\infty(f) + \frac{1}{2}\,\mathrm{ord}_i(f) + \frac{1}{3}\,\mathrm{ord}_\rho(f) + \sum_P{}^{*}\,\mathrm{ord}_P(f) = \frac{k}{12}\,,$$

where $\sum_P^*$ means a summation over the points of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ distinct from the classes of i and $\rho = e^{\frac{\pi i}{3}}$.

Proof. This formula can be proved integrating the logarithmic derivative of f along a contour which contains in its interior *essentially* one representative of each orbit of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ and using the residue theorem. For a detailed proof, see theorem 3 of chapter VII of Serre's book [16]. □

Corollary 1.9. *$\Delta(z)$ does not vanish in $\mathbb{H}$ and has a simple zero at $i\infty$.*

Proof. Since Δ is a cusp form of weight 12, $\mathrm{ord}_P(\Delta) \geq 0$ for all $P \in \mathbb{H}\backslash\,\mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{ord}_\infty(\Delta) \geq 1$. By theorem 1.8, these numbers add up to 1: this is only possible if all the inequalities are equalities. □

Proposition 1.10. *Multiplication by Δ defines an isomorphism of $M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$ onto $S_k(\mathrm{SL}_2(\mathbb{Z}))$ for every $k \in \mathbb{Z}$.*

Proof. Clearly, if $f \in M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$, then $f\Delta \in S_k(\mathrm{SL}_2(\mathbb{Z}))$. For the converse, let $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$. We set $g = \frac{f}{\Delta}$, which is a meromorphic modular form of weight $k-12$. By the previous result, $\mathrm{ord}_P(g) = \mathrm{ord}_P(f) \geq 0$ for every $P \in \mathbb{H}$ and $\mathrm{ord}_\infty(g) = \mathrm{ord}_\infty(f) - 1 \geq 0$. In conclusion, $g \in M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$.

This proves that multiplication by Δ gives a bijection between $M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$ and $S_k(\mathrm{SL}_2(\mathbb{Z}))$, and it is obviously a linear map. □

Proposition 1.11. *Let k be an integer.*

(1) *$M_k(\mathrm{SL}_2(\mathbb{Z})) = 0$ if $k < 0$, k is odd or $k = 2$.*
(2) *$M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$ (that is, the only modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weight 0 are the constants).*
(3) *$M_k(\mathrm{SL}_2(\mathbb{Z}))$ has dimension 1 and G_k is a basis if $k = 4, 6, 8, 10$ or 14.*

Proof. Recall that, if f is a non-zero modular form of weight k, then

$$\operatorname{ord}_\infty(f) + \frac{1}{2}\operatorname{ord}_i(f) + \frac{1}{3}\operatorname{ord}_\rho(f) + \sum_p{}^{*} \operatorname{ord}_p(f) = \frac{k}{12}\,.$$

Since f is holomorphic, this means that $\frac{k}{12} \geq 0$. Moreover, k must be even because the least common denominator of the left-hand side is 6 and $k \neq 2$ because $\frac{1}{6}$ cannot be written in the form $a + \frac{b}{2} + \frac{c}{3}$ with $a, b, c \geq 0$.

If $k \leq 10$, then $k - 12 < 0$ and $S_k(\mathrm{SL}_2(\mathbb{Z})) = \{\,0\,\}$ by proposition 1.10. Hence, $\dim(M_k(\mathrm{SL}_2(\mathbb{Z}))) \leq 1$. But we already know that $1, G_4, G_6, G_8, G_{10}$ are non-zero modular forms for $\mathrm{SL}_2(\mathbb{Z})$ of weights $0, 4, 6, 8, 10$, respectively; this concludes the proof. □

Corollary 1.12. *For any integer k,*

$$\dim(M_k(\mathrm{SL}_2(\mathbb{Z}))) = \begin{cases} 0 & \text{if } k < 0,\ k \text{ is odd or } k = 2\,, \\ \left\lfloor \frac{k}{12} \right\rfloor & \text{if } k \geq 0,\ k \text{ is even and } \frac{k}{2} \equiv 1 \mod 6\,, \\ \left\lfloor \frac{k}{12} \right\rfloor + 1 & \text{if } k \geq 0,\ k \text{ is even and } \frac{k}{2} \not\equiv 1 \mod 6\,. \end{cases}$$

Proof. The result follows by induction on k (the inductive step is performed by increasing k to $k + 12$ using proposition 1.10). □

Corollary 1.13. *The q–expansion of Δ can be expressed as the infinite product*

$$\Delta(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24}\,, \qquad \text{where } q = e^{2\pi i z}\,.$$

Proof. Define

$$F(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24}\,.$$

It suffices to prove that $F \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$, as $\dim S_{12}(\mathrm{SL}_2(\mathbb{Z})) = 1$ and the coefficient of q in both the q–expansions of F and Δ is 1. Thus, we only need to prove that

$$F\Big(\frac{-1}{z}\Big) = z^{12}F(z)\,.$$

Observe that the logarithmic differential of F is

$$\frac{dF}{F} = \left[1 - 24\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} q^{mn}\right]\frac{dq}{q} = \left[1 - 24\sum_{n=1}^{\infty}\sigma_1(n)q^n\right]\frac{dq}{q} = E_2(z)\cdot 2\pi i\, dz\,.$$

Using proposition 1.7, we can compute

$$\frac{dF\left(\frac{-1}{z}\right)}{F\left(\frac{-1}{z}\right)} = 2\pi i E_2\Big(\frac{-1}{z}\Big)\frac{dz}{z^2} = 2\pi i E_2(z)\,dz + 12\frac{dz}{z} = \frac{dF(z)}{F(z)} + 12\frac{dz}{z}\,.$$

Hence, the functions $F\left(\frac{-1}{z}\right)$ and $z^{12}F(z)$ have the same logarithmic differential, which implies that $F\left(\frac{-1}{z}\right) = \lambda z^{12}F(z)$ for some constant $\lambda \in \mathbb{C}^\times$. Evaluating at $z = i$, we see that $0 \neq F(i) = \lambda F(i)$, whence $\lambda = 1$. □

Theorem 1.14. *Let k be a non-negative integer. The vector space $M_k(\mathrm{SL}_2(\mathbb{Z}))$ admits as a basis the family of monomials $E_4^\alpha E_6^\beta$ with α and β non-negative integers such that $4\alpha + 6\beta = k$. As a consequence, $M(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}[E_4, E_6]$.*

Proof. If k is odd, $M_k(\mathrm{SL}_2(\mathbb{Z})) = 0$ and the equation $4\alpha + 6\beta = k$ has no solutions, so we may assume that k is even.

First, we show that these monomials generate $M_k(\mathrm{SL}_2(\mathbb{Z}))$ by induction on k. This is clear for $k \leq 6$, so suppose that $k \geq 8$. Choose a pair (α_0, β_0) of non-negative integers such that $4\alpha_0 + 6\beta_0 = k$ (this is possible for $k \geq 4$). The modular form $g = E_4^{\alpha_0}E_6^{\beta_0}$, of weight k, is not a cusp form. Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$. Now $f - \frac{f(\infty)}{g(\infty)}g$ is a cusp form and, in particular, is of the form Δh for some $h \in M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$. We can apply the induction hypothesis to h and obtain thus f as a linear combination of the desired monomials. (Recall that, by definition, $\Delta = \frac{1}{1728}(E_4^3 - E_6^2)$.)

Now we check that these monomials are linearly independent. Suppose, for the sake of contradiction, that there exists a non-trivial linear combination

$$\sum_{4\alpha+6\beta=k} \lambda_{\alpha,\beta} E_4^\alpha E_6^\beta = 0\,.$$

Up to multiplying this linear relation by suitable powers of E_4 and of E_6, we may assume that k is a multiple of 12. Dividing by $E_6^{k/6}$, we obtain that

$$\sum_{4\alpha+6\beta=k} \lambda_{\alpha,\beta}\Big(\frac{E_4^3}{E_6^2}\Big)^{\alpha/3} = 0$$

(where $\frac{\alpha}{3}$ is an integer because $4\alpha = k - 6\beta$ is a multiple of 3). That is, the meromorphic function E_4^3/E_6^2 satisfies a non-trivial algebraic equation over $\mathbb{C}$ and, therefore, must be constant. This gives the desired contradiction because the q–expansion of E_4^3/E_6^2 is not constant. □

There is a more convenient basis for our purpose of reducing q–expansions modulo prime numbers. The next result is based on lemma 2.20 of Stein's book [20], which he attributes to V. Miller. (Unfortunately, the author of this work was unable to find the thesis of V. Miller where this result first appeared and so there is no proper citation.) Actually we only describe a basis *in echelon form*, whereas Miller's basis is *in diagonal form* (obtained after applying gaussian elimination in the obvious way).

Theorem 1.15. *Let k be an even integer ≥ 4 and let $d = \dim S_k(\mathrm{SL}_2(\mathbb{Z}))$. Choose two integers $\alpha, \beta \geq 0$ such that $4\alpha + 6\beta \leq 14$ and $4\alpha + 6\beta \equiv k \bmod 12$. Define $g_j = \Delta^j E_4^\alpha E_6^{2(d-j)+\beta}$ for $j = 0, 1, \ldots, d$. Every $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ whose q–expansion lies in $\mathbb{Z}[\![q]\!]$ can be expressed uniquely as a $\mathbb{Z}$–linear combination of $g_0, g_1, \ldots, g_d$.*

Proof. By corollary 1.12, we see that

$$d = \begin{cases} \left\lfloor \frac{k}{12} \right\rfloor - 1 & \text{if } k \equiv 2 \mod 12\,, \\ \left\lfloor \frac{k}{12} \right\rfloor & \text{if } k \not\equiv 2 \mod 12\,. \end{cases}$$

This, together with the conditions for $4\alpha + 6\beta$, shows that $k = 12d + 4\alpha + 6\beta$. We deduce that every g_j has weight k.

On the other hand, the q–expansions of E_4, E_6 and Δ lie in $\mathbb{Z}[\![q]\!]$ and their first non-zero coefficient is 1, whence the q–expansion of g_j lies in $\mathbb{Z}[\![q]\!]$ too and its first coefficients are $a_i(g_j) = 0$ for $i < j$ and $a_j(g_j) = 1$. In particular, the g_j for $j = 0, 1, \ldots, d$ are linearly independent over $\mathbb{C}$ and so form a basis of $M_k(\mathrm{SL}_2(\mathbb{Z}))$. Furthermore, if $f = \lambda_0 g_0 + \lambda_1 g_1 + \cdots + \lambda_d g_d$, then the first coefficients of the q–expansion of f are given by

$$a_j(f) = \lambda_j + \lambda_{j-1} a_j(g_{j-1}) + \cdots + \lambda_1 a_j(g_1) + \lambda_0 a_j(g_0)$$

for $j = 0, 1, \ldots, d$. We conclude that the q–expansion of f lies in $\mathbb{Z}[\![q]\!]$ if and only if $\lambda_0, \lambda_1, \ldots, \lambda_d \in \mathbb{Z}$ (for example, using gaussian elimination on the triangular system of linear equations). □

1.2 Hecke operators

Hecke operators play a fundamental role in the theory of modular forms. They are a family of averaging operators acting on the space of modular forms with particularly simple formulae for the action on q–expansions. Moreover, they are

multiplicative and satisfy certain recurrence relations. One can then show that the coefficients of the q–expansions of Hecke eigenforms (i.e., eigenvectors of the Hecke operators) satisfy similar relations.

Hecke operators are defined by means of the modular interpretation of modular forms. That is, we think of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ as a moduli space parametrizing complex elliptic curves with a nowhere vanishing differential 1–form. This interpretation is very important because it leads to Katz's generalization of the notion of modular forms.

A lattice in $\mathbb{C}$ is a subgroup of the form $\Lambda = \Lambda(\omega_1, \omega_2) = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$, where ω_1 and ω_2 are complex numbers which are linearly independent over $\mathbb{R}$. We shall always assume (up to interchanging ω_1 and ω_2) that $\mathrm{Im}\left(\frac{\omega_1}{\omega_2}\right) > 0$. Let $\mathcal{L}$ denote the set of lattices in $\mathbb{C}$. One checks easily that $\Lambda(\omega_1, \omega_2) = \Lambda(\omega_1', \omega_2')$ if and only if $\omega_1' = a\omega_1 + b\omega_2$ and $\omega_2' = c\omega_1 + d\omega_2$ for some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$.

Given a lattice Λ, we can form the one-dimensional complex torus $\mathbb{C}/\Lambda$ with a holomorphic differential $\omega = dz$ (where z is the coordinate on $\mathbb{C}$). Consider the function

$$\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{l \in \Lambda\backslash\{0\}} \left[\frac{1}{(z-l)^2} - \frac{1}{l^2} \right]$$

and the constants

$$g_2(\Lambda) = 60 \sum_{l \in \Lambda\backslash\{0\}} \frac{1}{l^4} \quad \text{and} \quad g_3(\Lambda) = 140 \sum_{l \in \Lambda\backslash\{0\}} \frac{1}{l^6}\,.$$

We have an analytic isomorphism from $\mathbb{C}/\Lambda$ to the complex elliptic curve

$$E\colon Y^2Z = 4X^3 - g_2(\Lambda)XZ^2 - g_3(\Lambda)Z^3$$

given by the map

$$z + \Lambda \mapsto \begin{cases} (\wp(z;\Lambda) : \wp'(z;\Lambda) : 1) & \text{if } z \notin \Lambda\,, \\ (0:1:0) & \text{if } z \in \Lambda\,, \end{cases}$$

under which ω corresponds to $\frac{dx}{y}$. Conversely, given an elliptic curve E over $\mathbb{C}$ together with a nowhere vanishing holomorphic differential ω, we can construct a lattice

$$\Lambda(E,\omega) = \left\{ \int_\gamma \omega : \gamma \in H_1(E,\mathbb{Z}) \right\}.$$

These two constructions are inverse to each other and so give a correspondence

between lattices Λ and pairs (E, ω) consisting of an elliptic curve with a nowhere vanishing holomorphic differential (see section I.6 of Koblitz's book [15] for more details).

One can show that two elliptic curves $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic if and only if $\Lambda = \lambda\Lambda'$ for some $\lambda \in \mathbb{C}^\times$. Therefore, we consider the (left) action of $\mathbb{C}^\times$ on $\mathcal{L}$ by homotheties and a lattice $\Lambda(\omega_1, \omega_2)$ is equivalent to $\Lambda(\tau, 1)$ in $\mathbb{C}^\times\backslash\mathcal{L}$, where $\tau = \frac{\omega_1}{\omega_2} \in \mathbb{H}$. Thus, we write $\Lambda(\tau) = \Lambda(\tau, 1)$ for all $\tau \in \mathbb{H}$. We see that $\Lambda(\tau)$ and $\Lambda(\tau')$ coincide in $\mathbb{C}^\times\backslash\mathcal{L}$ if and only if there exists $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that $\tau' = \frac{a\tau+b}{c\tau+d}$. Hence, there is a bijective correspondence between the elements of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ and the isomorphism classes of elliptic curves. $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ is said to be a moduli space of elliptic curves over $\mathbb{C}$.

Throughout the remainder of this section, let k be a fixed integer.

Definition 1.16. The *modular function* associated with $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ is the function $F\colon \mathcal{L} \to \mathbb{C}$ defined by

$$F(\Lambda(\omega_1, \omega_2)) = \omega_2^{-k} f\Big(\frac{\omega_1}{\omega_2}\Big).$$

Remarks.

(1) The function F is well-defined. Indeed, for every $\Lambda(\omega_1, \omega_2) \in \mathcal{L}$ and every $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$,

$$\begin{aligned} F(\Lambda(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)) &= (c\omega_1 + d\omega_2)^{-k} f\Big(\frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2}\Big) \\ &= \omega_2^{-k}\Big(c\frac{\omega_1}{\omega_2} + d\Big)^{-k} f\left(\frac{a\frac{\omega_1}{\omega_2} + b}{c\frac{\omega_1}{\omega_2} + d}\right) \\ &= F(\Lambda(\omega_1, \omega_2)). \end{aligned}$$

(2) The function F is homogeneous of degree $-k$: $F(\lambda\Lambda) = \lambda^{-k}F(\Lambda)$ for all $\Lambda \in \mathcal{L}$ and all $\lambda \in \mathbb{C}^\times$.

(3) We can recover f from F because $f(z) = F(\Lambda(z))$.

We use this interpretation of modular forms to define Hecke operators. To this aim, we first define operators on lattices, which then induce operators on functions on lattices. Write $\mathbb{Z}^{(\mathcal{L})}$ for the free abelian group generated by the elements of $\mathcal{L}$.

Definition 1.17. Let $n \in \mathbb{N}$. The *Hecke operator* $\mathrm{T}_n\colon \mathbb{Z}^{(\mathcal{L})} \to \mathbb{Z}^{(\mathcal{L})}$ is the unique $\mathbb{Z}$–linear operator which maps each lattice Λ to the sum of all of its sublattices of

index n:

$$\mathrm{T}_n[\Lambda] = \sum_{[\Lambda:\Lambda']=n} [\Lambda'] .$$

Similarly, the *homothety operator* $\mathrm{R}_n \colon \mathbb{Z}^{(\mathcal{L})} \to \mathbb{Z}^{(\mathcal{L})}$ is the $\mathbb{Z}$–linear operator defined by $\mathrm{R}_n[\Lambda] = [n\Lambda]$.

Proposition 1.18. *The Hecke operators and the homothety operators satisfy the following identities:*

(1) $\mathrm{R}_m \circ \mathrm{R}_n = \mathrm{R}_n \circ \mathrm{R}_m = \mathrm{R}_{mn}$ *for all* $m, n \in \mathbb{N}$;
(2) $\mathrm{R}_m \circ \mathrm{T}_n = \mathrm{T}_n \circ \mathrm{R}_m$ *for all* $m, n \in \mathbb{N}$;
(3) $\mathrm{T}_m \circ \mathrm{T}_n = \mathrm{T}_n \circ \mathrm{T}_m = \mathrm{T}_{mn}$ *for all* $m, n \in \mathbb{N}$ *such that* $(m, n) = 1$, *and*
(4) $\mathrm{T}_{p^n} \circ \mathrm{T}_p = \mathrm{T}_{p^{n+1}} + p\,\mathrm{T}_{p^{n-1}} \circ \mathrm{R}_p$ *for all primes* p *and all* $n \in \mathbb{N}$.

Proof. The first two identities are trivial.

To prove the third identity, fix a lattice Λ. For every sublattice Λ'' of Λ of index mn, there exists a unique sublattice Λ' of Λ containing Λ'' and such that $[\Lambda : \Lambda'] = n$ and $[\Lambda' : \Lambda''] = m$. Indeed, Λ/Λ'' is an abelian group of order mn which decomposes uniquely as the direct sum of a group of order m and a group of order n (because $(m, n) = 1$). Therefore, $\mathrm{T}_{mn}[\Lambda] = (\mathrm{T}_m \circ \mathrm{T}_n)[\Lambda]$.

Finally, we prove the last identity with a similar argument. Let Λ be a lattice. We observe that $(\mathrm{T}_{p^n} \circ \mathrm{T}_p)[\Lambda]$, $\mathrm{T}_{p^{n+1}}[\Lambda]$ and $(\mathrm{T}_{p^{n-1}} \circ \mathrm{R}_p)[\Lambda]$ are all formal sums of sublattices of Λ of index p^{n+1}. One such sublattice Λ'' occurs exactly a times in the first sum, exactly once in the second sum and exactly b times in the third sum, so we have to prove that $a = 1 + pb$. To do so, we distinguish two cases.

If Λ'' is not contained in $p\Lambda$, it is clear that $b = 0$. On the other hand, a is the number of sublattices Λ' of Λ containing Λ'' with $[\Lambda : \Lambda'] = p$. Such a lattice Λ' contains $p\Lambda$, and its image in $\Lambda/p\Lambda$ is of order p and contains the image of Λ'', which must also be of order p. Thus, there is exactly one possible Λ' with these properties, which means that $a = 1$.

If Λ'' is contained in $p\Lambda$, we get that $b = 1$. But every sublattice Λ' of Λ of index p contains $p\Lambda$ and so Λ'' too. Therefore, a coincides with the number of sublattices of Λ of index p (or, equivalently, with the number of subgroups of $\Lambda/p\Lambda \simeq (\mathbb{Z}/p\mathbb{Z})^2$ of index p), and this number is $\frac{p^2-1}{p-1} = p + 1$. □

Corollary 1.19. *The* $\mathbb{Z}$*–algebra generated by the* T_p *and* R_p *for all primes* p *contains all the* T_n, $n \in \mathbb{N}$, *and is commutative.*

Definition 1.20. There is an action of Hecke operators and homothety operators on the set of functions $F\colon \mathcal{L} \to \mathbb{C}$ which are homogeneous of degree $-k$: for all $n \in \mathbb{N}$, we define $\mathrm{T}_n F$ and $\mathrm{R}_n F$ to be the functions given by

$$\mathrm{T}_n F(\Lambda) = F(\mathrm{T}_n[\Lambda]) = \sum_{[\Lambda:\Lambda']=n} F(\Lambda')$$

and $\mathrm{R}_n F(\Lambda) = F(\mathrm{R}_n[\Lambda]) = n^{-k}F(\Lambda)$.

Proposition 1.21. *Let $F\colon \mathcal{L} \to \mathbb{C}$ be a homogeneous function of degree $-k$.*

(1) $\mathrm{T}_m \mathrm{T}_n F = \mathrm{T}_{mn} F$ *for all $m, n \in \mathbb{N}$ such that $(m, n) = 1$.*
(2) $\mathrm{T}_p \mathrm{T}_{p^n} F = \mathrm{T}_{p^{n+1}} F + p^{1-k} \mathrm{T}_{p^{n-1}} F$ *for all primes p and all $n \in \mathbb{N}$.*

Proof. It is immediate from proposition 1.18. □

Definition 1.22. Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and let $F\colon \mathcal{L} \to \mathbb{C}$ be the associated modular function (as in definition 1.16). For every $n \in \mathbb{N}$, we define $\mathrm{T}_n f\colon \mathbb{H} \to \mathbb{C}$ to be the function associated with $n^{k-1} \mathrm{T}_n F$:

$$\mathrm{T}_n f(z) = n^{k-1} \mathrm{T}_n F(\Lambda(z, 1)) .$$

Remarks.

(1) The factor n^{k-1} is introduced so that some formulae have integer coefficients.
(2) One can check that $\mathrm{T}_n f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$, as $\mathrm{T}_n f(z)$ is defined as a linear combination of values of f. This will become apparent once we give explicit formulae for the action of T_n.

Proposition 1.23. *Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$.*

(1) $\mathrm{T}_m \mathrm{T}_n f = \mathrm{T}_{mn} f$ *for all $m, n \in \mathbb{N}$ such that $(m, n) = 1$.*
(2) $\mathrm{T}_p \mathrm{T}_{p^n} f = \mathrm{T}_{p^{n+1}} f + p^{k-1} \mathrm{T}_{p^{n-1}} f$ *for all primes p and all $n \in \mathbb{N}$.*

Proof. It is immediate from the definition of $\mathrm{T}_n f$ and proposition 1.21 (taking into account the additional factor). □

It is often convenient to understand the action of Hecke operators in terms of the definitions given in the previous section. The following results provide simpler descriptions of these functions and even precise formulae to compute them.

Lemma 1.24. *Let A be a 2×2 matrix with entries in $\mathbb{Z}$ and $\det(A) = n > 0$. There exists $U \in \mathrm{SL}_2(\mathbb{Z})$ such that $UA = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with $ad = n$, $a \geq 1$ and $0 \leq b < d$. Moreover, the integers a, b and d are uniquely determined.*

Proof. It is possible to put A into upper triangular form by using elementary operations of the following types: adding a multiple of one row to another and swapping two rows. Since these operations are invertible, they correspond to left multiplication by a matrix in $\mathrm{SL}_2(\mathbb{Z})$. Now we can assume, up to multiplication by $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, that the diagonal entries are positive. Finally, adding a suitable multiple of the second row to the first one, we obtain a matrix of the form $UA = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with the required properties.

For the uniqueness, observe that a is the greatest common divisor of the elements in the first column of A (the operations performed to obtain an upper triangular form coincide with Euclid's algorithm). Now, $d = \frac{n}{a}$ and b is obviously uniquely determined modulo d. □

Proposition 1.25. *Let $n \in \mathbb{N}$ and let $M(n)$ be the set of 2×2 matrices with entries in $\mathbb{Z}$ and determinant n. Let $X(n)$ be the subset of $M(n)$ consisting of matrices of the form $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with $a \geq 1$ and $0 \leq b < d$. If $\Lambda = \Lambda(\omega_1, \omega_2)$, then the map*

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mapsto \Lambda(a\omega_1 + b\omega_2, d\omega_2)$$

is a bijection between $X(n)$ and the set of sublattices of Λ of index n.

Proof. If $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in X(n)$, then $\Lambda(a\omega_1 + b\omega_2, d\omega_2)$ has index n in Λ because $ad = n$. Conversely, if Λ' is a sublattice of Λ of index n, then every basis of Λ' must be of the form $(\omega_1', \omega_2') = (a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$ for some $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M(n)$. By lemma 1.24, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is $\mathrm{SL}_2(\mathbb{Z})$–equivalent to exactly one element of $X(n)$. □

Corollary 1.26. *Let $n \in \mathbb{N}$ and let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$. Then,*

$$\mathrm{T}_n f(z) = n^{k-1} \sum_{a,b,d} d^{-k} f\Big(\frac{az + b}{d}\Big)$$

where the sum is over the triples of integers a, b and d such that $a \geq 1$, $ad = n$ and $0 \leq b < d$.

Theorem 1.27. *Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ with q–expansion*

$$f(z) = \sum_{m=0}^{\infty} c(m) q^m \,.$$

For every $n \in \mathbb{N}$, $\mathrm{T}_n f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and its q–expansion is

$$\mathrm{T}_n f(z) = \sum_{m=0}^{\infty} c_n(m) q^m ,$$

where

$$c_n(m) = \sum_{a \Big| (n,m)} a^{k-1} c\Big(\frac{mn}{a^2}\Big)$$

(the last sum is over the positive divisors of (n,m)) for all $m \in \mathbb{N}$. In particular, if $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$, then $\mathrm{T}_n f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ too.

Proof. Let F be the modular function associated with f. For each $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$,

$$\begin{aligned} \mathrm{T}_n f\Big(\frac{az+b}{cz+d}\Big) &= n^{k-1} \mathrm{T}_n F\Big(\Lambda\Big(\frac{az+b}{cz+d}, 1\Big)\Big) \\ &= n^{k-1} \mathrm{T}_n F((cz+d)^{-1}\Lambda(az+b, cz+d)) \\ &= (cz+d)^k n^{k-1} \mathrm{T}_n F(\Lambda(z,1)) = (cz+d)^k \mathrm{T}_n f(z) . \end{aligned}$$

Moreover, by corollary 1.26, we see that $\mathrm{T}_n f$ is holomorphic on $\mathbb{H}$ because f is. It remains to prove that $\mathrm{T}_n f$ is holomorphic at ∞, which we do by computing its q–expansion explicitly.

We can write

$$\mathrm{T}_n f(z) = n^{k-1} \sum_{a \geq 1} \sum_{ad=n} \sum_{0 \leq b < d} d^{-k} \sum_{m=0}^{\infty} c(m) e^{2\pi i m \frac{az+b}{d}} .$$

But, for fixed a and d, the sum $\sum_{0 \leq b < d} e^{2\pi i b m/d}$ is 0 unless $d \mid m$, in which case it is d. Thus, setting $m' = \frac{m}{d}$,

$$\mathrm{T}_n f(z) = n^{k-1} \sum_{a,d,m'} d^{-k+1} c(m'd) e^{2\pi i a m' z} .$$

In the previous expression, we can collect powers of $e^{2\pi i z}$ and write $t = am'$ to obtain that

$$c_n(t) = \sum_{a \Big| (n,t)} a^{k-1} c\Big(\frac{n}{a}\frac{t}{a}\Big)$$

(the sum is over the positive divisors of (n,t)). □

We could have defined Hecke operators by means of the formula in theorem 1.27, which is enough if one is interested only in formal manipulations of

power series. However, it is not at all clear that, when one applies this formula to the q–expansion of a modular form, the resulting power series is again the q–expansion of a modular form. That is why the more abstract point of view presented here is more convenient.

Corollary 1.28. *Consider $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ non-constant and with q–expansion*

$$f(z) = \sum_{m=0}^{\infty} c_m q^m .$$

If f is an eigenvector of all the T_n (i.e., an eigenform*), say $\mathrm{T}_n f = \lambda_n f$ with $\lambda_n \in \mathbb{C}$ for each $n \in \mathbb{N}$, then $c_1 \neq 0$ and $c_m = \lambda_m c_1$ for all $m \in \mathbb{N}$.*

Proof. The coefficient of q in the q–expansion of $\mathrm{T}_n f$ is c_n, by theorem 1.27. But, since $\mathrm{T}_n f = \lambda_n f$, this number is also $\lambda_n c_1$. Finally, if c_1 were zero, then c_m would be zero for all $m \in \mathbb{N}$, thus contradicting the assumption that f is not a constant. □

Corollary 1.29. *Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ with q–expansion*

$$f(z) = \sum_{m=0}^{\infty} c_m q^m .$$

If f is an eigenform with $c_1 = 1$, then

(1) *$c_m c_n = c_{mn}$ for all $m, n \in \mathbb{N}$ with $(m, n) = 1$ and*
(2) *$c_p c_{p^n} = c_{p^{n+1}} + p^{k-1} c_{p^{n-1}}$ for all primes p and all $n \in \mathbb{N}$.*

Proof. These relations follow immediately from corollary 1.28 and proposition 1.23. □

Example 1.30. As the spaces $M_4(\mathrm{SL}_2(\mathbb{Z}))$, $M_6(\mathrm{SL}_2(\mathbb{Z}))$ and $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ have dimension 1, we deduce that E_4, E_6 and Δ are eigenforms. In fact, one can prove that all Eisenstein series are eigenforms.

1.3 Modular forms modulo p

The exposition in this section follows closely section 3 of Swinnerton-Dyer's article [21]. The same results are also presented in section 1 of Serre's article [17].

To begin with, we fix some notation. Throughout this section, let p be a fixed prime number. Write v_p for the p–adic valuation on $\mathbb{Q}$ and consider its

valuation ring $\mathbb{Z}_{(p)}$ (this is the ring of rational numbers with denominators prime to p). We can identify modular forms with their q–expansions and regard each $M_k(\mathrm{SL}_2(\mathbb{Z}))$, $k \in \mathbb{Z}$, as a subspace of the ring of formal power series $\mathbb{C}[\![q]\!]$. We consider those modular forms whose q–expansions have coefficients in $\mathbb{Z}_{(p)}$: set

$$M_k^p(\mathrm{SL}_2(\mathbb{Z})) = \left\{ f \in M_k(\mathrm{SL}_2(\mathbb{Z})) : f = \sum_{n=0}^{\infty} a_n q^n \in \mathbb{Z}_{(p)}[\![q]\!] \right\}$$

for each $k \in \mathbb{Z}$ and

$$M^p(\mathrm{SL}_2(\mathbb{Z})) = \bigoplus_{k \in \mathbb{Z}} M_k^p(\mathrm{SL}_2(\mathbb{Z}))\,.$$

But elements of $\mathbb{Z}_{(p)}$ can be reduced mod p; that is, the residue field of $\mathbb{Z}_{(p)}$ is $\mathbb{F}_p$. Thus, given

$$f = \sum_{n=0}^{\infty} a_n q^n \in \mathbb{Z}_{(p)}[\![q]\!]\,,$$

we can form

$$\widetilde{f} = \sum_{n=0}^{\infty} \widetilde{a}_n q^n \in \mathbb{F}_p[\![q]\!]\,,$$

where $\widetilde{a}_n$ denotes the image of a_n in $\mathbb{F}_p$. This leads us to the definition of modular forms mod p.

Definition 1.31. For each $k \in \mathbb{Z}$, let

$$\widetilde{M}_k^p(\mathrm{SL}_2(\mathbb{Z})) = \{\, \widetilde{f} \in \mathbb{F}_p[\![q]\!] : f \in M_k^p(\mathrm{SL}_2(\mathbb{Z}))\,\}\,.$$

The $\mathbb{F}_p$–algebra of *modular forms modulo* p for $\mathrm{SL}_2(\mathbb{Z})$ is

$$\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z})) = \sum_{k \in \mathbb{Z}} \widetilde{M}_k^p(\mathrm{SL}_2(\mathbb{Z}))\,.$$

Our objective is to determine the structure of the algebra $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$. Write

$$P = E_2 = 1 - 24 \sum_{n=1}^{\infty} \sigma_1(n) q^n\,,$$

$$Q = E_4 = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n\,,$$

$$R = E_6 = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n\,.$$

(See examples 1.5 and 1.6 for the formulae.)

By theorem 1.15, every $f \in M_k^p(\mathrm{SL}_2(\mathbb{Z}))$ can be expressed uniquely as an isobaric polynomial of weight k in Q, R and Δ (which have weights 4, 6 and 12, respectively) with coefficients in $\mathbb{Z}_{(p)}$. Therefore, $M^p(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{Z}_{(p)}[Q, R, \Delta]$ (both regarded as subrings of $\mathbb{Z}_{(p)}[\![q]\!]$) and it only remains to find the algebraic relations satisfied by $\widetilde{Q}$, $\widetilde{R}$ and $\widetilde{\Delta}$ in $\mathbb{F}_p[\![q]\!]$. But, from the explicit formulae above, two cases are trivial.

Theorem 1.32. *If $p = 2$ or $p = 3$, then $\widetilde{P} = \widetilde{Q} = \widetilde{R} = 1$ and $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{F}_p[\widetilde{\Delta}]$ is the algebra of polynomials in one variable $\widetilde{\Delta}$ with coefficients in $\mathbb{F}_p$ (i.e., it is isomorphic to $\mathbb{F}_p[X]$).*

From now on, we assume that $p \geq 5$. In this case, $p \nmid 1728$ and, from the equation $1728\Delta = Q^3 - R^2$, we deduce that $M^p(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{Z}_{(p)}[Q, R]$. Also, Q and R satisfy no non-trivial algebraic relations over $\mathbb{C}$, as we saw in the proof of theorem 1.14. Thus, we have surjective ring homomorphisms

$$\begin{aligned} M^p(\mathrm{SL}_2(\mathbb{Z})) \cong \mathbb{Z}_{(p)}[X, Y] &\twoheadrightarrow \mathbb{F}_p[X, Y] \twoheadrightarrow \widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z})) \\ \Phi(X, Y) &\mapsto \widetilde{\Phi}(X, Y) \mapsto \widetilde{\Phi}(\widetilde{Q}, \widetilde{R}) \end{aligned}$$

given by reduction mod p and we want to determine the kernel of the last arrow. But, before doing so, we need to introduce some additional structure on the algebra of modular forms.

Consider the operator

$$\theta = q\frac{d}{dq}$$

acting on $\mathbb{C}[\![q]\!]$:

$$\theta\left(\sum_{n=0}^{\infty} a_n q^n\right) = \sum_{n=1}^{\infty} n a_n q^n \,.$$

Theorem 1.33 (Ramanujan). *Let $k \in \mathbb{Z}$.*

(1) *For every $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$, $(12\theta - kP)f \in M_{k+2}(\mathrm{SL}_2(\mathbb{Z}))$.*

(2) *We have identities*

$$\theta P = \frac{1}{12}(P^2 - Q)\,, \qquad \theta Q = \frac{1}{3}(PQ - R)\,,$$
$$\theta R = \frac{1}{2}(PR - Q^2)\,, \qquad \theta \Delta = P\Delta\,.$$

Proof. Observe that, in terms of z,

$$\theta = \frac{1}{2\pi i}\frac{d}{dz}.$$

Take $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and set $F = (12\theta - kP)f$. It is clear that F is holomorphic (also at the cusps) and that $F(z+1) = F(z)$, so it only remains to prove that $F\left(\frac{-1}{z}\right) = z^{k+2}F(z)$. Differentiating the equation $f\left(\frac{-1}{z}\right) = z^k f(z)$, we get that

$$f'\Big(\frac{-1}{z}\Big) = kz^{k+1}f(z) + z^{k+2}f'(z).$$

Using these formulae and proposition 1.7, we obtain that

$$\begin{aligned} F\Big(\frac{-1}{z}\Big) &= \frac{12}{2\pi i}f'\Big(\frac{-1}{z}\Big) - kP\Big(\frac{-1}{z}\Big)f\Big(\frac{-1}{z}\Big) \\ &= \frac{12}{2\pi i}z^{k+2}f'(z) - kz^{k+2}P(z)f(z) = z^{k+2}F(z). \end{aligned}$$

This concludes the proof of (1).

Similarly, set $G = 12\theta P - P^2$, which is holomorphic (also at the cusps) and satisfies that $G(z+1) = G(z)$ because P has these properties. We differentiate the equation $P\left(\frac{-1}{z}\right) = z^2P(z) + \frac{12z}{2\pi i}$ (see proposition 1.7) to obtain that

$$P'\Big(\frac{-1}{z}\Big) = z^4P'(z) + 2z^3P(z) + \frac{12z^2}{2\pi i}$$

and then compute

$$G\Big(\frac{-1}{z}\Big) = \frac{12}{2\pi i}P'\Big(\frac{-1}{z}\Big) - P\Big(\frac{-1}{z}\Big)^2 = \frac{12}{2\pi i}z^4P(z) - z^4P(z)^2 = z^4G(z).$$

Therefore, $G \in M_4(\mathrm{SL}_2(\mathbb{Z}))$. But this space has dimension 1 and $G(\infty) = -1$, so $G = -Q$. Finally, by (1), we obtain that $G_1 = (12\theta - 4P)Q \in M_6(\mathrm{SL}_2(\mathbb{Z}))$, $G_2 = (12\theta - 6P)R \in M_8(\mathrm{SL}_2(\mathbb{Z}))$ and $G_3 = (12\theta - 12P)\Delta \in M_{14}(\mathrm{SL}_2(\mathbb{Z}))$. As these spaces have dimension 1 and $G_1(\infty) = -4$, $G_2(\infty) = -6$ and $G_3(\infty) = 0$, we conclude that $G_1 = -4R$, $G_2 = -6Q^2$ and $G_3 = 0$. □

Definition 1.34. We define ∂ to be the derivation on $M(\mathrm{SL}_2(\mathbb{Z}))$ acting on each graded component $M_k(\mathrm{SL}_2(\mathbb{Z}))$, $k \in \mathbb{Z}$, by $12\theta - kP$.

Remarks.

(1) The first part of theorem 1.33 shows that ∂ is well-defined and one checks easily that ∂ is a derivation.

(2) By the second part of theorem 1.33, $\partial Q = -4R$ and $\partial R = -6Q^2$. Hence, ∂ acts on $M^p(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{Z}_{(p)}[Q,R]$, which is isomorphic to the polynomial ring $\mathbb{Z}_{(p)}[X,Y]$ via $X \mapsto Q$ and $Y \mapsto R$. Thus, we obtain an induced derivation ∂ on $\mathbb{Z}_{(p)}[X,Y]$ defined by $\partial X = -4Y$ and $\partial Y = -6X^2$. The same equations define a derivation on $\mathbb{F}_p[X,Y]$ which we call ∂ as well.

(3) If $f \in M_k^p(\mathrm{SL}_2(\mathbb{Z}))$, we write $\partial\widetilde{f}$ for $\partial f \mod p$, which lies in $\widetilde{M}_{k+2}^p(\mathrm{SL}_2(\mathbb{Z}))$. That is, $\partial\widetilde{f} = (12\theta - k\widetilde{P})\widetilde{f}$ in $\mathbb{F}_p[\![q]\!]$.

Theorem 1.35. *Let $k \in \mathbb{N}$.*

(1) *If $p-1 \mid 2k$, then $pB_{2k} \in \mathbb{Z}_{(p)}$ and $pB_{2k} \equiv -1 \mod p$ (Clausen–von Staudt congruence). In particular, $v_p(B_{2k}) = -1$.*

(2) *If $p-1 \nmid 2k$, then $B_{2k}/(2k) \in \mathbb{Z}_{(p)}$ and its residue class $\mod p$ depends only on $2k \mod p-1$. That is,*

$$\frac{B_{2k}}{2k} \equiv \frac{B_{2k'}}{2k'} \mod p \quad \text{if } 2k \equiv 2k' \not\equiv 0 \mod p-1$$

(Kummer's congruence).

Proof. See theorems 4 and 5 of section 5.8 of Borevich and Shafarevich's book [1]. □

Corollary 1.36.

(1) $E_{p-1} \in M_{p-1}^p(\mathrm{SL}_2(\mathbb{Z}))$ *and* $\widetilde{E}_{p-1} = 1$.

(2) $E_{p+1} \in M_{p+1}^p(\mathrm{SL}_2(\mathbb{Z}))$ *and* $\widetilde{E}_{p+1} = \widetilde{P}$.

Proof. Recall that

$$E_{2k}(z) = 1 - 2\frac{2k}{B_{2k}}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)q^n\,.$$

Since $v_p\big(2\frac{p-1}{B_{p-1}}\big) = 1$, we get (1). On the other hand,

$$\frac{B_{p+1}}{p+1} \equiv \frac{B_2}{2} \equiv \frac{-1}{12} \not\equiv 0 \mod p\,.$$

Now (2) follows from this congruence and the fact that, by Fermat's little theorem, $\sigma_i(n) \equiv \sigma_j(n) \mod p$ if $i \equiv j \mod p-1$. □

Corollary 1.37. *The algebra $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$ is stable under θ.*

Proof. If $f \in \widetilde{M}_k^p(\mathrm{SL}_2(\mathbb{Z}))$, we can write

$$12\theta f = k\widetilde{P}f + \partial f = k\widetilde{E}_{p+1}f + \widetilde{E}_{p-1}\partial f$$

and observe that $\widetilde{E}_{p+1}f$ and $\widetilde{E}_{p-1}\partial f$ belong to $\widetilde{M}_{k+p+1}^p(\mathrm{SL}_2(\mathbb{Z}))$. □

Definition 1.38. We define $A, B \in \mathbb{Z}_{(p)}[X, Y]$ to be the unique polynomials satisfying that $A(Q, R) = E_{p-1}$ and $B(Q, R) = E_{p+1}$.

These are all the elements required to determine the structure of $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$. We are now in a position to prove the main theorem.

Lemma 1.39. *$\partial\widetilde{A} = \widetilde{B}$ and $\partial\widetilde{B} = -X\widetilde{A}$.*

Proof. Since $\widetilde{A}(\widetilde{Q}, \widetilde{R}) = \widetilde{E}_{p-1} = 1$, we see that $\theta\widetilde{A}(\widetilde{Q}, \widetilde{R}) = 0$ and so

$$\partial\widetilde{A}(\widetilde{Q}, \widetilde{R}) = \widetilde{P}\widetilde{A}(\widetilde{Q}, \widetilde{R}) = \widetilde{P} = \widetilde{E}_{p+1} = \widetilde{B}(\widetilde{Q}, \widetilde{R})\,.$$

That is to say, the modular form $\partial A(Q, R) - B(Q, R)$ lies in $pM_{p+1}^p(\mathrm{SL}_2(\mathbb{Z}))$ or, equivalently, $\partial A - B \in p\mathbb{Z}_{(p)}[X, Y]$. Therefore, $\partial\widetilde{A} = \widetilde{B}$.

Similarly,

$$\partial\widetilde{B}(\widetilde{Q}, \widetilde{R}) = (12\theta - \widetilde{P})\widetilde{P} = -\widetilde{Q} = -\widetilde{Q}\widetilde{A}(\widetilde{Q}, \widetilde{R})\,,$$

where the second inequality follows from theorem 1.33. Thus, the modular form $\partial B(Q, R) + QA(Q, R)$ lies in $pM_{p+3}^p(\mathrm{SL}_2(\mathbb{Z}))$ and so $\partial B + XA \in p\mathbb{Z}_{(p)}[X, Y]$. In conclusion, $\partial\widetilde{B} = -X\widetilde{A}$. □

Lemma 1.40. *The polynomial $\widetilde{A}$ has no repeated factors in $\overline{\mathbb{F}}_p[X, Y]$ and is prime to $\widetilde{B}$.*

Proof. Recall that $\widetilde{A}$ is an isobaric polynomial of weight $p - 1$, where X and Y have weights 4 and 6, respectively. Thus, the factors appearing in its decomposition must be of the form $X^3 - cY^2$ with $c \in \overline{\mathbb{F}}_p^\times$, X or Y.

Suppose, for the sake of contradiction, that the polynomial $\widetilde{A}$ is exactly divisible by $(X^3 - cY^2)^n$ for some $c \in \overline{F}_p^\times$ and some $n > 1$. Since $\widetilde{A}(\widetilde{Q}, \widetilde{R}) = 1$ and $\widetilde{Q}^3 - \widetilde{R}^2 \in q\mathbb{F}_p[\![q]\!]$, we see that $c \neq 1$. Then, $\partial(X^3 - cY^2) = 12(c - 1)X^2Y$ is prime to $X^3 - cY^2$ and so $(X^3 - cY^2)^{n-1}$ divides $\partial\widetilde{A} = \widetilde{B}$ exactly. In the same way, we see that $(X^3 - cY^2)^{n-2}$ divides $\partial\widetilde{B} = -X\widetilde{A}$ exactly, contradicting the hypothesis.

Analogously, if $\widetilde{A}$ is exactly divisible by X^n (resp. Y^n) for some $n > 1$, we deduce that $-X\widetilde{A}$ is exactly divisible by X^{n-2} (resp. Y^{n-2}) using that $\partial X = -4Y$ is prime to X (resp. $\partial Y = -6X^2$ is prime to Y), which is a contradiction.

We have seen that the factors of $\widetilde{A}$ have multiplicity $n = 1$ and that they appear with multiplicity $n - 1 = 0$ in $\partial\widetilde{A} = \widetilde{B}$ (i.e., they do not divide $\widetilde{B}$). □

Theorem 1.41. *The algebra $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$ of modular forms* mod p *is naturally isomorphic to $\mathbb{F}_p[X,Y]/(\widetilde{A}-1)$ and has a natural grading with values in $\mathbb{Z}/(p-1)\mathbb{Z}$.*

Proof. We have a surjective ring homomorphism

$$\begin{aligned}\mathbb{F}_p[X,Y] &\twoheadrightarrow \widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))\\ \phi(X,Y) &\mapsto \phi(\widetilde{Q},\widetilde{R})\end{aligned}$$

and we have to prove that its kernel $\mathfrak{a}$ is generated by $\widetilde{A}-1$. Since $\widetilde{A}(\widetilde{Q},\widetilde{R}) = 1$, it is clear that $(\widetilde{A}-1) \subseteq \mathfrak{a}$ and we only need to prove that the inclusion is in fact an equality.

Observe that $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$ is a subring of $\mathbb{F}_p[\![q]\!]$ and, in particular, an integral domain, which implies that $\mathfrak{a}$ is a prime ideal. However, $\mathfrak{a}$ is not maximal: if it were, $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$ would be finite by Hilbert's nullstellensatz, but theorem 1.15 gives arbitrarily many $\mathbb{F}_p$–linearly independent elements if we take large enough weights. Now, since the ring $\mathbb{F}_p[X,Y]$ has dimension 2, it suffices to prove that $(\widetilde{A}-1)$ is a prime ideal or, equivalently, that $\widetilde{A}-1$ is an irreducible polynomial.

Suppose, for the sake of contradiction, that $\widetilde{A}-1$ is not irreducible and let ϕ be one of its irreducible factors. Consider a decomposition $\phi = \phi_n + \phi_{n-1} + \cdots + \phi_0$, where ϕ_k is an isobaric polynomial of weight k (X and Y have weights 4 and 6, respectively) and $\phi_n \neq 0$. In particular, $n < p-1$ because $\phi \mid \widetilde{A}-1$. Take a primitive $(p-1)$–th root of unity ζ in $\overline{\mathbb{F}}_p$. Considering the weights, we see that $\widetilde{A}(\zeta^4X,\zeta^6Y) = \widetilde{A}(X,Y)$ but $\phi_n(\zeta^4X,\zeta^6Y) = \zeta^n\phi_n(X,Y) \neq \phi_n(X,Y)$. Thus, $\phi(X,Y)$ and $\phi(\zeta^4X,\zeta^6Y)$ are two distinct factors of $\widetilde{A}(X,Y)-1$. Therefore, we can write

$$\widetilde{A}(X,Y) - 1 = \phi(X,Y)\phi(\zeta^4X,\zeta^6Y)\psi(X,Y)$$

for some polynomial $\psi = \psi_m + \psi_{m-1} + \cdots + \psi_0$. Looking at the terms of highest weight in this equation, we find that

$$\widetilde{A}(X,Y) = \phi_n(X,Y)\phi_n(\zeta^4X,\zeta^6Y)\psi_m(X,Y) = \zeta^n\phi_n(X,Y)^2\psi_m(X,Y)$$

and this contradicts lemma 1.40.

Finally, since A has weight $p-1$, the grading on $M^p(\mathrm{SL}_2(\mathbb{Z}))$ induces a grading

$$\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z})) = \bigoplus_{\alpha \in \mathbb{Z}/(p-1)\mathbb{Z}} \left[\sum_{k \in \alpha} \widetilde{M}^p_k(\mathrm{SL}_2(\mathbb{Z})) \right].$$

□

Example 1.42. We list the first few cases.

(1) If $p = 5$, $E_{p-1} = Q$ and $A(X, Y) = X$, so $\widetilde{M}^5(\mathrm{SL}_2(\mathbb{Z})) \cong \mathbb{F}_5[Y]$.
(2) If $p = 7$, $E_{p-1} = R$ and $A(X, Y) = Y$, so $\widetilde{M}^7(\mathrm{SL}_2(\mathbb{Z})) \cong \mathbb{F}_7[X]$.
(3) If $p = 11$, $E_{p-1} = QR$ and $A(X, Y) = XY$, so $\widetilde{M}^{11}(\mathrm{SL}_2(\mathbb{Z})) \cong \mathbb{F}_{11}[X, X^{-1}]$.

As we have seen, the weights of the modular forms induce a grading on $\widetilde{M}^p(\mathrm{SL}_2(\mathbb{Z}))$. Even more, we can use weights to define a filtration.

Definition 1.43. Let $\alpha \in \mathbb{Z}/(p-1)\mathbb{Z}$ and consider a non-zero element

$$\widetilde{f} \in \sum_{k \in \alpha} \widetilde{M}^p_k(\mathrm{SL}_2(\mathbb{Z})).$$

By multiplying each summand by suitable powers of $\widetilde{A}(\widetilde{Q}, \widetilde{R})$, we can assume that $\widetilde{f} \in \widetilde{M}^p_k(\mathrm{SL}_2(\mathbb{Z}))$ for some $k \in \alpha$. We say that $\widetilde{f}$ is of *exact filtration* k if $\widetilde{f} \in \widetilde{M}^p_k(\mathrm{SL}_2(\mathbb{Z}))$ but $\widetilde{f} \notin \widetilde{M}^p_{k'}(\mathrm{SL}_2(\mathbb{Z}))$ for any $k' < k$; in this case, we write $w(\widetilde{f}) = k$. We make the convention that $w(0) = -\infty$.

Proposition 1.44. *Let $k \in \mathbb{Z}$ and let $f \in M^p_k(\mathrm{SL}_2(\mathbb{Z}))$ such that $\widetilde{f} \neq 0$. Consider the unique polynomial $\Phi \in \mathbb{Z}_{(p)}[X, Y]$ satisfying that $\Phi(Q, R) = f$.*

(1) *$w(\widetilde{f}) < k$ if and only if $\widetilde{A}$ divides $\widetilde{\Phi}$.*
(2) *$w(\theta\widetilde{f}) \leq w(\widetilde{f}) + p + 1$, with equality if and only if $w(\widetilde{f}) \not\equiv 0 \bmod p$.*

Proof. (1) is immediate from theorem 1.41. For (2), assume that $w(\widetilde{f}) = k$ and, as we saw in corollary 1.37, we can write

$$12\theta\widetilde{f} = \widetilde{A}(\widetilde{Q}, \widetilde{R})\partial\widetilde{f} + k\widetilde{B}(\widetilde{Q}, \widetilde{R})\widetilde{f}$$

and $A(Q, R)\partial f + kB(Q, R)f \in M^p_{k+p+1}(\mathrm{SL}_2(\mathbb{Z}))$. Hence, $w(\theta\widetilde{f}) \leq k + p + 1$. In addition, using (1) and that $\widetilde{A}$ and $\widetilde{B}$ are relatively prime (by lemma 1.40), we find that $\widetilde{A}$ divides $\widetilde{A}\partial\widetilde{\Phi} + k\widetilde{B}\widetilde{\Phi}$ if and only if $k = 0$ in $\mathbb{F}_p$. Again by (1), we have just proved that $w(\theta\widetilde{f}) = k + p + 1$ if and only if $k \equiv 0 \bmod p$. □

We finish with an interesting result which can be proved only with the tools from section 1.1.

Proposition 1.45. *Let $k \in \mathbb{N} \cup \{0\}$. Two elements $\widetilde{f_1}$ and $\widetilde{f_2}$ of $\widetilde{M}_k^p(\mathrm{SL}_2(\mathbb{Z}))$ are equal if and only if the coefficients of q^n in $\widetilde{f_1}$ and $\widetilde{f_2}$ are equal for every $n \leq \frac{k}{12}$.*

Proof. The condition is obviously necessary. Suppose that it holds. By theorem 1.15, we have a basis $g_0, g_1, \ldots, g_d$ of $M_k^p(\mathrm{SL}_2(\mathbb{Z}))$ with $d \leq \frac{k}{12}$. Moreover, in the proof of the theorem, we saw that the first coefficients of their q–expansions are $a_i(g_j) = 0$ for $i < j$ and $a_j(g_j) = 1$. Thus, the coefficients expressing $\widetilde{f_1} - \widetilde{f_2}$ in terms of the basis $\widetilde{g_0}, \widetilde{g_1}, \ldots, \widetilde{g_d}$ can be computed solving a triangular linear system of equations given by the coefficients of q^n for $n \leq d$, which are all 0. Therefore, $\widetilde{f_1} - \widetilde{f_2} = 0$. □

Chapter 2

Katz's theory of modular forms

In this chapter we present Katz's theory of modular forms. The approach of Katz is much more geometric than the one taken in chapter 1 and allows us to generalize the notion of modular forms to all kinds of base rings and work algebraically. In fact, this geometric context is also useful to define p–adic modular forms and other generalizations which we do not treat in this work.

Katz's modular forms are *rules* which assign values to elliptic curves with certain additional structures. One can then define q–expansions as the values assigned to a particular elliptic curve defined over the ring of power series on q, the Tate curve. As in the classical case, a modular form is uniquely determined by its q–expansions (although the proof in this case is more difficult). On the other hand, modular forms must satisfy some extra conditions which allow us to interpret them as *twisted functions* on elliptic curves: more precisely, they correspond to global sections of certain line bundles over elliptic curves. (A similar interpretation for classical modular forms is already hinted at in section 1.2.) It turns out that (in many cases) there is a universal elliptic curve from which all the other elliptic curves can be obtained and then modular forms correspond to global sections of some line bundle over the universal elliptic curve.

This chapter reproduces the main concepts explained in the first chapter of Katz's paper [8] and gives pointers to the proofs involving general elliptic curves and moduli spaces in Katz and Mazur's book [12], as explaining all that theory would take us too much afar from our objective of studying modular forms.

2.1 Definitions

We have seen in section 1.2 that (classical) modular forms (of level 1) can be interpreted as certain functions associating a complex number with each pair (E, ω), where E is an elliptic curve over $\mathbb{C}$ and ω is a nowhere vanishing holomorphic differential on E. The key fact is that the space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ parametrizes isomorphism classes of elliptic curves over $\mathbb{C}$. Katz vastly generalized this idea to define modular forms over arbitrary base schemes. Moreover, the geometric

viewpoint of Katz allows for further generalizations and constructions, some of which are presented in his article [8]. Here we only reproduce some basic definitions from section 1 of ibid.

Definition 2.1. An *elliptic curve* over a scheme S is a proper smooth morphism of schemes $\pi\colon E \to S$ whose geometric fibres are connected curves of genus 1 together with a section $e\colon S \to E$. We write $\underline{\omega}_{E/S} = \pi_*(\Omega^1_{E/S})$.

Remarks.

(1) The geometric fibres of $\pi\colon E \to S$ are elliptic curves over algebraically closed fields in the usual sense.

(2) As is the case with elliptic curves defined over algebraically closed fields, the S–scheme E admits a unique structure of abelian group scheme for which e is the identity section. See theorems 2.1.2 and 2.5.1 of Katz and Mazur's book [12] for the details.

(3) The invertible sheaf $\Omega^1_{E/S}$ is fibrewise of degree 0 and Serre–Grothendieck duality defines a canonical trace map $\mathrm{Tr}\colon \mathrm{R}^1\pi_*(\Omega^1_{E/S}) \to \mathscr{O}_S$ which is an isomorphism compatible with arbitrary base change. Therefore, $\underline{\omega}_{E/S}$ is an invertible sheaf whose formation is compatible with arbitrary base change and is dual to $\mathrm{R}^1\pi_*(\mathscr{O}_E)$. In particular, we can find a basis ω for $\underline{\omega}_{E/S}$ locally on S.

Fix $N \in \mathbb{N}$. Later we define modular forms for the group

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}.$$

One can check that $\Gamma(N)\backslash\mathbb{H}$ parametrizes isomorphism classes of pairs (E, α), where E is an elliptic curve over $\mathbb{C}$ and $\alpha = (\alpha_1, \alpha_2)$ is a basis of the N–torsion subgroup $E[N]$ of E with the property that $e_n(\alpha_1, \alpha_2) = e^{2\pi i/N}$ for the Weil pairing $e_n\colon E[N] \times E[N] \to \mu_N$. Recall that $E[N]$ is the kernel of the homomorphism $[N]\colon E \to E$ given by multiplication by N and that (over any algebraically closed field in which N is invertible) $E[N]$ is a free $(\mathbb{Z}/N\mathbb{Z})$–module of rank 2.

More generally, given an elliptic curve E over an arbitrary scheme S, we can still consider the homomorphism $[N]\colon E \to E$ given by multiplication by N and its kernel $E[N]$. In general, $E[N]$ is a finite flat abelian group scheme of order N^2 over S. Moreover, $E[N]$ is étale over S if and only if N is invertible in $\mathrm{H}^0(S, \mathscr{O}_S)$ or, equivalently, S is a $\mathbb{Z}\left[\frac{1}{N}\right]$–scheme. In that case, there exists a finite étale covering S' of S such that $E_{S'}[N]$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})_{S'} \times_{S'} (\mathbb{Z}/N\mathbb{Z})_{S'}$

over S', where $(\mathbb{Z}/N\mathbb{Z})_{S'}$ is the constant cyclic group scheme of order N over S'. (See theorem 2.3.1 of Katz and Mazur's book [12] for a proof of these claims about the structure of $E[N]$.)

Definition 2.2. Let $N \in \mathbb{N}$. A *level* $\Gamma(N)$*–structure* on an elliptic curve E over a scheme S is an isomorphism $\alpha_N \colon E[N] \to (\mathbb{Z}/N\mathbb{Z})_S \times_S (\mathbb{Z}/N\mathbb{Z})_S$ of group schemes over S.

Definition 2.3. Let $N \in \mathbb{N}$ and $k \in \mathbb{Z}$. A *modular form* for $\Gamma(N)$ of weight k is a *rule* f which assigns to each triple $(E/R, \omega, \alpha_N)$, consisting of an elliptic curve E over a ring R together with a basis ω of $\mathrm{H}^0(\mathrm{Spec}(R), \underline{\omega}_{E/R})$ and a level $\Gamma(N)$–structure α_N on E, an element $f(E/R, \omega, \alpha_N) \in R$ satisfying the following conditions:

(i) $f(E/R, \omega, \alpha_N)$ depends only on the R–isomorphism class of the triple $(E/R, \omega, \alpha_N)$;

(ii) $f(E_{R'}/R', \omega_{R'}, \alpha_{N,R'}) = g(f(E/R, \omega, \alpha_N))$ for every ring homomorphism $g \colon R \to R'$ (i.e., f commutes with base change), and

(iii) $f(E/R, \lambda\omega, \alpha_N) = \lambda^{-k} f(E/R, \omega, \alpha_N)$ for all $\lambda \in R^\times$ (i.e., f is homogeneous of degree $-k$ in the *second variable*).

Remarks.

(1) Observe that condition (iii) implies that $f(E/R, \omega, \alpha_N)\omega^{\otimes k}$ is a global section of $\underline{\omega}_{E/R}^{\otimes k}$ independent of the choice of ω. Thus, we can define a *modular form* for $\Gamma(N)$ of weight k alternatively as a *rule* f which assigns to each pair $(E/S, \alpha_N)$, consisting of an elliptic curve E over a scheme S and a level $\Gamma(N)$–structure α_N on E, an element $f(E/S, \alpha_N) \in \mathrm{H}^0(S, \underline{\omega}_{E/S}^{\otimes k})$ satisfying the following conditions:
   (i) $f(E/S, \alpha_N)$ depends only on the S–isomorphism class of the pair $(E/S, \alpha_N)$, and
   (ii) $f(E_{S'}/S', g^*(\alpha_N)) = g^*(f(E/S, \alpha_N))$ for every morphism $g \colon S' \to S$ of schemes (i.e., f commutes with base change).

(2) If $N = 1$, we omit the level structure because it is trivial.

(3) If $M \mid N$, every level $\Gamma(N)$–structure induces a level $\Gamma(M)$–structure in the obvious way. Hence, we can view modular forms for $\Gamma(M)$ as modular forms for $\Gamma(N)$.

(4) If we consider only base rings R (or base schemes S) defined over a fixed ring R_0 and only base changes by R_0–morphisms, we obtain the notion of a

modular form for $\Gamma(N)$ of weight k defined over R_0; we define $F(R_0;\Gamma(N),k)$ to be the R_0–module of such modular forms.

The previous definition of modular forms is analogous to the classical one except that we have not required any *condition at* ∞ yet.

2.2 The Tate curve and q–expansions

In this section we introduce Tate's curve via computations over the complex numbers, following section A1.2 of Katz's article [8]. This curve is then used to define the q–expansions of a modular form.

Fix $\tau \in \mathbb{H}$ and write $q_\tau = e^{2\pi i\tau}$. As in the beginning of section 1.2, we have an elliptic curve E_τ over $\mathbb{C}$ corresponding to $\mathbb{C}/\Lambda(\tau)$, where $\Lambda(\tau) = \mathbb{Z}\tau \oplus \mathbb{Z}$, defined by the affine equation

$$E_\tau\colon \widetilde{y}^2 = 4\widetilde{x}^3 - 60G_4(\tau)\widetilde{x} - 140G_6(\tau) = 4\widetilde{x}^3 - \frac{(2\pi i)^4}{12}E_4(\tau)\widetilde{x} - \frac{(2\pi i)^6}{216}E_6(\tau)$$

(see example 1.5). The discriminant of E_τ is

$$\Big(\frac{(2\pi i)^4}{12}E_4(\tau)\Big)^3 - 27\Big(\frac{(2\pi i)^6}{216}E_6(\tau)\Big)^2 = (2\pi)^{12}\Delta(\tau)$$

and it has a nowhere vanishing holomorphic differential $\frac{d\widetilde{x}}{\widetilde{y}}$ corresponding to dz.

Using the q–expansions of E_4 and E_6, we obtain an equation in $\mathbb{C}[\![q_\tau]\!]$ for E_τ. We regard q as a formal variable and consider the elliptic curve

$$E\colon \widetilde{y}^2 = 4\widetilde{x}^3 - \frac{(2\pi i)^4}{12}\Big(1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n\Big)\widetilde{x} - \frac{(2\pi i)^6}{216}\Big(1 - 504\sum_{n=1}^{\infty}\sigma_5(n)q^n\Big)$$

(see example 1.5) defined over $\mathbb{C}(\!(q)\!)$ with discriminant

$$(2\pi)^{12}q\prod_{n=1}^{\infty}(1-q^n)^{24}$$

(see corollary 1.13). Observe that the discriminant is invertible in $\mathbb{C}(\!(q)\!)$ but not in $\mathbb{C}[\![q]\!]$, even if the equation of E is defined over $\mathbb{C}[\![q]\!]$.

On the other hand, there is an analytic isomorphism

$$\mathbb{C}/\Lambda(\tau) \to \mathbb{C}^\times/q_\tau^{\mathbb{Z}}$$

$$z \mapsto t = e^{2\pi i z}$$

(where $q_\tau^{\mathbb{Z}}$ is the subgroup of $\mathbb{C}^\times$ generated by q_τ). We can express $\widetilde{x} = \wp(z;\Lambda(\tau))$ and $\widetilde{y} = \wp'(z;\Lambda(\tau))$ in terms of q_τ and t.

Proposition 2.4. *We have the following identities:*

$$\frac{1}{(2\pi i)^2}\wp(z;\Lambda(\tau)) = \sum_{n\in\mathbb{Z}} \frac{q_\tau^n t}{(1-q_\tau^n t)^2} + \frac{1}{12} - 2\sum_{n=1}^{\infty} \frac{q_\tau^n}{(1-q_\tau^n)^2}\,,$$
$$\frac{1}{(2\pi i)^3}\wp'(z;\Lambda(\tau)) = \sum_{n\in\mathbb{Z}} \frac{q_\tau^n t(1+q_\tau^n t)}{(1-q_\tau^n t)^3}\,.$$

Proof. See lemma 6.1 and theorem 6.2 of chapter I of Silverman's book [19]. □

It is convenient to make a change of variables to remove the powers of $2\pi i$ and the denominators in the formulae above. More precisely, we set

$$\frac{1}{(2\pi i)^2}\widetilde{x} = x + \frac{1}{12} \quad \text{and} \quad \frac{1}{(2\pi i)^3}\widetilde{y} = 2y + x$$

and, substituting in the equation of the elliptic curve above and clearing common factors, we obtain the equation

$$E_\tau\colon y^2 + xy = x^3 + a_4(q_\tau)x + a_6(q_\tau)$$

with

$$a_4(q_\tau) = \frac{1-E_4(\tau)}{48} = -5\Big(\frac{E_4(\tau)-1}{240}\Big) = \sum_{n=1}^{\infty} -5\sigma_3(n)q_\tau^n\,,$$
$$a_6(q_\tau) = \frac{1-3E_4(\tau)-2E_6(\tau)}{1728} = -\frac{5}{12}\Big(\frac{E_4(\tau)-1}{240}\Big) - \frac{7}{12}\Big(\frac{E_6(\tau)-1}{-504}\Big)$$
$$= \sum_{n=1}^{\infty} \frac{-5\sigma_3(n)-7\sigma_5(n)}{12} q_\tau^n\,.$$

(See example 1.5 for the formulae for E_4 and E_6.) Observe that the coefficients of q_τ^n, $n \in \mathbb{N}$, in the previous equation are integers because $d^3 \equiv d^5 \bmod 12$ for every $d \in \mathbb{Z}$. The discriminant of E_τ is

$$-a_6(q_\tau) + a_4(q_\tau)^2 + 72a_4(q_\tau)a_6(q_\tau) - 64a_4(q_\tau)^3 - 432a_6(q_\tau)^2 = \cdots = \Delta(\tau)$$

and it has a nowhere vanishing holomorphic differential $\omega = \frac{dx}{2y+x}$. Moreover,

we can write in terms of q_τ and t

$$x = \frac{1}{(2\pi i)^2}\widetilde{x} - \frac{1}{12} = \sum_{n\in\mathbb{Z}} \frac{q_\tau^n t}{(1-q_\tau^n t)^2} - 2\sum_{n=1}^{\infty} \frac{q_\tau^n}{(1-q_\tau^n)^2}\,,$$
$$y = \frac{1}{2(2\pi i)^3}\widetilde{y} - \frac{1}{2}x = \sum_{n\in\mathbb{Z}} \frac{(q_\tau^n t)^2}{(1-q_\tau^n t)^3} + \sum_{n=1}^{\infty} \frac{q_\tau^n}{(1-q_\tau^n)^2}\,,$$
$$\omega = \frac{dx}{2y+x} = (2\pi i)\frac{d\widetilde{x}}{\widetilde{y}} = 2\pi i\, dz = \frac{dt}{t}\,.$$

Regarding q as a formal variable, we obtain the Tate curve.

Definition 2.5. The *Tate curve* is the elliptic curve over $\mathbb{Z}((q))$ given by the affine equation

$$\mathrm{Tate}(q)\colon y^2 = x^3 + a_4(q)x + a_6(q)\,,$$

where

$$a_4(q) = \sum_{n=1}^{\infty} -5\sigma_3(n)q^n \quad \text{and} \quad a_6(q) = \sum_{n=1}^{\infty} \frac{-5\sigma_3(n) - 7\sigma_5(n)}{12}q^n\,,$$

together with its *canonical differential*

$$\omega_{\mathrm{can}} = \frac{dx}{2y+x}\,.$$

Remark. The equation defining $\mathrm{Tate}(q)$ has coefficients in $\mathbb{Z}[\![q]\!]$. However, the discriminant of $\mathrm{Tate}(q)$ is

$$\Delta(q) = q\prod_{n=1}^{\infty}(1-q^n)^{24}\,,$$

which is invertible in $\mathbb{Z}((q))$ but not in $\mathbb{Z}[\![q]\!]$. In addition, its j–invariant is

$$j(q) = \frac{1}{q} + 744 + 196884q + \cdots \in \frac{1}{q}\mathbb{Z}[\![q]\!]\,.$$

We use the Tate curve to define the q–expansions of modular forms, following Katz. But first let us briefly explain the relation between the Tate curve and the q–expansion of a modular form in the sense of section 1.1.

Let R be the subring of $\mathbb{C}((q))$ consisting of Laurent series of functions which are holomorphic on $\{\, q \in \mathbb{C} : 0 < |q| < 1 \,\}$. The equation defining $\mathrm{Tate}(q)$ defines also an elliptic curve over R, which abusing notation we call again

$\mathrm{Tate}(q)$. Take $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$. The q–expansion of f is an element $\widehat{f} \in R$ which, in Katz's definition of modular forms, corresponds to $f(\mathrm{Tate}(q), \omega_{\mathrm{can}})$. Moreover, the morphism of $\mathbb{C}$–algebras $q \mapsto q_\tau = e^{2\pi i \tau} : R \to \mathbb{C}$ yields a cartesian diagram

$$\begin{array}{ccc} E_\tau & \longrightarrow & \mathrm{Tate}(q) \\ \downarrow & \ulcorner & \downarrow \\ \mathrm{Spec}(\mathbb{C}) & \longrightarrow & \mathrm{Spec}(R) \end{array}$$

and in this situation the condition that f commutes with base change is just that $f(\tau) = f(E_\tau, \omega) = \widehat{f}(q_\tau)$ (cf. definition 1.16), which is the equation defining the q–expansion in the classical case. On the other hand, the natural inclusion $R \hookrightarrow \mathbb{C}((q))$ yields another cartesian square

$$\begin{array}{ccc} \mathrm{Tate}(q) \otimes_{\mathbb{Z}((q))} \mathbb{C}((q)) & \longrightarrow & \mathrm{Tate}(q) \\ \downarrow & \ulcorner & \downarrow \\ \mathrm{Spec}(\mathbb{C}((q))) & \longrightarrow & \mathrm{Spec}(R) \end{array}$$

which allows us to view $\widehat{f}$ as a formal series in $\mathbb{C}((q))$.

Let $N \in \mathbb{N}$. We have to take into consideration some level $\Gamma(N)$–structures. Consider a primitive N–th root of unity ζ_N. We observe that, via the isomorphism $E_\tau \cong \mathbb{C}^\times / q_\tau^{\mathbb{Z}}$, the subgroup $E_\tau[N]$ corresponds to $\{\, \zeta_N^i q_\tau^{j/N} : 0 \leq i, j \leq N-1 \,\}$. Replacing t with $\zeta_N^i q_\tau^{j/N}$ in the previous formulae for x and y, we see that this point has coordinates in $\mathbb{Z}[\![q_\tau^{1/N}]\!] \otimes_{\mathbb{Z}} \mathbb{Z}[\zeta_N]$. On the other hand, $\mathrm{Tate}(q)[N]$ is étale over the base ring if N is invertible in it, in which case we can think of the group scheme $\mathrm{Tate}(q)[N]$ (of order N^2) in terms of its points. All in all, the level $\Gamma(N)$–structures on $\mathrm{Tate}(q)$ are defined over $\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} \mathbb{Z}\left[\zeta_N, \frac{1}{N}\right]$. (Abusing notation, we write $\mathrm{Tate}(q)$ for any base change of the Tate curve.)

Definition 2.6. Let $N \in \mathbb{N}$ and $k \in \mathbb{Z}$. Let f be a modular form for $\Gamma(N)$ of weight k defined over a ring R_0 which contains N^{-1} and a primitive N–th root of unity ζ_N. For each level $\Gamma(N)$–structure α_N on the curve $\mathrm{Tate}(q)$ over $\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} R_0$, the *$q$–expansion* of f at α_N is the Laurent series

$$\widehat{f}_{\alpha_N}(q) = f(\mathrm{Tate}(q)/(\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} R_0), \omega_{\mathrm{can}}, \alpha_N) \in \mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} R_0\,.$$

Definition 2.7. Let $N \in \mathbb{N}$ and $k \in \mathbb{Z}$. Consider a ring R_0 in which N is not nilpotent and let ζ_N be a primitive N–th root of unity. We say that $f \in F(R_0; \Gamma(N), k)$ is

holomorphic at ∞ (resp. is a *cusp form*) if its image in $F\big(R_0\big[\zeta_N, \frac{1}{N}\big]; \Gamma(N), k\big)$ has all its q–expansions in $\mathbb{Z}[\![q^{1/N}]\!] \otimes_{\mathbb{Z}} R_0\big[\zeta_N, \frac{1}{N}\big]$ (resp. in $q^{1/N}\mathbb{Z}[\![q^{1/N}]\!] \otimes_{\mathbb{Z}} R_0\big[\zeta_N, \frac{1}{N}\big]$). We write $M(R_0; \Gamma(N), k)$ (resp. $S(R_0; \Gamma(N), k)$) for the R_0–module of modular forms for $\Gamma(N)$ of weight k defined over R_0 which are holomorphic at ∞ (resp. cusp forms for $\Gamma(N)$ of weight k defined over R_0).

2.3 The modular curves

In the previous section, we have encountered an elliptic curve *parametrizing* all the elliptic curves over $\mathbb{C}$. We want to generalize this in order to redefine modular forms in a more geometric way. To do this, one needs to solve some moduli problems.

In this section we briefly introduce these moduli problems and state the main representability results without proof. The properties we use later are stated in sections 1.4 and 1.5 of Katz's article [8]. The general theory in a much more comprehensive way (and the proofs) can be found in Katz and Mazur's book [12].

Definition 2.8. Let R_0 be a ring. We define $\mathrm{Ell}\,/R_0$ to be the category

(i) whose objects are elliptic curves E/S, where S is an R_0–scheme, and

(ii) whose morphisms from E'/S' to E/S are cartesian squares

$$\begin{array}{ccc} E' & \xrightarrow{\varphi} & E \\ \downarrow{\scriptstyle \pi'} & \ulcorner & \downarrow{\scriptstyle \pi} \\ S' & \xrightarrow{f} & S \end{array}$$

of R_0–morphisms (i.e., the induced morphism $(\varphi, \pi') \colon E' \to E \times_S S'$ is an isomorphism).

Definition 2.9. Let R_0 be a ring. A *moduli problem for elliptic curves* over R_0 is a functor $\mathcal{P} \colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$. Given $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$, the elements of $\mathcal{P}(E/S)$ are called *level $\mathcal{P}$–structures*.

We say that $\mathcal{P}$ is *representable* if there exists $\mathbb{E}/\mathfrak{M}(\mathcal{P}) \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$ such that $\mathcal{P} \cong \mathrm{Hom}_{\mathrm{Ell}\,/R_0}(\,\cdot\,, \mathbb{E}/\mathfrak{M}(\mathcal{P}))$. We say that $\mathcal{P}$ is *relatively representable* if, for every $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$, the functor $\mathcal{P}_{E/S} \colon (\mathrm{Sch}\,/S)^{\mathrm{op}} \to \mathrm{Set}$ given by $\mathcal{P}_{E/S}(T) = \mathcal{P}(E_T/T)$, where $E_T = E \times_S T$, is represented by an S–scheme $\mathcal{P}_{E/S}$. (We use the same notation for the scheme and the functor it represents.)

Lemma 2.10. *Let R_0 be a ring and let $\mathcal{P}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ be a moduli problem. If $\mathcal{P}$ is represented by $\mathbb{E}/\mathfrak{M}(\mathcal{P})$, then the R_0–scheme $\mathfrak{M}(\mathcal{P})$ represents the functor $\widetilde{\mathcal{P}}\colon (\mathrm{Sch}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ given by*

$$\widetilde{\mathcal{P}}(S) = \{\, [(E/S, \alpha)] : E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0) \text{ and } \alpha \in \mathcal{P}(E/S) \,\},$$

where $[(E/S, \alpha)]$ is the S–isomorphism class of the pair $(E/S, \alpha)$.

Proof. Given an R_0–morphism $f\colon S \to \mathfrak{M}(\mathcal{P})$, we can form the fibre product

$$\begin{array}{ccc} E = \mathbb{E} \times_{\mathfrak{M}(\mathcal{P})} S & \xrightarrow{\varphi} & \mathbb{E} \\ \downarrow{\scriptstyle \pi} & \ulcorner & \downarrow \\ S & \xrightarrow{f} & \mathfrak{M}(\mathcal{P}) \end{array}$$

and we obtain $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$ defined up to S–isomorphism. Moreover, $(\varphi, f) \in \mathrm{Hom}_{\mathrm{Ell}\,/R_0}(E/S, \mathbb{E}/\mathfrak{M}(\mathcal{P}))$ yields $\alpha \in \mathcal{P}(E/S)$, by hypothesis. □

Remark. In particular, setting $S = \mathfrak{M}(\mathcal{P})$ and $f = \mathrm{id}_{\mathfrak{M}(\mathcal{P})}$ in the previous proof, we obtain a *universal pair* $(\mathbb{E}/\mathfrak{M}(\mathcal{P}), \alpha_{\mathrm{univ}})$.

Throughout the rest of this section, let $N \in \mathbb{N}$. Let R_0 be a ring in which N is not nilpotent. Our objective is to study the representability of the moduli problem $\Gamma(N)_{R_0}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ given by

$$\Gamma(N)_{R_0}(E/S) = \{\, \alpha_N \text{ level } \Gamma(N)\text{–structure on } E/S \,\}$$

(here, we refer to level $\Gamma(N)$–structures in the sense of definition 2.2). We write $\Gamma(N) = \Gamma(N)_{\mathbb{Z}}$.

Theorem 2.11 (relative representability). *Let E/S be an elliptic curve. The functor $\Gamma(N)_{E/S}$ is represented by a finite S–scheme. If, in addition, S is a $\mathbb{Z}\left[\frac{1}{N}\right]$–scheme, then $\Gamma(N)_{E/S}$ is represented by a finite étale S–scheme.*

Proof. See proposition 1.6.5 and theorems 3.6.0 and 3.7.1 of Katz and Mazur's book [12]. □

Definition 2.12. Let R_0 be a ring and let P be a property of morphisms of schemes. We say that a moduli problem $\mathcal{P}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ *has property P* if it is relatively representable and, for every $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$, the morphism of schemes $\mathcal{P}_{E/S} \to S$ has property P.

Corollary 2.13. *The moduli problem $\Gamma(N)_{\mathbb{Z}[1/N]}$ is finite and étale.*

Definition 2.14. Let R_0 be a ring. A moduli problem $\mathcal{P}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ is called *rigid* if, for every $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$ and every $\alpha \in \mathcal{P}(E/S)$, the pair $(E/S, \alpha)$ has no non-trivial automorphisms (i.e., the group $\mathrm{Aut}(E/S)$ acts freely on $\mathcal{P}(E/S)$).

Theorem 2.15 (rigidity). *Let S be a connected scheme and let $E/S \in \mathrm{Ob}(\mathrm{Ell}\,/\mathbb{Z})$. Let $\varphi\colon E \to E$ be an automorphism of E over S. If $N \geq 3$ and φ restricts to the identity on $E[N]$, then $\varphi = \mathrm{id}_E$.*

Proof. See corollary 2.7.2 of Katz and Mazur's book [12]. □

Theorem 2.16. *Let R_0 be a ring and let $\mathcal{P}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ be a moduli problem which is relatively representable and affine. The moduli problem $\mathcal{P}$ is representable if and only if it is rigid. In that case, the R_0–scheme $\mathfrak{M}(\mathcal{P})$ representing $\widetilde{\mathcal{P}}$ (see lemma 2.10) is affine.*

Proof. See theorem 4.7.0 of Katz and Mazur's book [12]. □

Theorem 2.17. *Let R_0 be a ring and let $\mathcal{P}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ be a moduli problem. If $\mathcal{P}$ is relatively representable, rigid, affine and étale, then it is representable by some $\mathbb{E}/\mathfrak{M}(\mathcal{P}) \in \mathrm{Ob}(\mathrm{Ell}\,/R_0)$ and $\mathfrak{M}(\mathcal{P})$ is a smooth affine curve over R_0.*

Proof. See corollary 4.7.1 of Katz and Mazur's book [12]. □

Corollary 2.18. *If $N \geq 3$, the moduli problem $\Gamma(N)_{\mathbb{Z}[1/N]}$ is representable by some $\mathbb{E}/\mathfrak{M}(\Gamma(N)) \in \mathrm{Ob}\big(\mathrm{Ell}\,/\mathbb{Z}\big[\frac{1}{N}\big]\big)$ and $\mathfrak{M}(\Gamma(N))$ is a smooth affine curve over $\mathbb{Z}\big[\frac{1}{N}\big]$. Furthermore, over a $\mathbb{Z}\big[\frac{1}{N}\big]$–algebra R_0, the moduli problem $\Gamma(N)_{R_0}$ is representable by $\mathbb{E}_{R_0}/\mathfrak{M}(\Gamma(N))_{R_0}$, where $\cdot_{R_0}$ is obtained from $\cdot$ by the base change $\mathbb{Z}\big[\frac{1}{N}\big] \to R_0$.*

Definition 2.19. For $N \geq 3$, consider the universal pair $(\mathbb{E}/\mathfrak{M}(\Gamma(N)), \alpha_{\mathrm{univ}})$ for the moduli problem $\Gamma(N)_{\mathbb{Z}[1/N]}$ (see the remark after lemma 2.10). We write $Y(N)$ for the smooth affine curve $\mathfrak{M}(\Gamma(N))$ over $\mathbb{Z}\big[\frac{1}{N}\big]$ and call it the *modular curve* (without cusps) for $\Gamma(N)$. We call $\mathbb{E}/Y(N)$ the *universal elliptic curve* for $\Gamma(N)$ and α_{univ} its *universal level $\Gamma(N)$–structure*.

Suppose for the rest of this section that $N \geq 3$. Since modular forms commute with base change and every elliptic curve with a level $\Gamma(N)$–structure can be obtained as a pull-back of the universal elliptic curve, we see that every $f \in F\big(\mathbb{Z}\big[\frac{1}{N}\big]; \Gamma(N), k\big)$ for some $k \in \mathbb{Z}$ is uniquely determined by its value $f(\mathbb{E}/Y(N), \alpha_{\mathrm{univ}}) \in \mathrm{H}^0(Y(N), \underline{\omega}^{\otimes k}_{\mathbb{E}/Y(N)})$. Thus we obtain an alternative definition of modular forms which can be generalized in the following way.

Definition 2.20. Let K be a $\mathbb{Z}\left[\frac{1}{N}\right]$–module and let $k \in \mathbb{Z}$. A *modular form* for $\Gamma(N)$ of weight k with coefficients in K is an element of $\mathrm{H}^0(Y(N), \underline{\omega}^{\otimes k}_{\mathbb{E}/Y(N)} \otimes_{\mathbb{Z}[1/N]} K)$ (cf. definition 2.3). We also write $F(K;\Gamma(N),k) = \mathrm{H}^0(Y(N), \underline{\omega}^{\otimes k}_{\mathbb{E}/Y(N)} \otimes_{\mathbb{Z}[1/N]} K)$ and

$$F(K;\Gamma(N)) = \bigoplus_{k' \in \mathbb{Z}} F(K;\Gamma(N),k')\,.$$

Remark. Let $f \in F(K;\Gamma(N),k)$. Consider an elliptic curve E/S with a level $\Gamma(N)$–structure α_N. There is a cartesian diagram

$$\begin{array}{ccc} E & \longrightarrow & \mathbb{E} \\ \downarrow & \ulcorner & \downarrow \\ S & \xrightarrow{g} & Y(N) \end{array}$$

such that $(E/S,\alpha_N) = g^*(\mathbb{E}/Y(N),\alpha_{\mathrm{univ}})$ and g is unique with this property. We define $f(E/S,\alpha_N) = g^*(f) \in \mathrm{H}^0(S, \underline{\omega}^{\otimes k}_{E/S} \otimes_{\mathbb{Z}[1/N]} K)$. This agrees with definition 2.3 when K is a ring. In particular, we can define the *q–expansions* of f in the same way as in definition 2.6.

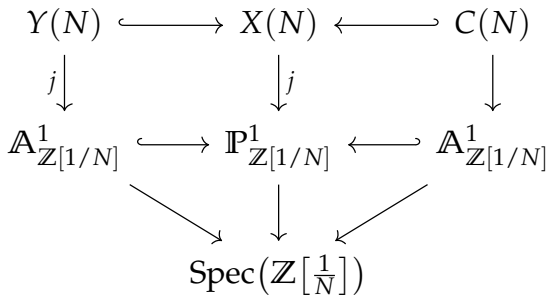
In definition 2.20, we use exactly the same notation as in definition 2.3 because the two notions of modular form agree when both make sense (i.e., when $N \geq 3$ and K is a ring). In this case, we can use the two definitions interchangeably and this sometimes gives us greater insight. For this reason, we would like to find a similar alternative definition for modular forms which are holomorphic at ∞. To this aim, we need to *extend* the sheaf $\underline{\omega}_{\mathbb{E}/Y(N)}$.

There is a morphism of $\mathbb{Z}\left[\frac{1}{N}\right]$–schemes $j\colon Y(N) \to \mathbb{A}^1_{\mathbb{Z}[1/N]}$ defined on points as follows. Consider a $\mathbb{Z}\left[\frac{1}{N}\right]$–algebra R and take $P \in Y(N)(R)$. By lemma 2.10, P corresponds to some isomorphism class $[(E/R,\alpha_N)]$, where E/R is an elliptic curve and α_N is a level $\Gamma(N)$–structure. Then, $j(P) \in \mathbb{A}^1_{\mathbb{Z}[1/N]}(R)$ is given by the j–invariant of E/R (i.e., the morphism of $\mathbb{Z}\left[\frac{1}{N}\right]$–algebras $\mathbb{Z}\left[\frac{1}{N}\right][j] \to R$ which maps j to the j–invariant of E/R, where j denotes the variable defining $\mathbb{A}^1_{\mathbb{Z}[1/N]}$). The morphism j is finite and flat.

Moreover, the affine j–line $\mathbb{A}^1_{\mathbb{Z}[1/N]}$ can be canonically embedded in the projective j–line $\mathbb{P}^1_{\mathbb{Z}[1/N]} = \mathrm{Proj}\big(\mathbb{Z}\left[\frac{1}{N}\right][j]\big)$. After composition, we obtain a morphism $Y(N) \to \mathbb{P}^1_{\mathbb{Z}[1/N]}$.

Definition 2.21. The *modular curve* (with cusps) $X(N)$ over $\mathbb{Z}\left[\frac{1}{N}\right]$ is the normalization of $\mathbb{P}^1_{\mathbb{Z}[1/N]}$ in $Y(N)$.

We obtain a commutative diagram



where the affine line $\mathbb{A}^1_{\mathbb{Z}[1/N]}$ in the last column is $\mathrm{Spec}\big(\mathbb{Z}\big[\frac{1}{N}\big][j^{-1}]\big)$ and $C(N)$ is the closed subscheme $X(N) \setminus Y(N)$ of $X(N)$.

Definition 2.22. The $\mathbb{Z}\big[\frac{1}{N}\big]$–scheme $C(N) = X(N) \setminus Y(N)$ is called the scheme of *cusps*.

Theorem 2.23. *The modular scheme $X(N)$ is a proper smooth curve over $\mathbb{Z}\big[\frac{1}{N}\big]$ and its closed subscheme $C(N)$ is finite étale over $\mathbb{Z}\big[\frac{1}{N}\big]$. Moreover, locally in a neighbourhood of the cusps, the invertible sheaf $\Omega^1_{X(N)/Z[1/N]}(\log(C(N)))$ of differential 1–forms with at worst simple poles along the cusps has a basis $\frac{dj^{-1}}{j^{-1}}$.*

Proof. See theorem 8.6.8 and corollary 10.9.2 of Katz and Mazur's book [12]. □

The modular curve $X(N)$ can be defined in an alternative way. There is a notion of generalized elliptic curves and one can define a moduli problem for generalized elliptic curves analogous to $\Gamma(N)$. It turns out that such a moduli problem is representable over $\mathbb{Z}\big[\frac{1}{N}\big]$ by a universal generalized elliptic curve over a base scheme which coincides with $X(N)$. This is the approach taken in Deligne and Rapoport's article [5], even though that article makes use of algebraic spaces, stacks and other mathematical tools the author of this work is not familiar with.

Let R_0 be a $\mathbb{Z}\big[\frac{1}{N}\big]$–algebra. Let S be an R_0–scheme and consider an elliptic curve E/S with a level $\Gamma(N)$–structure α_N. Observe that the $\mathbb{Z}/N\mathbb{Z}$–module $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ has a canonical basis $((1,0),(0,1))$ and so the isomorphism $\alpha_N \colon E[N] \to (\mathbb{Z}/N\mathbb{Z})_S \times_S (\mathbb{Z}/N\mathbb{Z})_S$ induces locally on S a canonical basis (P,Q) of $E[N]$. Moreover, we have the Weil pairing

$$e_N \colon E[N] \times_S E[N] \to \mu_{N,S}$$

(see section 2.8 of Katz and Mazur's book [12]) and $e_N(P,Q)$ is a primitive N–th root of unity (see theorem 5.6.3 of ibid.) which we call the *determinant* of α_N. This definition can also be extended to generalized elliptic curves.

Now assume that R_0 contains a primitive N–th root of unity ζ_N (e.g., if $R_0 = \mathbb{Z}\left[\frac{1}{N}, \zeta_N\right]$). We have a moduli problem $\Gamma(N)_{R_0,\zeta_N}\colon (\mathrm{Ell}\,/R_0)^{\mathrm{op}} \to \mathrm{Set}$ given by

$$\Gamma(N)_{R_0,\zeta_N}(E/S) = \{\, \alpha_N \text{ level } \Gamma(N)\text{–structure of determinant } \zeta_N \text{ on } E/S \,\}$$

(see paragraph 9.4.3.1 of Katz and Mazur's book [12]). This moduli problem is representable by $\mathbb{E}_{R_0,\zeta_N}/Y(N)_{R_0,\zeta_N}$, where $Y(N)_{R_0,\zeta_N}$ is a closed subscheme of $Y(N)_{R_0}$ (the locus of points corresponding to the level $\Gamma(N)$–structures of determinant ζ_N) and $\mathbb{E}_{R_0,\zeta_N}$ is the pull-back of $\mathbb{E}_{R_0}$ by the closed immersion $Y(N)_{R_0,\zeta_N} \hookrightarrow Y(N)_{R_0}$ (see proposition 9.1.7 of ibid.). In particular, we write $Y(N)_{\zeta_N} = Y(N)_{\mathbb{Z}[1/N,\zeta_N],\zeta_N}$ and $\mathbb{E}_{\zeta_N} = \mathbb{E}_{\mathbb{Z}[1/N,\zeta_N],\zeta_N}$. Analogously, one can define a closed subscheme $X(N)_{\zeta_N}$ of $X(N)_{\mathbb{Z}[1/N,\zeta_N]}$ using generalized elliptic curves.

Theorem 2.24. *The curve $Y(N)_{\zeta_N}$ (resp. $X(N)_{\zeta_N}$) is an affine (resp. proper) smooth geometrically connected curve over $\mathbb{Z}\left[\frac{1}{N}, \zeta_N\right]$.*

Proof. See corollary 10.9.2 of Katz and Mazur's book [12]. □

From now on, set $R_0 = \mathbb{Z}\left[\frac{1}{N}, \zeta_N\right]$. Theorem 2.24 implies that $Y(N)_{R_0}$ (resp. $X(N)_{R_0}$) is a disjoint union of $\varphi(N) = |(\mathbb{Z}/N\mathbb{Z})^\times|$ geometrically connected components, one for each primitive N–th root of unity. Thus, over R_0, the scheme of cusps $C(N)_{R_0}$ becomes a disjoint union of points: it is a divisor on a curve given by the equation $j^{-1} = 0$. Write $\widehat{C(N)}$ for the formal completion of $X(N)$ along the cusps (i.e., along the divisor defining the cusps). We have a cartesian diagram

$$\begin{array}{ccccccc}
\widehat{C(N)}_{R_0} & \hookrightarrow & X(N)_{R_0} & \hookleftarrow & Y(N)_{R_0} & \hookleftarrow & \widehat{C(N)}_{R_0,j^{-1}} \\
\downarrow & \ulcorner & \downarrow & & \downarrow & \urcorner & \downarrow \\
\mathrm{Spec}(R_0[\![j^{-1}]\!]) & \hookrightarrow & \mathrm{Proj}(R_0[j]) & \hookleftarrow & \mathrm{Spec}(R_0[j]) & \hookleftarrow & \mathrm{Spec}(R_0((j^{-1})))
\end{array}$$

where $\widehat{C(N)}_{R_0,j^{-1}}$ is the open subscheme of $\widehat{C(N)}_{R_0}$ where j^{-1} is invertible (as the morphism $\mathrm{Spec}(R_0((j^{-1}))) \to \mathrm{Proj}(R_0[j])$ factors through $\mathrm{Spec}(R_0[\![j^{-1}]\!])$).

Lemma 2.25. *The scheme $\widehat{C(N)}_{R_0}$ is the normalization of $R_0[\![j^{-1}]\!]$ in $\widehat{C(N)}_{R_0,j^{-1}}$.*

Proof. See lemma 8.11.2 of Katz and Mazur's book [12]. □

There is a unique R_0–algebra isomorhism $R_0[\![j^{-1}]\!] \to R_0[\![q]\!]$ which is continuous for the adic topologies and which maps j^{-1} to the inverse of the j–invariant of the Tate curve (see the remark after definition 2.5). That is,

$$j^{-1} \mapsto q(1 - 744q + \cdots).$$

This extends to an isomorphism $R_0((j^{-1})) \to R_0((q))$. Via these isomorphisms, we can view $\widehat{C(N)}_{R_0}$ as a scheme over $R_0[\![q]\!]$ and $\widehat{C(N)}_{R_0,j^{-1}}$ as a scheme over $R_0((q))$. Using this, one can prove the following result.

Theorem 2.26. *There is a canonical bijection between the cusps over R_0 (i.e., the points of $C(N)_{R_0}$) and the isomorphism classes of level $\Gamma(N)$–structures on* $\mathrm{Tate}(q)$ *over $\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} R_0$.*

Proof. In fact, there is a much more precise way to describe the cusps in terms of the Tate curve; see proposition 8.11.7 of Katz and Mazur's book [12]. □

This suggests that the q–expansions of modular forms should be somehow related to the cusps.

Theorem 2.27. *There is an invertible sheaf of modules $\underline{\omega}$ on $X(N)$ whose restriction to $Y(N)$ is $\underline{\omega}_{\mathbb{E}/Y(N)}$ and whose sections over the formal completion $R_0[\![q]\!]$ at each cusp (after base change to R_0) correspond to the $R_0[\![q^{1/N}]\!]$–multiples of ω_{can} on* $\mathrm{Tate}(q)$ *(via the correspondence between cusps and the Tate curve implied in theorem 2.26).*

Proof. See section 10.13 (especially proposition 10.13.4) of Katz and Mazur's book [12]. □

Alternatively, $\underline{\omega}$ can be constructed in the same way as $\underline{\omega}_{\mathbb{E}/Y(N)}$ but using generalized elliptic curves. In that case, one can prove that the cusps correspond to $\mathrm{Tate}(q)$ over $\mathbb{Z}[\![q^{1/N}]\!]$ (recall that the Tate curve is not an elliptic curve if q is not invertible, but in this case it is at least a generalized elliptic curve).

Let $f \in \mathrm{H}^0(Y(N), \underline{\omega}^{\otimes k}_{\mathbb{E}/Y(N)} \otimes_{\mathbb{Z}[1/N]} R_0) = F(R_0; \Gamma(N), k)$ for some $k \in \mathbb{Z}$. Via the correspondence between cusps and $\Gamma(N)$–level structures on $\mathrm{Tate}(q)$ over $\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} R_0$, we see that the q–expansions of f are holomorphic (i.e., contain no negative powers of q) if and only if f extends to $\mathrm{H}^0(X(N), \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} R_0)$. This leads us to the following generalization of the definition of modular forms holomorphic at ∞.

Definition 2.28. Let K be a $\mathbb{Z}\left[\frac{1}{N}\right]$–module and let $k \in \mathbb{Z}$. A *modular form* for $\Gamma(N)$ of weight k with coefficients in K and *holomorphic at* ∞ is an element of $\mathrm{H}^0(X(N), \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} K)$ (cf. definition 2.20 and definition 2.7). We also write $M(K;\Gamma(N),k) = \mathrm{H}^0(X(N), \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} K)$ and

$$M(K;\Gamma(N)) = \bigoplus_{k' \in \mathbb{Z}} M(K;\Gamma(N);k')\,.$$

Remark. Let $f \in M(K;\Gamma(N),k)$. At each cusp, f has a *q–expansion* given by $f\big(\mathrm{Tate}(q)\,/\,\big(\mathbb{Z}((q^{1/N})) \otimes_{\mathbb{Z}} \mathbb{Z}\left[\frac{1}{N},\zeta_N\right]\big), \omega_{\mathrm{can}}, \alpha_N\big)$, for some level $\Gamma(N)$–structure α_N associated with the cusp, which lies in $\mathbb{Z}[\![q^{1/N}]\!] \otimes_{\mathbb{Z}} \mathbb{Z}\left[\frac{1}{N},\zeta_N\right] \otimes_{\mathbb{Z}[1/N]} K$ (by theorems 2.26 and 2.27).

2.4 The q–expansion principle

In this section we prove a very important result called the q–expansion principle, which asserts that every modular form is uniquely determined by its q–expansions. The proof of this result is section 1.6 of Katz's article [8].

Theorem 2.29. *Let $N, k \in \mathbb{Z}$ with $N \geq 3$. Let ζ_N denote a primitive N–th root of unity and let K be a $\mathbb{Z}\left[\frac{1}{N}\right]$–module. Consider $f \in M(K;\Gamma(N),k)$. If each connected component of $X(N)_{\mathbb{Z}[1/N,\zeta_N]}$ has at least one cusp at which the corresponding q–expansion of f vanishes identically, then $f = 0$.*

Proof. Observe that $\underline{\omega}$ is a locally free $\mathscr{O}_{X(N)}$–module and that $X(N)$ is flat over $\mathbb{Z}\left[\frac{1}{N}\right]$. Thus, for every short exact sequence

$$0 \to F' \to F \to F'' \to 0$$

of $\mathbb{Z}\left[\frac{1}{N}\right]$–modules, we obtain a short exact sequence

$$0 \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} F' \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} F \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} F'' \to 0$$

of $\mathscr{O}_{X(N)}$–modules and, after taking sheaf cohomology, an exact sequence

$$0 \to M(F';\Gamma(N),k) \to M(F;\Gamma(N);k) \to M(F'',\Gamma(N),k)\,.$$

We can regard K as a $\mathbb{Z}\left[\frac{1}{N}\right]$–submodule of the dual ring $D(K) = \mathbb{Z}\left[\frac{1}{N}\right] \oplus K$, where the multiplication is given by $(a,x)(b,y) = (ab, ay + bx)$, and we get an

inclusion $M(K;\Gamma(N),k) \hookrightarrow M(D(K);\Gamma(N),k)$. Hence, it suffices to prove the theorem when K is a ring over $\mathbb{Z}\left[\frac{1}{N}\right]$. Next, we can express K as the inductive limit of its finitely generated $\mathbb{Z}\left[\frac{1}{N}\right]$–subalgebras, which are noetherian. But the formation of the cohomology of quasi-coherent sheaves of modules on a noetherian scheme commutes with inductive limits, so we can assume that K is a noetherian ring.

If $\mathfrak{p}$ is a prime ideal in K, the canonical morphism $K \to K_{\mathfrak{p}}$ is flat and, by flat base change on cohomology, $M(K_{\mathfrak{p}};\Gamma(N),k) \cong M(K;\Gamma(N),k) \otimes_K K_{\mathfrak{p}}$. In fact, by studying the localizations of K at all its prime ideals, it suffices to prove the theorem when K is a noetherian local ring. In this case, the completion of K with respect to its maximal ideal $\mathfrak{m}$ gives a faithfully flat morphism $K \to \widehat{K}_{\mathfrak{m}}$ and so we can replace K with $\widehat{K}_{\mathfrak{m}}$. That is, we assume that $(K,\mathfrak{m})$ is a complete noetherian local ring. In particular, $K = \varprojlim(K/\mathfrak{m}^n)$, where we take the projective limit of the artinian local rings $K/\mathfrak{m}^n$ for $n \in \mathbb{N}$ with the canonical projections. By the theorem on formal functions, it suffices to prove the theorem when K is an artinian local ring over $\mathbb{Z}\left[\frac{1}{N}\right]$.

We regard f as a global section of the sheaf on $X(N)_K$ obtained from $\underline{\omega}^{\otimes k}$ by base change. The hypothesis of the theorem is that the germs of the section f at one cusp in each geometrically connected component of $X(N)_{K\otimes\mathbb{Z}[1/N,\zeta_N]}$ are 0. This means that we can find open neighbourhoods of these cusps on which f vanishes. But the union of such open neighbourhoods gives an open subset U of $X(N)_K$ which meets all the irreducible components (because a smooth geometrically connected curve is geometrically integral). In particular, U is dense in $X(N)_K$.

Suppose, for the sake of contradiction, that $f \neq 0$. Then, the support of f is a non-empty closed subset Z of $X(N)_K$ with $Z \cap U = \emptyset$. In particular, Z contains no irreducible components (hence no maximal points) of $X(N)_K$. Let z be a maximal point of Z and consider the local ring $A = \mathscr{O}_{X(N)_K,z}$ with its maximal ideal $\mathfrak{m} = \mathfrak{m}_z$. There is a canonical morphism

$$\mathrm{Spec}(A) = \mathrm{Spec}(\mathscr{O}_{X(N)_K,z}) \to X(N)_K$$

which is a homeomorphism of $\mathrm{Spec}(A)$ onto the set of points of $X(N)_K$ which are generalizations of z. We deduce that $\{\,\mathfrak{m}\,\} \subsetneq \mathrm{Spec}(A)$ because some maximal point of $X(N)_K$ must be a generalization of z distinct from z. On the other hand, the sheaf $\underline{\omega}^{\otimes k}$ is an invertible $\mathscr{O}_{X(N)}$–module, so the stalk of its base change to K at z is (non-canonically) isomorphic to A. Fix one such isomorphism and

let a denote the image of f_z in A. Since z is a maximal point of the support of f, the support of a must be $\{\mathfrak{m}\}$. This means that $a \neq 0$ in A but, for every $\mathfrak{p} \in \operatorname{Spec}(A) \setminus \{\mathfrak{m}\}$, the image of a in the localization $A_\mathfrak{p}$ is 0. Now take $b \in \mathfrak{m}$. Observe that $\operatorname{Spec}(A_b)$ is a proper open subset of $\operatorname{Spec}(A)$ and so the image of a in A_b must be 0. That is to say, $b^n a = 0$ for some $n \in \mathbb{N}$ and, in particular, b is a zero divisor. But this holds for all the elements of $\mathfrak{m}$, which implies that A has depth 0. Moreover, $X(N)_K$ is smooth over an artinian local ring K and so it is Cohen–Macaulay. Therefore, A has dimension 0 (equal to the depth), thus contradicting the fact that $\{\mathfrak{m}\} \subsetneq \operatorname{Spec}(A)$. □

Corollary 2.30 (the q–expansion principle). *Let $N, k \in \mathbb{Z}$ with $N \geq 3$. Let ζ_N denote a primitive N–th root of unity and let K be a $\mathbb{Z}\left[\frac{1}{N}\right]$–module with a submodule L. Consider $f \in M(K; \Gamma(N), k)$. If each connected component of $X(N)_{\mathbb{Z}[1/N, \zeta_N]}$ has at least one cusp at which all the coefficients of the corresponding q–expansion of f lie in $L \otimes_{\mathbb{Z}[1/N]} \mathbb{Z}\left[\frac{1}{N}, \zeta_N\right]$, then $f \in M(L; \Gamma(N), k)$.*

Proof. Since $\underline{\omega}$ is a locally free $\mathscr{O}_{X(N)}$–module and $X(N)$ is flat over $\mathbb{Z}\left[\frac{1}{N}\right]$, we obtain a short exact sequence

$$0 \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} L \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} K \to \underline{\omega}^{\otimes k} \otimes_{\mathbb{Z}[1/N]} K/L \to 0$$

of $\mathscr{O}_{X(N)}$–modules. Taking sheaf cohomology yields an exact sequence

$$0 \to M(L; \Gamma(N), k) \to M(K; \Gamma(N), k) \to M(K/L; \Gamma(N), k)\,.$$

The image of f in $M(K/L; \Gamma(N), k)$ satisfies the hypothesis of theorem 2.29 and so is 0. Therefore, f lies in the image of $M(L; \Gamma(N), k) \hookrightarrow M(K; \Gamma(N), k)$. □

Corollary 2.31. *Theorem 2.29 and corollary 2.30 hold also for modular forms which are not holomorphic at ∞ (that is, replacing $M(K; \Gamma(N), k)$ with $F(K; \Gamma(N), k)$ and $M(L; \Gamma(N), k)$ with $F(L; \Gamma(N), k)$).*

Proof. Consider the modular discriminant $\Delta \in M(K; \Gamma(N), 12)$. Observe that Δ is defined first as a cusp form over $\mathbb{C}$, but its q–expansion has integer coefficients (see example 1.5) and so we can regard it as a cusp form over $\mathbb{Z}\left[\frac{1}{N}\right]$ by corollary 2.30 and then consider its image in $M(K; \Gamma(N), 12)$ by base change. Moreover, Δ is invertible in $F(K; \Gamma(N))$ because the discriminant of an elliptic curve must be invertible.

Let $f \in F(K;\Gamma(N),k)$ and choose $r \gg 0$ such that $f\Delta^r$ is holomorphic at ∞. The q–expansion of f at a cusp is 0 if and only if the q–expansion of $f\Delta^r$ at the same cusp is also 0. Thus, if each connected component of $X(N)_{\mathbb{Z}[1/N,\zeta_N]}$ has one cusp at which the corresponding q–expansion of f is 0, we can apply theorem 2.29 to deduce that $f\Delta^r = 0$ and so that $f = 0$.

Now the proof of corollary 2.30 works exactly in the same way if we replace $\underline{\omega}$ with $\underline{\omega}_{\mathbb{E}/Y(N)}$. □

Chapter 3

Some geometric tools

This chapter presents some geometric tools which we use in the next chapter. To begin with, we describe the absolute and relative Frobenius morphisms for schemes in positive characteristic and, focusing on elliptic curves, the duality between the Frobenius and the Verschiebung morphisms. Next, we recall the definition of (algebraic) de Rham cohomology and a couple of important filtrations on it, again focusing on the case of elliptic curves. Then, we recall the constructions of the Gauss–Manin connection and of the Kodaira–Spencer isomorphism. Finally, we exhibit some computations on the Tate curve using complex analytic tools.

The concepts introduced in this chapter can be defined in quite general situations, but we only use them for elliptic curves and so we simplify some things. Furthermore, we use some deep theorems which we only state without proof (but giving appropriate references to the proofs).

The first three sections of this chapter explain most of the concepts summarized in Kedlaya's notes [14] but with more detailed explanations in some parts. More precise references are given later. The computations at the end are essentially an extended version of sections A1.3 and A1.4 of Katz's article [8].

3.1 The Frobenius morphisms

Throughout this section, let p be a prime number and let S be an $\mathbb{F}_p$–scheme. That is, for every open subset U of S and every section $a \in \Gamma(U, \mathscr{O}_S)$, $pa = 0$. In what follows, we define the Frobenius morphisms following section 12.1 of Katz and Mazur's book [12].

Definition 3.1. The *absolute Frobenius endomorphism* of S is the morphism of schemes (over $\mathbb{F}_p$) $\mathrm{Frob}_S \colon S \to S$ which is the identity on the underlying topological spaces and such that $\mathrm{Frob}_S^\sharp \colon \mathscr{O}_S \to \mathscr{O}_S$ is given, for every open subset U of S, by the map $a \mapsto a^p$ on $\Gamma(U, \mathscr{O}_S)$.

Remark. The map

$$\begin{aligned} \mathrm{Frob}_S^\sharp \colon \Gamma(U, \mathscr{O}_S) &\to \Gamma(U, \mathscr{O}_S) \\ a &\mapsto a^p \end{aligned}$$

is a morphism of rings because $p\Gamma(U, \mathscr{O}_S) = 0$.

If S is affine, say $S = \mathrm{Spec}(R)$ for some $\mathbb{F}_p$–algebra R, then Frob_S corresponds to the morphism of rings

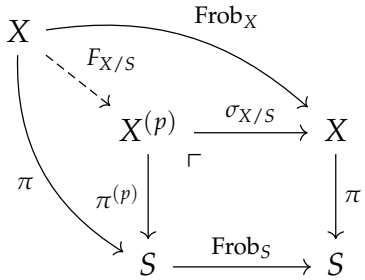
$$\begin{aligned} R &\to R \\ a &\mapsto a^p \end{aligned}$$

(which is also called the Frobenius endomorphism of R).

Now consider a morphism of $\mathbb{F}_p$–schemes $\pi \colon X \to S$. It is clear that the diagram

$$\begin{array}{ccc} X & \xrightarrow{\mathrm{Frob}_X} & X \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ S & \xrightarrow{\mathrm{Frob}_S} & S \end{array}$$

is commutative. Observe, however, that Frob_X is not an S–morphism in general. Thus, if we want to regard X as an S–scheme, we need to give a different definition of the Frobenius morphism. Let $\pi^{(p)} \colon X^{(p)} \to S$ be the pull-back of $\pi \colon X \to S$ under $\mathrm{Frob}_S \colon S \to S$. There is a unique S–morphism $F_{X/S} \colon X \to X^{(p)}$ which makes the diagram



commutative.

Definition 3.2. The S–morphism $F_{X/S} \colon X \to X^{(p)}$ constructed above is called the *relative Frobenius morphism* of X over S. (When there is no possible confusion, we write $F = F_{X/S}$.)

The construction of the Frobenius morphisms can be iterated. That is to say, for each $n \in \mathbb{N}$, we can consider $\mathrm{Frob}_S^n = \mathrm{Frob}_S \circ \overset{(n)}{\ldots} \circ \mathrm{Frob}_S \colon S \to S$ and also $F^n = F_{X^{(p^{n-1})}/S} \circ \cdots \circ F_{X^{(p)}/S} \circ F_{X/S} \colon X \to X^{(p^n)}$ (where $X^{(p^n)}$ is obtained from X by pull-back under Frob_S^n). These iterated Frobenius morphisms have mostly the same properties as Frob_S and $F_{X/S}$ and are useful if one works over $\mathbb{F}_{p^n}$ for $n \in \mathbb{N}$.

Let us describe the situation locally. Suppose that $S = \mathrm{Spec}(R)$ and that $X = \mathrm{Spec}(B)$. The morphism $\pi \colon X \to S$ endows B with the structure of an R–algebra, by means of which we can identify B with $R[X_i : i \in I]/(f_j : j \in J)$, where X_i for $i \in I$ is a family of indeterminates and f_j for $j \in J$ is a family of polynomials in these indeterminates. For every polynomial $f \in R[X_i : i \in I]$, which is of the form

$$f = \sum_{\nu \in \mathbb{N}^{(I)}} a_\nu X^\nu \,,$$

we write

$$f^{(p)} = \sum_{\nu \in \mathbb{N}^{(I)}} a_\nu^p X^\nu$$

(i.e., $f^{(p)}$ is the polynomial obtained from f by raising its coefficients, but not the variables, to the p–th power). By construction, $X^{(p)} = \mathrm{Spec}(B^{(p)})$, where $B^{(p)} = B \otimes_R R$. Here, we regard R as an R–algebra by means of $\mathrm{Frob}_S^\sharp$ (not the identity). That is, $a_\nu X^\nu \otimes 1 = X^\nu \otimes a_\nu^p$ in $B^{(p)}$. Therefore, we can identify $B^{(p)}$ with $R[X_i : i \in I]/(f_j^{(p)} : j \in J)$ and $\sigma_{X/S}^\sharp \colon B \to B^{(p)}$ is induced by $f \mapsto f^{(p)}$. Since $\mathrm{Frob}_X^\sharp \colon B \to B$ is given by $f \mapsto f^p$, we conclude that $F^\sharp \colon B^{(p)} \to B$ is induced by the morphism of R–algebras $R[X_i : i \in I] \to R[X_i : i \in I]$ defined by $X_i \mapsto X_i^p$. Using the local description of the Frobenius morphisms, one can easily see that $F_{X/S} \circ \sigma_{X/S} = \mathrm{Frob}_{X^{(p)}}$. On the other hand, by definition, we have that $\sigma_{X/S} \circ F_{X/S} = \mathrm{Frob}_X$.

We give an alternative interpretation of the ring $B^{(p)}$ following section V.2 of the preliminary version of van der Geer and Moonen's book [7]. This alternative interpretation is useful later when we define a *dual* of F.

Define $T^p(B) = B \otimes_R \overset{(p)}{\ldots} \otimes_R B$ and consider the subalgebra $S^p(B)$ of $T^p(B)$ consisting of the symmetric tensors (i.e., those which are invariant under the action of the symmetric group $\mathfrak{S}_p$ by permutations). Let $S \colon T^p(B) \to S^p(B)$ be the morphism of R–modules given by

$$S(b_1 \otimes \cdots \otimes b_p) = \sum_{\sigma \in \mathfrak{S}_p} b_{\sigma(1)} \otimes \cdots \otimes b_{\sigma(p)} \,.$$

Observe that, for every $s \in S^p(B)$ and every $t \in T^p(B)$, $S(st) = sS(t)$. Therefore, $J = S(T^p(B))$ is an ideal of $S^p(B)$. We obtain a well-defined map

$$\begin{aligned} \varphi_{B/R} \colon B^{(p)} &\to S^p(B)/J \\ b \otimes a &\mapsto a(b \otimes \overset{(p)}{\ldots} \otimes b) \mod J \end{aligned}$$

which, in fact, is a morphism of R–algebras (recall that, in $B^{(p)} = B \otimes_R R$, we view R as an R–module through the Frobenius endomorphism).

Lemma 3.3. *If B is flat over R, then $\varphi_{B/R} \colon B^{(p)} \to S^p(B)/J$ is an isomorphism of rings.*

Proof. Since $\varphi_{B/R}$ is a ring homomorphism, it suffices to prove that it is bijective. In fact, we prove that it is an isomorphism of R–modules.

First, suppose that B is a free R–module with a basis $(e_i)_{i \in I}$. The tensors $e_{i_1} \otimes \ldots \otimes e_{i_p}$ for $(i_1, \ldots, i_p) \in I^p$ form a basis of $T^p(B)$. For each $(i_1, \ldots, i_p) \in I^p$, take its stabilizer $H \subseteq \mathfrak{S}_p$ and define

$$s_{i_1,\ldots,i_p} = \sum_{\overline{\sigma} \in H \backslash \mathfrak{S}_p} e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(p)}} .$$

The symmetric tensors obtained in this way span $S^p(B)$. Now observe that $S(e_i \otimes \cdots \otimes e_i) = p!\, s_{i,\ldots,i} = 0$ for every $i \in I$ and $S(e_{i_1} \otimes \cdots \otimes e_{i_p}) = u s_{i_1,\ldots,i_p}$ for some $u \in R^\times$ if $(i_1, \ldots, i_p) \neq (i_1, \ldots, i_1)$. Therefore, the tensors $e_i \otimes \cdots \otimes e_i$ for $i \in I$ form a basis of $S^p(B)/J$ and it is clear from this that $\varphi_{B/R}$ is an isomorphism in this case.

In the general case, B is a filtered direct limit of free modules (of finite rank) by Lazard's theorem. But the functor $\varinjlim$ is exact and commutes with tensor products, so the result follows from the previous case. □

We now focus on the case of an elliptic curve E over S. Let us recall some basic definitions and results on elliptic curves which can be found in Katz and Mazur's book [12].

Definition 3.4. Let E and E' be two elliptic curves over a scheme S. A homomorphism $f \colon E \to E'$ of group schemes over S is called an *isogeny* if it is surjective, finite and locally free. In this case, $\mathrm{Ker}(f)$ is a finite locally free group scheme over S of locally constant rank and, if this rank is constant equal to d, we say that f has *degree* d.

Definition 3.5. Let $f\colon E \to E'$ be an isogeny of degree d between two elliptic curves E and E' over a scheme S. We say that an isogeny $f'\colon E' \to E$ of degree d is *dual* to f if $f' \circ f = [d]$ (here, $[d]\colon E \to E$ is the homomorphism given by multiplication by d).

Remark. If $f' \circ f = [d]$, then $f \circ f' \circ f = f \circ [d] = [d] \circ f$ and so $f \circ f' = [d]$. Indeed, since f is an isogeny, it is faithfully flat and, in particular, an epimorphism of schemes.

Theorem 3.6. *Let $f\colon E \to E'$ be an isogeny of degree d between two elliptic curves E and E' over a scheme S. There exists an isogeny $f'\colon E' \to E$ which is dual to f.*

Proof. See theorem 2.6.1 of Katz and Mazur's book [12]. □

Lemma 3.7. *Let S be an $\mathbb{F}_p$–scheme and let E be an elliptic curve over S. The relative Frobenius morphism $F\colon E \to E^{(p)}$ is an isogeny of degree p.*

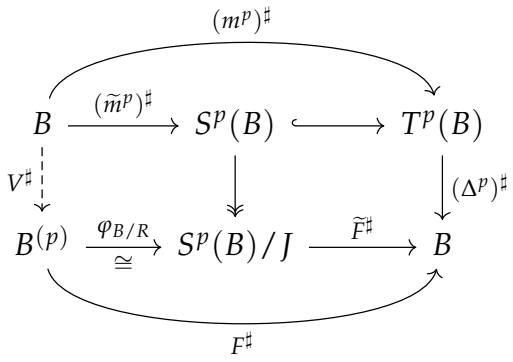
Proof. Using that elliptic curves have dimension 1 over the base, it is clear from its local expression that F is surjective, finite and locally free of rank p. □

Definition 3.8. Let E be an elliptic curve over an $\mathbb{F}_p$–scheme S. Consider the relative Frobenius morphism $F_{E/S}\colon E \to E^{(p)}$. Its dual isogeny $V_{E/S}\colon E^{(p)} \to E$ is called the *Verschiebung morphism* of E over S. (When there is no possible confusion, we write $V = V_{E/S}$.)

In fact, the construction of the Verschiebung morphism works more generally. Next, we give an alternative more direct construction following section V.2 of the preliminary version of van der Geer and Moonen's book [7].

From now on, suppose that X is a commutative group scheme which is flat over S (e.g., if X/S is an elliptic curve). Write $m\colon X \times_S X \to X$ for the morphism corresponding to the group law and $\Delta\colon X \to X \times_S X$ for the diagonal morphism. Write also $m^p\colon X \times_S \overset{(p)}{\cdots} \times_S X \to X$ for the p–fold group law and $\Delta^p\colon X \to X \times_S \overset{(p)}{\cdots} \times_S X$ for the p–fold diagonal. We work locally: consider affine open subsets $U = \operatorname{Spec}(B)$ of X and $W = \operatorname{Spec}(R)$ of S such that $\pi(U) \subseteq W$. We have seen that $U^{(p)} = \operatorname{Spec}(B^{(p)})$ can be identified with $\operatorname{Spec}(S^p(B)/J)$ through $\varphi_{B/R}$. Then, the ring homomorphism $F^\sharp\colon B^{(p)} \to B$ factors through $\varphi_{B/R}$; that is to say, we can express $F^\sharp = \widetilde{F}^\sharp \circ \varphi_{B/R}$ with $\widetilde{F}^\sharp\colon S^p(B)/J \to B$ defined by $\widetilde{F}^\sharp(b_1 \otimes \cdots \otimes b_p \bmod J) = b_1 \dots b_p$. Since the group scheme is commutative, the morphism $(m^p)^\sharp\colon B \to T^p(B)$ factors through $S^p(B)$; let $(\widetilde{m}^p)^\sharp\colon B \to S^p(B)$

be the induced morphism. Now, we define $V^\sharp\colon B \to B^{(p)}$ to be the unique morphism which makes the diagram



commutative. This construction can be performed on affine open coverings of X and S and, after gluing together the resulting morphisms, we obtain a morphism $V = V_{X/S}\colon X^{(p)} \to X$ of S–schemes. The next result shows that this morphism must coincide with the Verschiebung morphism as defined in definition 3.8 when X is an elliptic curve (that is why we have used the same notation).

Corollary 3.9. *The composition $V \circ F$ is the homomorphism $[p]\colon X \to X$ given by multiplication by p.*

Proof. We can check it locally using the same notation as in the construction above. But we have seen that $F^\sharp \circ V^\sharp = (\Delta^p)^\sharp \circ (m^p)^\sharp$ (see the commutative diagram defining $V^\sharp$). The result follows from this because $[p] = m^p \circ \Delta^p$. □

3.2 De Rham cohomology

We begin this section by briefly recalling the definition of algebraic de Rham cohomology. We mostly follow section 1 of Kedlaya's notes [14] and the summary in sections 1.1 and 1.2 of Wedhorn's notes [22].

Let $\mathcal{A}$ and $\mathcal{B}$ be two abelian categories and suppose that $\mathcal{A}$ has enough injectives (e.g., the category of abelian sheaves on a scheme). Let $T\colon \mathcal{A} \to \mathcal{B}$ be a left-exact functor. We say that $M \in \mathrm{Ob}(\mathcal{A})$ is *T–acyclic* if $\mathrm{R}^iT(M) = 0$ for every $i > 0$, where R^iT is the i–th right derived functor of T.

Now consider a complex $C^\bullet$ of objects in $\mathcal{A}$ such that $C^k = 0$ for every $k < 0$. Choose a T–acyclic resolution $C^\bullet \to I^\bullet$ (i.e., a quasi-isomorphism to $I^\bullet$ with I^k T–acyclic for each $k \geq 0$ and $I^k = 0$ for each $k < 0$). Then, the *right hyper-derived functors* $\mathbb{R}^iT$ are given by $\mathbb{R}^iT(C^\bullet) = \mathrm{H}^i(T(I^\bullet))$. (The result is independent of the choice of $I^\bullet$.)

Definition 3.10. Let $\pi\colon X \to S$ be a morphism of schemes and consider the relative de Rham complex

$$\Omega^{\bullet}_{X/S} = (0 \to \mathscr{O}_X \to \Omega^1_{X/S} \to \Omega^2_{X/S} \to \cdots).$$

For each $n \geq 0$, the *n–th (relative) de Rham cohomology* of X over S is defined to be $\mathrm{H}^n_{\mathrm{dR}}(X/S) = \mathbb{R}^n\pi_*(\Omega^{\bullet}_{X/S})$.

Now fix a morphism of schemes $\pi\colon X \to S$. In order to compute the de Rham cohomology of X over S, we can choose a Cartan–Eilenberg resolution $\mathscr{D}^{\bullet,\bullet}$ of $\Omega^{\bullet}_{X/S}$. Thus, we get a double complex

$$\begin{array}{ccccccc}
 & & \vdots & & \vdots & & \vdots \\
 & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & \mathscr{D}^{0,1} & \longrightarrow & \mathscr{D}^{1,1} & \longrightarrow & \mathscr{D}^{2,1} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & \mathscr{D}^{0,0} & \longrightarrow & \mathscr{D}^{1,0} & \longrightarrow & \mathscr{D}^{2,0} & \longrightarrow & \cdots \\
 & & \uparrow & & \uparrow & & \uparrow \\
0 & \longrightarrow & \mathscr{O}_X & \longrightarrow & \Omega^1_{X/S} & \longrightarrow & \Omega^2_{X/S} & \longrightarrow & \cdots
\end{array}$$

in which the columns are injective resolutions. We can then form the associated total complex $\mathscr{C}^{\bullet}$, with

$$\mathscr{C}^k = \bigoplus_{a+b=k} \mathscr{D}^{a,b},$$

which is an injective resolution of $\Omega^{\bullet}_{X/S}$. In particular, for each $n \in \mathbb{Z}$,

$$\mathrm{H}^n_{\mathrm{dR}}(X/S) = \mathrm{R}^n\pi_*(\mathscr{C}^{\bullet}) = \mathrm{H}^n(\pi_*(\mathscr{C}^{\bullet})) = \mathscr{H}^n(\pi_*(\mathscr{C}^{\bullet})).$$

We have (at least) two natural filtrations ${}_{\mathrm{I}}\mathrm{Fil}$ and ${}_{\mathrm{II}}\mathrm{Fil}$ on $\mathscr{C}^{\bullet}$, namely

$${}_{\mathrm{I}}\mathrm{Fil}^i\,\mathscr{C}^k = \bigoplus_{a \geq i} \mathscr{D}^{a,k-a} \quad \text{and} \quad {}_{\mathrm{II}}\mathrm{Fil}^j\,\mathscr{C}^k = \bigoplus_{b \geq j} \mathscr{D}^{k-b,b},$$

which induce filtrations

$${}_{\mathrm{I}}\mathrm{Fil}^i\,\mathrm{H}^n_{\mathrm{dR}}(X/S) = \mathrm{Im}\big(\mathrm{R}^n\pi_*({}_{\mathrm{I}}\mathrm{Fil}^i\,\mathscr{C}^{\bullet}) \to \mathrm{R}^n\pi_*(\mathscr{C}^{\bullet}) = \mathrm{H}^n_{\mathrm{dR}}(X/S)\big)$$

and

$${}_{\mathrm{II}}\mathrm{Fil}^j\,\mathrm{H}^n_{\mathrm{dR}}(X/S) = \mathrm{Im}\big(\mathrm{R}^n\pi_*({}_{\mathrm{II}}\mathrm{Fil}^j\,\mathscr{C}^{\bullet}) \to \mathrm{R}^n\pi_*(\mathscr{C}^{\bullet}) = \mathrm{H}^n_{\mathrm{dR}}(X/S)\big)$$

on the de Rham cohomology of X over S. These filtrations give rise to two spectral sequences ${}_{\mathrm{I}}E$ and ${}_{\mathrm{II}}E$ abutting to the de Rham cohomology. They are sometimes called the *first and second spectral sequences of hypercohomology*. Some of their terms are:

$$\begin{aligned}
{}_{\mathrm{I}}E_0^{a,b} &= {}_{\mathrm{I}}\mathrm{Gr}^a\, \pi_*(\mathscr{C}^{a+b}) = \pi_*(\mathscr{D}^{a,b}), & {}_{\mathrm{II}}E_0^{a,b} &= {}_{\mathrm{II}}\mathrm{Gr}^a\, \pi_*(\mathscr{C}^{a+b}) = \pi_*(\mathscr{D}^{b,a}),\\
{}_{\mathrm{I}}E_1^{a,b} &= \mathrm{R}^b\pi_*(\Omega^a_{X/S}), & {}_{\mathrm{II}}E_1^{a,b} &= \pi_*(\mathscr{H}^b(\mathscr{D}^{\bullet,a})),\\
{}_{\mathrm{I}}E_2^{a,b} &= \mathscr{H}^a(\mathrm{R}^b\pi_*(\Omega^\bullet_{X/S})), & {}_{\mathrm{II}}E_2^{a,b} &= \mathrm{R}^a\pi_*(\mathscr{H}^b(\Omega^\bullet_{X/S})),\\
{}_{\mathrm{I}}E_\infty^{a,b} &= {}_{\mathrm{I}}\mathrm{Gr}^a\, \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S), & {}_{\mathrm{II}}E_\infty^{a,b} &= {}_{\mathrm{II}}\mathrm{Gr}^a\, \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S)\,.
\end{aligned}$$

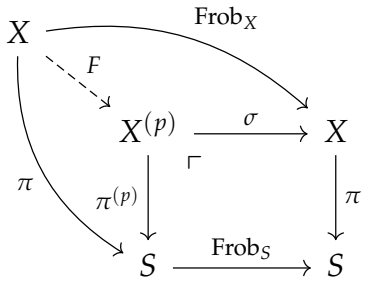
(Here, $\mathscr{H}^k$ denotes the k–th cohomology sheaf of a complex.) We usually write

$$\begin{aligned}
{}_{\mathrm{I}}E_1^{a,b} = \mathrm{R}^b\pi_*(\Omega^a_{X/S}) &\implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S)\,,\\
{}_{\mathrm{II}}E_2^{a,b} = \mathrm{R}^a\pi_*(\mathscr{H}^b(\Omega^\bullet_{X/S})) &\implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S)\,.
\end{aligned}$$

Definition 3.11. We call ${}_{\mathrm{I}}\mathrm{Fil}$ the *Hodge filtration* and ${}_{\mathrm{I}}E$ the *Hodge–de Rham spectral sequence* for X/S. We call ${}_{\mathrm{II}}\mathrm{Fil}$ the *conjugate filtration* and ${}_{\mathrm{II}}E$ the *conjugate spectral sequence* for X/S.

Next, we focus in the case we are most interested in. The main results are stated without proof. This second part is based on sections 2.1 to 2.3 of Katz's article [10], also summarized in sections 1.4 to 1.6 of Wedhorn's notes [22].

Throughout the rest of this section, let p be a fixed prime number. Let S be an $\mathbb{F}_p$–scheme and let $\pi\colon X \to S$ be an S–scheme. Consider the absolute Frobenius endomorphism $\mathrm{Frob}_S\colon S \to S$ and let $X^{(p)}$ be the pull-back of X under Frob_S. Let $F\colon X \to X^{(p)}$ be the relative Frobenius morphism of X/S. We have a commutative diagram



(see section 3.1). We want to use the additional structure in order to interpret the terms of ${}_{\mathrm{II}}E_2$ in an alternative way.

Theorem 3.12. *There is a unique isomorphism of $\mathscr{O}_{X^{(p)}}$–modules*

$$C_k^{-1}\colon \Omega^k_{X^{(p)}/S} \to \mathscr{H}^k(F_*(\Omega^\bullet_{X/S}))$$

for every $k \geq 0$ satisfying the following conditions on sections over open subsets U of $X^{(p)}$:

(i) $C_0^{-1}(1) = 1$;

(ii) $C_{k+k'}^{-1}(\omega \wedge \omega') = C_k^{-1}(\omega) \wedge C_{k'}^{-1}(\omega')$ *for every* $\omega \in \Gamma(U, \Omega^k_{X^{(p)}/S})$ *and every* $\omega' \in \Gamma(U, \Omega^{k'}_{X^{(p)}/S})$, *and*

(iii) $C_1^{-1}(d\sigma^*(x)) = [x^{p-1}\,dx]$ *for all* $x \in \Gamma(F^{-1}(U), \Omega^1_{X/S})$ *(where $[\,\cdot\,]$ denotes the cohomology class).*

Proof. See theorem 7.2 of Katz's article [11]. □

Definition 3.13. For every $k \geq 0$, the k–th *Cartier isomorphism* for X/S is the isomorphism of $\mathscr{O}_{X^{(p)}}$–modules $C_k\colon \mathscr{H}^k(F_*(\Omega^\bullet_{X/S})) \to \Omega^k_{X^{(p)}/S}$ given by theorem 3.12.

Now suppose that X is smooth and proper over S. As F is a homeomorphism and $\pi = \pi^{(p)} \circ F$,

$$\mathrm{R}^a\pi_*(\mathscr{H}^b(\Omega^\bullet_{X/S})) \cong \mathrm{R}^a\pi^{(p)}_*(F_*(\mathscr{H}^b(\Omega^\bullet_{X/S}))) \cong \mathrm{R}^a\pi^{(p)}_*(\mathscr{H}^b(F_*(\Omega^\bullet_{X/S})))\,.$$

In addition, the Cartier isomorphism induces an isomorphism

$$\mathrm{R}^a\pi^{(p)}_*(\mathscr{H}^b(F_*(\Omega^\bullet_{X/S}))) \cong \mathrm{R}^a\pi^{(p)}_*(\Omega^b_{X^{(p)}/S})$$

and so we usually rewrite the conjugate spectral sequence as

$${}_{\mathrm{II}}E_2^{a,b} = \mathrm{R}^a\pi^{(p)}_*(\Omega^b_{X^{(p)}/S}) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S)\,.$$

Also, the canonical isomorphism $\sigma^*(\Omega^b_{X/S}) \cong \Omega^b_{X^{(p)}/S}$ induces an isomorphism

$$\mathrm{R}^a\pi^{(p)}_*(\Omega^b_{X^{(p)}/S}) \cong \mathrm{R}^a\pi^{(p)}_*(\sigma^*(\Omega^b_{X/S}))$$

and, if $\mathrm{R}^a\pi_*(\Omega^b_{X/S})$ is flat over $\mathscr{O}_S$ (e.g., if it is a locally free $\mathscr{O}_S$–module), the flat base change theorem yields an isomorphism

$$\mathrm{R}^a\pi^{(p)}_*(\sigma^*(\Omega^b_{X/S})) \cong \mathrm{Frob}_S^*(\mathrm{R}^a\pi_*(\Omega^b_{X/S}))\,.$$

Theorem 3.14. *Suppose that X is smooth and proper over S, as above. If the sheaves $\mathrm{R}^a\pi_*(\Omega^b_{X/S})$ are locally free $\mathscr{O}_S$–modules of finite rank for all $a, b \in \mathbb{Z}$ and the Hodge–de Rham spectral sequence*

$$ {}_{\mathrm{I}}E_1^{a,b} = \mathrm{R}^b\pi_*(\Omega^a_{X/S}) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S) $$

degenerates at ${}_{\mathrm{I}}E_1$, then the conjugate sequence

$$ {}_{\mathrm{II}}E_2^{a,b} = \mathrm{Frob}_S^*(\mathrm{R}^a\pi_*(\Omega^b_{X/S})) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S) $$

degenerates at ${}_{\mathrm{II}}E_2$.

Proof. See proposition 2.3.2 of Katz's article [10]. □

Theorem 3.15. *If X is an elliptic curve over S, then the Hodge–de Rham spectral sequence*

$$ {}_{\mathrm{I}}E_1^{a,b} = \mathrm{R}^b\pi_*(\Omega^a_{X/S}) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S) $$

degenerates at ${}_{\mathrm{I}}E_1$.

Proof. See theorem 4.1.3 and corollary 4.1.5 of Deligne and Illusie's article [4].□

Finally, consider an elliptic curve $\pi\colon E \to S$ (where S is still an $\mathbb{F}_p$–scheme). In this case, the Hodge filtration gives a short exact sequence

$$ 0 \to {}_{\mathrm{I}}\mathrm{Fil}^1\,\mathrm{H}^1_{\mathrm{dR}}(E/S) \to {}_{\mathrm{I}}\mathrm{Fil}^0\,\mathrm{H}^1_{\mathrm{dR}}(E/S) \to {}_{\mathrm{I}}\mathrm{Gr}^0\,\mathrm{H}^1_{\mathrm{dR}}(E/S) \to 0\,. $$

But, since E/S is an elliptic curve, ${}_{\mathrm{I}}\mathrm{Fil}^2\,\mathrm{H}^1_{\mathrm{dR}}(E/S) = 0$. Using theorem 3.15 to express ${}_{\mathrm{I}}\mathrm{Gr}^0\,\mathrm{H}^1_{\mathrm{dR}}(E/S)$ and ${}_{\mathrm{I}}\mathrm{Gr}^1\,\mathrm{H}^1_{\mathrm{dR}}(E/S)$ in terms of ${}_{\mathrm{I}}E_1$, the previous short exact sequence becomes

$$ 0 \to \mathrm{R}^0\pi_*(\Omega^1_{E/S}) \to \mathrm{H}^1_{\mathrm{dR}}(E/S) \to \mathrm{R}^1\pi_*(\mathscr{O}_E) \to 0 $$

and we call it again the *Hodge filtration* for E/S. Similarly, the conjugate filtration together with theorem 3.14 yields a short exact sequence

$$ 0 \to \mathrm{R}^1\pi^{(p)}_*(\mathscr{O}_{E^{(p)}}) \to \mathrm{H}^1_{\mathrm{dR}}(E^{(p)}/S) \to \mathrm{R}^0\pi^{(p)}_*(\Omega^1_{E^{(p)}/S}) \to 0 $$

which we call again the *conjugate filtration* for E/S. These short exact sequences are functorial for S–morphisms $E \to E'$ and their formation commutes with base change by $\mathbb{F}_p$–morphisms $T \to S$.

We have presented these results in characteristic p because it is what we need later in chapter 4. In characteristic 0 there is no Cartier isomorphism and so we cannot express the conjugate spectral sequence as above. However, the Hodge–de Rham sequence always degenerates at ${}_{\mathrm{I}}E$ in characteristic 0. We put it more precisely in the following result.

Theorem 3.16. *Let S be a $\mathbb{Q}$–scheme and let $\pi\colon X \to S$ be a proper smooth morphism of schemes. The sheaves $\mathrm{R}^b\pi_*(\Omega^a_{X/S})$ are locally free $\mathscr{O}_S$–modules of finite rank for all $a, b \in \mathbb{Z}$ and the Hodge–de Rham spectral sequence*

$$ {}_{\mathrm{I}}E_1^{a,b} = \mathrm{R}^b\pi_*(\Omega^a_{X/S}) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/S) $$

degenerates at ${}_{\mathrm{I}}E_1$.

Proof. See theorem 5.5 of Deligne's article [3]. □

3.3 The Gauss–Manin connection

This section explains the construction of the Gauss–Manin connection on de Rham cohomology following sections 1 to 3 of Katz and Oda's article [13]. Throughout this section, let $S \to T$ be a smooth morphism of schemes.

Definition 3.17. Let $\mathscr{E}$ be a quasi-coherent $\mathscr{O}_S$–module. A *connection* on $\mathscr{E}$ is a homomorphism $\nabla\colon \mathscr{E} \to \mathscr{E} \otimes_{\mathscr{O}_S} \Omega^1_{S/T}$ of abelian sheaves on S such that, for every open subset U of S and every $f \in \Gamma(U, \mathscr{O}_S)$ and $e \in \Gamma(U, \mathscr{E})$,

$$ \nabla(fe) = f\nabla(e) + e \otimes d_{S/T}(f) $$

(here, $d_{S/T}\colon \mathscr{O}_S \to \Omega^1_{S/T}$ is the universal derivation of S/T). For each derivation $D \in \mathscr{D}er_{\mathscr{O}_T}(\mathscr{O}_S, \mathscr{O}_S)$, we write $\nabla(D)\colon \mathscr{E} \to \mathscr{E}$ for the composition

$$ \mathscr{E} \xrightarrow{\nabla} \mathscr{E} \otimes_{\mathscr{O}_S} \Omega^1_{S/T} \xrightarrow{\mathrm{id}_{\mathscr{E}} \otimes \widetilde{D}} \mathscr{E} \otimes_{\mathscr{O}_S} \mathscr{O}_S \cong \mathscr{E}\,, $$

where $\widetilde{D}$ is the unique element of $\mathscr{H}om_{\mathscr{O}_S}(\Omega^1_{S/T}, \mathscr{O}_S)$ such that $D = \widetilde{D} \circ d_{S/T}$.

Remarks.

(1) On sections $f \in \Gamma(U, \mathscr{O}_S)$ and $e \in \Gamma(U, \mathscr{E})$ for some open subset U of S, we have that

$$ \nabla(D)(fe) = D(f)e + \nabla(D)(e) \otimes d_{S/T}(f) $$

by definition.

(2) For each $k \in \mathbb{N}$, ∇ induces a connection $\nabla\colon \mathscr{E}^{\otimes k} \to \mathscr{E}^{\otimes k} \otimes_{\mathscr{O}_S} \Omega^1_{S/T}$ by the Leibniz rule on sections. That is, for an open subset U of S and sections $e_1, \dots, e_k \in \Gamma(U, \mathscr{E})$,

$$\nabla(e_1 \otimes \cdots \otimes e_k) = \nabla(e_1) \otimes e_2 \otimes \cdots \otimes e_k + \cdots + e_1 \otimes e_2 \otimes \cdots \otimes \nabla(e_k)\,.$$

This in turn induces a connection on $\operatorname{Sym}^k \mathscr{E}$.

Now consider a smooth T–morphism $\pi\colon X \to S$. By smoothness, we get a short exact sequence of $\mathscr{O}_X$–modules

$$0 \to \pi^*(\Omega^1_{S/T}) \to \Omega^1_{X/T} \to \Omega^1_{X/S} \to 0\,.$$

We can construct the associated *Koszul filtration* Fil on the de Rham complex $\Omega^\bullet_{X/T}$, given by $\operatorname{Fil}^i \Omega^\bullet_{X/T} = \operatorname{Im}\big(\pi^*(\Omega^i_{S/T}) \otimes_{\mathscr{O}_X} \Omega^{\bullet-i}_{X/T} \to \Omega^\bullet_{X/T}\big)$. As the sheaves $\Omega^i_{X/T}$ (resp. $\Omega^i_{S/T}$) for $i \geq 0$ are locally free $\mathscr{O}_X$–modules (resp. $\mathscr{O}_S$–modules) by smoothness, we see that $\operatorname{Gr}^i \Omega^\bullet_{X/T} \cong \pi^*(\Omega^i_{S/T}) \otimes_{\mathscr{O}_X} \Omega^{\bullet-i}_{X/S}$.

Applying the functor $\mathbb{R}^0\pi_*$ from the category of complexes of $\mathscr{O}_X$–modules to the category of $\mathscr{O}_S$–modules (whose right derived functors are $\mathbb{R}^i\pi_*$ for $i \geq 0$), this filtration induces a spectral sequence E abutting to the de Rham cohomology of X over T with terms

$$E_1^{a,b} = \mathbb{R}^{a+b}\pi_*(\pi^*(\Omega^a_{S/T}) \otimes_{\mathscr{O}_X} \Omega^{\bullet-a}_{X/S}) = \mathbb{R}^b\pi_*(\pi^*(\Omega^a_{S/T}) \otimes_{\mathscr{O}_X} \Omega^\bullet_{X/S})\,.$$

But we observe that $\Omega^a_{S/T}$ is a locally free $\mathscr{O}_S$–module and that the differential in the complex $\pi^*(\Omega^a_{S/T}) \otimes_{\mathscr{O}_X} \Omega^\bullet_{X/S}$ is $\pi^{-1}(\mathscr{O}_S)$–linear. Therefore,

$$\mathbb{R}^b\pi_*(\pi^*(\Omega^a_{S/T}) \otimes_{\mathscr{O}_X} \Omega^\bullet_{X/S}) \cong \Omega^a_{S/T} \otimes_{\mathscr{O}_S} \mathbb{R}^b\pi_*(\Omega^\bullet_{X/S}) = \Omega^a_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S)$$

and, using this isomorphism, we write

$$E_1^{a,b} = \Omega^a_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/T)\,.$$

Let us focus on the terms of E_1. For each $b \geq 0$, we have a complex $E_1^{\bullet,b}$ of the form

$$0 \longrightarrow \mathrm{H}^b_{\mathrm{dR}}(X/S) \xrightarrow{d_1^{0,b}} \Omega^1_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S) \xrightarrow{d_1^{1,b}} \Omega^2_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S) \longrightarrow \cdots$$

which we write $\Omega^\bullet_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S)$. For every $b \geq 0$, one checks that the differential $d_1^{0,b}\colon \mathrm{H}^b_{\mathrm{dR}}(X/S) \to \Omega^1_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S)$ is induced by a morphism $\mathscr{H}^b(\Omega^\bullet_{X/S}) \to \mathscr{H}^b(\pi^*(\Omega^1_{S/T}) \otimes_{\mathscr{O}_X} \Omega^\bullet_{X/S})$ defined on sections as follows. Take a section $[\omega]$ of $\mathscr{H}^b(\Omega^\bullet_{X/S})$ represented by ω in $\Omega^b_{X/S}$ and choose a lift $\widetilde{\omega}$ of ω to $\Omega^b_{X/T}$. Computing the connecting morphism of cohomology, we see that $d_{X/T}(\widetilde{\omega})$ is a section of $\mathrm{Fil}^1 \Omega^{b+1}_{X/T} = \mathrm{Im}\big(\pi^*(\Omega^1_{S/T}) \otimes_{\mathscr{O}_X} \Omega^b_{X/T} \to \Omega^{b+1}_{X/T}\big)$. We get a section $[d_{X/T}(\widetilde{\omega})]$ of $\mathscr{H}^b(\pi^*(\Omega^1_{S/T}) \otimes_{\mathscr{O}_X} \Omega^\bullet_{X/T})$ after projecting to the quotient $\mathrm{Gr}^1 \Omega^{b+1}_{X/T}$. The image of $[\omega]$ is then $[d_{X/T}(\widetilde{\omega})]$. From this construction, we deduce that $d_1^{0,b}$ is a connection on $\mathrm{H}^b_{\mathrm{dR}}(X/S)$. That is, for every open subset U of S and every $f \in \Gamma(U, \mathscr{O}_S)$ and $\eta \in \Gamma(U, \mathrm{H}^b_{\mathrm{dR}}(X/S))$,

$$d_1^{0,b}(f\eta) = d_{S/T}(f) \otimes \eta + f d_1^{0,b}(\eta)\,.$$

Definition 3.18. The *Gauss–Manin connection* on $\mathrm{H}^n_{\mathrm{dR}}(X/S)$ is the differential $\nabla = d_1^{0,n}\colon \mathrm{H}^n_{\mathrm{dR}}(X/S) \to \Omega^1_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^n_{\mathrm{dR}}(X/S)$ appearing in the first page of the spectral sequence

$$E_1^{a,b} = \Omega^a_{S/T} \otimes_{\mathscr{O}_S} \mathrm{H}^b_{\mathrm{dR}}(X/S) \implies \mathrm{H}^{a+b}_{\mathrm{dR}}(X/T)$$

described above.

Since E is the spectral sequence of a filtered object, the differentials

$$d_1^{a,b}\colon E_1^{a,b} \to E_1^{a+1,b}$$

can be computed as the connecting homomorphisms of the functors $\mathbb{R}^b\pi_*$ on the short exact sequence

$$0 \to \mathrm{Gr}^{a+1} \Omega^\bullet_{X/T} \to \mathrm{Fil}^a \Omega^\bullet_{X/T} / \mathrm{Fil}^{a+2} \Omega^\bullet_{X/T} \to \mathrm{Gr}^a \Omega^\bullet_{X/T} \to 0\,.$$

In the case we are most interested in, this exact sequence is quite simple. Suppose that $T = \mathrm{Spec}(K)$ for some field K, that S is a smooth curve over K and that $\pi\colon E = X \to S$ is an elliptic curve. Then, $\Omega^2_{S/K} = 0$ and so $\mathrm{Fil}^2 \Omega^\bullet_{E/K} = 0$. Hence, for $a = 0$, we have a short exact sequence

$$0 \to \pi^*(\Omega^1_{S/K}) \otimes_{\mathscr{O}_E} \Omega^{\bullet-1}_{E/S} \to \Omega^\bullet_{E/K} \to \Omega^\bullet_{E/S} \to 0$$

of complexes of $\mathscr{O}_E$–modules. Applying the functor $\mathbb{R}^0\pi_*$, we obtain a connecting

homomorphism

$$\begin{array}{ccc} \delta\colon \mathbb{R}^1\pi_*(\Omega^\bullet_{E/S}) & \longrightarrow & \mathbb{R}^2\pi_*(\pi^*(\Omega^1_{S/K}) \otimes_{\mathscr{O}_E} \Omega^{\bullet-1}_{E/S}) \\ \| & & \|\wr \\ \nabla\colon \mathrm{H}^1_{\mathrm{dR}}(E/S) & \longrightarrow & \Omega^1_{S/K} \otimes_{\mathscr{O}_S} \mathrm{H}^1_{\mathrm{dR}}(E/S) \end{array}$$

which is precisely the Gauss–Manin connection on $\mathrm{H}^1_{\mathrm{dR}}(E/S)$.

Now assume that K has positive characteristic p. By theorem 3.15, the Hodge filtration yields a short exact sequence

$$0 \to \pi_*(E/S) \to \mathrm{H}^1_{\mathrm{dR}}(E/S) \to \mathrm{R}^1\pi_*(\mathscr{O}_E) \to 0\,.$$

Set $\underline{\omega}_{E/S} = \pi_*(E/S)$. There is a canonical isomorphism $\mathrm{R}^1\pi_*(\mathscr{O}_E) \cong \underline{\omega}^{\otimes -1}_{E/S}$ given by Serre–Grothendieck duality, so we rewrite the Hodge filtration as

$$0 \to \underline{\omega}_{E/S} \to \mathrm{H}^1_{\mathrm{dR}}(E/S) \to \underline{\omega}^{\otimes -1}_{E/S} \to 0\,.$$

Using this, we define a morphism $\kappa\colon \underline{\omega}_{E/S} \to \underline{\omega}^{\otimes -1}_{E/S} \otimes_{\mathscr{O}_S} \Omega^1_{S/K}$ of $\mathscr{O}_S$–modules given by the composition

$$\underline{\omega}_{E/S} \hookrightarrow \mathrm{H}^1_{\mathrm{dR}}(E/S) \xrightarrow{\nabla} \mathrm{H}^1_{\mathrm{dR}}(E/S) \otimes_{\mathscr{O}_S} \Omega^1_{S/K} \twoheadrightarrow \underline{\omega}^{\otimes -1}_{E/S} \otimes_{\mathscr{O}_S} \Omega^1_{S/K}$$

(where ∇ is the Gauss–Manin connection).

Definition 3.19. In the above situation, $\kappa\colon \underline{\omega}_{E/S} \to \underline{\omega}^{\otimes -1}_{E/S} \otimes_{\mathscr{O}_S} \Omega^1_{S/K}$ induces a morphism of $\mathscr{O}_S$–modules $\mathrm{KS}\colon \underline{\omega}^{\otimes 2}_{E/S} \to \Omega^1_{S/K}$ which is called the *Kodaira–Spencer morphism* for E/S.

Theorem 3.20. *Let $N \geq 3$ such that $p \nmid N$ (i.e., N is invertible in K). Let $\mathbb{E}_K/Y(N)_K$ be the universal elliptic curve for $\Gamma(N)$ over K (see definition 2.19). The Kodaira–Spencer morphism $\mathrm{KS}\colon \underline{\omega}^{\otimes 2}_{\mathbb{E}_K/Y(N)_K} \to \Omega^1_{Y(N)_K/K}$ for $\mathbb{E}_K/Y(N)_K$ is an isomorphism of $\mathscr{O}_{Y(N)_K}$–modules.*

Proof. See lemma 7 of Diamond and Taylor's article [6]. □

3.4 Computations for the Tate curve

In the previous sections of this chapter we have introduced several geometric tools abstractly. In this section we apply the previous theory to the case of the

Tate curve and exhibit some explicit computations working mostly over the complex numbers (like in the introduction of the Tate curve in section 2.2). We follow sections A1.3 and A1.4 of Katz's article [8].

Unless otherwise stated, we use the same notation as in section 2.2. Let R be the subring of $\mathbb{C}((q))$ consisting of Laurent series of functions which are holomorphic on $\{\, q \in \mathbb{C} : 0 < |q| < 1 \,\}$. We regard R as a $\mathbb{C}$–algebra via the inclusion of constants. The Tate curve $\mathrm{Tate}(q)$ is defined over R by the affine equation

$$\mathrm{Tate}(q)\colon\ Y^2 = X^3 + a_4(q)X + a_6(q)$$

given in definition 2.5. (Here, unlike in section 2.2, we use capital letters for the affine coordinates X and Y because later we want to use x and y for the coordinates of the complex plane.) Our objective is to compute the Gauss–Manin connection $\nabla\colon\ \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \to \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \otimes_R \Omega^1_{R/\mathbb{C}}$.

For each $\tau \in \mathbb{H}$, we have an elliptic curve E_τ over $\mathbb{C}$ which corresponds to the compact Riemann surface $\mathbb{C}/\Lambda(\tau)$, where $\Lambda(\tau) = \mathbb{Z}\tau \oplus \mathbb{Z}$. It can be obtained from $\mathrm{Tate}(q)$ by pull-back under the map $q \mapsto q_\tau = e^{2\pi i\tau} : R \to \mathbb{C}$; that is, we have a cartesian diagram

$$\begin{array}{ccc} E_\tau & \longrightarrow & \mathrm{Tate}(q) \\ \downarrow & \ulcorner & \downarrow \pi \\ \mathrm{Spec}(\mathbb{C}) & \longrightarrow & \mathrm{Spec}(R) \end{array}$$

where the lower arrow corresponds to the morphism $R \to \mathbb{C}$ of $\mathbb{C}$–algebras defined by $q \mapsto q_\tau$. These morphisms are all the $\mathbb{C}$–points of $\mathrm{Spec}(R)$ over $\mathrm{Spec}(\mathbb{C})$, so we can study many properties of $\mathrm{Tate}(q)/R$ by studying $E_\tau/\mathbb{C}$ for every $\tau \in \mathbb{H}$ (i.e., by studying every fibre). In addition, we can use the complex analytic structure of each $E_\tau/\mathbb{C}$ (i.e., of $\mathbb{C}/\Lambda(\tau)$).

From now on, let τ denote a coordinate on $\mathbb{H}$ and set $a_\tau = \mathrm{Re}(\tau)$ and $b_\tau = \mathrm{Im}(\tau)$, so that $\tau = a_\tau + ib_\tau$. It is also convenient to interpret R as the ring of holomorphic functions $\mathbb{H} \to \mathbb{C}$ (using the change of variables $q = e^{2\pi i\tau}$). Consider a fundamental parallelogram P_τ for $\Lambda(\tau)$ centred at the origin. Let $\gamma_{1,\tau}$ be the line segment from the vertex $\frac{-1-\tau}{2}$ to the vertex $\frac{-1+\tau}{2}$ and let $\gamma_{2,\tau}$ be the line segment from the vertex $\frac{-1-\tau}{2}$ to the vertex $\frac{1-\tau}{2}$, as shown in figure 3.1. The singular homology classes of these two paths (which we write again $\gamma_{1,\tau}$ and $\gamma_{2,\tau}$) form a $\mathbb{Z}$–basis of $\mathrm{H}_1(E_\tau, \mathbb{Z})$ (where we identify E_τ with $\mathbb{C}/\Lambda(\tau)$). Furthermore, $\mathrm{H}_1(E_\tau, \mathbb{C}) \cong \mathrm{H}_1(E_\tau, \mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{C}$.

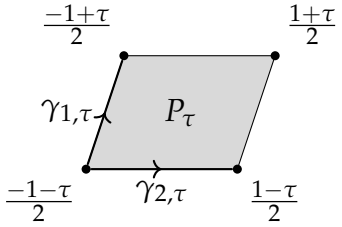


Figure 3.1: A basis of $\mathrm{H}_1(\mathbb{C}/\Lambda(\tau),\mathbb{Z})$.

There is a perfect pairing

$$\begin{aligned}(\,\cdot\,,\,\cdot\,)\colon \mathrm{H}_1(E_\tau,\mathbb{C})\times \mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C}) &\longrightarrow \mathbb{C}\\ ([\gamma],\theta) &\longmapsto \int_\gamma \theta\end{aligned}$$

which allows us to identify $\mathrm{H}_1(E_\tau,\mathbb{C})$ with the dual of $\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})$. Thus, we write $(\gamma^\vee_{1,\tau},\gamma^\vee_{2,\tau})$ for the basis of $\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})$ dual to $(\gamma_{1,\tau},\gamma_{2,\tau})$. Let z denote a coordinate on $\mathbb{C}/\Lambda(\tau)$ and set $x=\mathrm{Re}(z)$ and $y=\mathrm{Im}(z)$, so that $z=x+iy$. One checks easily that, in coordinates,

$$\gamma^\vee_{1,\tau}=\frac{1}{b_\tau}\,dy \quad\text{ and }\quad \gamma^\vee_{2,\tau}=\,dx-\frac{a_\tau}{b_\tau}\,dy\,.$$

That is, the forms defined by these formulae satisfy that $(\gamma_{i,\tau},\gamma^\vee_{j,\tau})=\delta_{ij}$.

On the other hand, there is another perfect pairing

$$\begin{aligned}\langle\,\cdot\,,\,\cdot\,\rangle\colon \mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})\times \mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C}) &\longrightarrow \mathbb{C}\\ (\theta_1,\theta_2) &\longmapsto \int_{P_\tau}\theta_1\wedge\theta_2\end{aligned}$$

which allows us to identify $\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})$ with its own dual. Putting these two dualities together, we obtain a basis $(\varphi_{1,\tau},\varphi_{2,\tau})$ of $\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})$ satisfying that $\langle\varphi_{i,\tau},\theta\rangle=(\gamma_{i,\tau},\theta)$ for all $\theta\in\mathrm{H}^1_{\mathrm{dR}}(E_\tau/\mathbb{C})$, $i\in\{\,1,2\,\}$. Using that

$$\int_{P_\tau} dx\wedge dy=b_\tau$$

(this is the area of P_τ), one checks easily that, in coordinates,

$$\varphi_{1,\tau}=\,dx-\frac{a_\tau}{b_\tau}\,dy \quad\text{ and }\quad \varphi_{2,\tau}=-\frac{1}{b_\tau}\,dy\,.$$

That is, the forms defined by these formulae satisfy that $\langle \varphi_{i,\tau}, \gamma_{j,\tau}^{\vee} \rangle = \delta_{ij}$. Also, $\langle \varphi_{1,\tau}, \varphi_{1,\tau} \rangle = 0 = \langle \varphi_{2,\tau}, \varphi_{2,\tau} \rangle$ and $\langle \varphi_{2,\tau}, \varphi_{1,\tau} \rangle = 1 = -\langle \varphi_{1,\tau}, \varphi_{2,\tau} \rangle$.

In this way, by letting τ vary, we obtain a basis (φ_1, φ_2) of $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R)$. It is quite easy to compute the Gauss–Manin connection in terms of this basis. Indeed, as we have seen at the end of section 3.3, the Gauss–Manin connection is the connecting homomorphism obtained from the short exact sequence

$$0 \to \Omega^{\bullet-1}_{\mathrm{Tate}(q)/R} \otimes_R \Omega^1_{R/\mathbb{C}} \to \Omega^{\bullet}_{\mathrm{Tate}(q)/\mathbb{C}} \to \Omega^{\bullet}_{\mathrm{Tate}(q)/R} \to 0$$

and the functor $\mathbb{R}^0\pi_*$, where $\pi\colon \mathrm{Tate}(q) \to \mathrm{Spec}(R)$ is the structure morphism. Hence, there is an exact sequence

$$\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/\mathbb{C}) \longrightarrow \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \xrightarrow{\nabla} \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \otimes_R \Omega^1_{R/\mathbb{C}}\,.$$

But, since φ_1 and φ_2 are defined in terms of complex (singular) homology, they are in the image of $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/\mathbb{C})$. Therefore, $\nabla(\varphi_1) = 0 = \nabla(\varphi_2)$.

We want to express the Gauss–Manin connection in terms of the more conventional basis (ω, η) of $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R)$ given by the cohomology classes of $\frac{dX}{Y}$ and $\frac{X\,dX}{Y}$, respectively. Again, we work analytically on fibres given by $\tau \in \mathbb{H}$. We have $\omega_\tau = \frac{dX}{Y} = dz$ and $\eta_\tau = \frac{X\,dX}{Y} = \wp(z;\Lambda(\tau))\,dz$. Define, for $i \in \{1,2\}$, $\omega_{i,\tau} = (\gamma_{i,\tau}, \omega_\tau) = \langle \varphi_{i,\tau}, \omega_\tau \rangle$ and $\eta_{i,\tau} = (\gamma_{i,\tau}, \eta_\tau) = \langle \varphi_{i,\tau}, \eta_\tau \rangle$. We can express $\omega_\tau = \omega_{2,\tau}\varphi_{1,\tau} - \omega_{1,\tau}\varphi_{2,\tau}$ and $\eta_\tau = \eta_{2,\tau}\varphi_{1,\tau} - \eta_{1,\tau}\varphi_{2,\tau}$ or, in matrix form,

$$\begin{pmatrix} \omega_\tau \\ \eta_\tau \end{pmatrix} = \begin{pmatrix} \omega_{2,\tau} & -\omega_{1,\tau} \\ \eta_{2,\tau} & -\eta_{1,\tau} \end{pmatrix} \begin{pmatrix} \varphi_{1,\tau} \\ \varphi_{2,\tau} \end{pmatrix}.$$

Furthermore, a straight-forward computation shows that

$$\omega_{1,\tau} = \int_{\gamma_{1,\tau}} dz = \tau \quad \text{and} \quad \omega_{2,\tau} = \int_{\gamma_{2,\tau}} dz = 1\,.$$

Lemma 3.21. *The period $\eta_{2,\tau}$ is $-\frac{\pi^2}{3}E_2(\tau)$.*

Sketch of the proof. We consider the Weierstrass zeta function

$$\zeta(z;\Lambda(\tau)) = \frac{1}{z} + \sum_{l \in \Lambda(\tau)}{}' \Big(\frac{1}{z-l} + \frac{1}{l} + \frac{z}{l^2} \Big)\,,$$

which satisfies that $\zeta'(z;\Lambda(\tau)) = -\wp(z;\Lambda(\tau))$. It turns out that the series for $\zeta(z;\Lambda(\tau))$ is absolutely convergent and so the order of summation is irrelevant.

The elements of $\Lambda(\tau)$ are of the form $m\tau + n$ for $m, n \in \mathbb{Z}$. With this notation, we consider the sum first over n and then over m. Now we can decompose the previous series as

$$\zeta(z;\Lambda(\tau)) = \sum_{m\in\mathbb{Z}} f_m(z) + g(z) + h(z)\,,$$

where

$$\begin{aligned} f_m(z) &= \frac{1}{z+m\tau} + \sum_{n=1}^{\infty}\Big(\frac{1}{z+m\tau+n} + \frac{1}{z+m\tau-n}\Big)\,,\\ g(z) &= \sum_{l\in\Lambda(\tau)}{}' \frac{1}{l} = 0\,,\\ h(z) &= \sum_{m\in\mathbb{Z}}\sum_{n\in\mathbb{Z}}{}' \frac{z}{(m\tau+n)^2} = z\frac{\pi^2}{3}E_2(\tau)\,. \end{aligned}$$

By definition, $\wp(z+l;\Lambda(\tau)) = \wp(z;\Lambda(\tau))$ for every $l \in \Lambda(\tau)$. This shows that the function $\zeta(z+l;\Lambda(\tau)) - \zeta(z;\Lambda(\tau))$ is constant (its derivative is 0) for every $l \in \Lambda(\tau)$. In particular,

$$\begin{aligned} \eta_{2,\tau} &= \int_{\gamma_{2,\tau}} \wp(z;\Lambda(\tau))\,dz = \zeta\Big(\frac{-1-\tau}{2};\Lambda(\tau)\Big) - \zeta\Big(\frac{1-\tau}{2};\Lambda(\tau)\Big)\\ &= \zeta\Big(\frac{-1}{2};\Lambda(\tau)\Big) - \zeta\Big(\frac{1}{2};\Lambda(\tau)\Big) = -2\zeta\Big(\frac{1}{2};\Lambda(\tau)\Big) \end{aligned}$$

(because $\zeta(z;\Lambda(\tau))$ is odd).

It remains to prove that

$$\sum_{m\in\mathbb{Z}} f_m\Big(\frac{1}{2}\Big) = 0\,.$$

But, using the formula at the beginning of proposition 1.4, we can write

$$f_m(z) = \pi\frac{\cos(\pi(z+m\tau))}{\sin(\pi(z+m\tau))} = \pi\frac{e^{2\pi iz}q_\tau^m + 1}{e^{2\pi iz}q_\tau^m - 1}\,.$$

Therefore,

$$\begin{aligned} \sum_{m\in\mathbb{Z}} f_m\Big(\frac{1}{2}\Big) &= f_0\Big(\frac{1}{2}\Big) + \sum_{m=1}^{\infty}\Big[f_m\Big(\frac{1}{2}\Big) + f_{-m}\Big(\frac{1}{2}\Big)\Big]\\ &= -\pi\sum_{m=1}^{\infty}\Big(\frac{1-q_\tau^m}{1+q_\tau^m} + \frac{1-q_\tau^{-m}}{1+q_\tau^{-m}}\Big) = 0\,, \end{aligned}$$

as required. □

Lemma 3.22. *The periods $\omega_{1,\tau}$, $\omega_{2,\tau}$, $\eta_{1,\tau}$ and $\eta_{2,\tau}$ satisfy the Legendre relation*

$$\eta_{1,\tau}\omega_{2,\tau} - \eta_{2,\tau}\omega_{1,\tau} = 2\pi i\,.$$

Sketch of the proof. We integrate $\zeta(z;\Lambda(\tau))$ along the boundary of the fundamental parallelogram P_τ. On the one hand, in P_τ, $\zeta(z;\Lambda(\tau))$ has only a simple pole with residue 1 at the origin. Hence, by the residue theorem,

$$\int_{\partial P_\tau} \zeta(z;\Lambda(\tau))\,dz = 2\pi i\,.$$

On the other hand, we know that $\wp(z;\Lambda(\tau))$ is periodic of period $\Lambda(\tau)$. Thus, since $1, \tau \in \Lambda(\tau)$, we observe that $\wp(z+1;\Lambda(\tau)) = \wp(z;\Lambda(\tau)) = \wp(z+\tau;\Lambda(\tau))$ and, by integrating these relations, we deduce that

$$\begin{aligned}\zeta(z+\tau;\Lambda(\tau)) - \zeta(z;\Lambda(\tau)) &= \zeta\Big(\frac{-1+\tau}{2};\Lambda(\tau)\Big) - \zeta\Big(\frac{-1-\tau}{2};\Lambda(\tau)\Big)\\ &= \int_{\gamma_{1,\tau}} -\wp(z;\Lambda(\tau))\,dz = -\eta_{1,\tau}\end{aligned}$$

and, similarly,

$$\zeta(z+1;\Lambda(\tau)) - \zeta(z;\Lambda(\tau)) = \int_{\gamma_{2,\tau}} -\wp(z;\Lambda(\tau))\,dz = -\eta_{2,\tau}\,.$$

Therefore, the integrals of $\zeta(z;\Lambda(\tau))$ along opposite sides of the parallelogram P_τ *almost* cancel out and we get that

$$\int_{\partial P_\tau} \zeta(z;\Lambda(\tau))\,dz = \int_{\gamma_{1,\tau}} -\eta_{2,\tau}\,dz - \int_{\gamma_{2,\tau}} -\eta_{1,\tau}\,dz = -\eta_{2,\tau}\omega_{1,\tau} + \eta_{1,\tau}\omega_{2,\tau}\,.$$

Equating the two expressions for the integral, we obtain the desired relation. □

The Legendre relation computes precisely the determinant of the matrix expressing (ω_τ, η_τ) in terms of $(\varphi_{1,\tau}, \varphi_{2,\tau})$. Inverting that matrix, we obtain that

$$2\pi i\begin{pmatrix}\varphi_{1,\tau}\\ \varphi_{2,\tau}\end{pmatrix} = \begin{pmatrix}\eta_{1,\tau} & -\omega_{1,\tau}\\ \eta_{2,\tau} & -\omega_{2,\tau}\end{pmatrix}\begin{pmatrix}\omega_\tau\\ \eta_\tau\end{pmatrix}.$$

Again, letting τ vary, we can regard $\omega_{i,\tau}$ and $\eta_{i,\tau}$ as functions of τ and we get in

this way $\omega_i, \eta_i \in R$, $i \in \{1,2\}$, such that

$$2\pi i\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix} = \begin{pmatrix}\eta_1 & -\omega_1\\ \eta_2 & -\omega_2\end{pmatrix}\begin{pmatrix}\omega\\ \eta\end{pmatrix}.$$

Next we want to apply the Gauss–Manin connection to the previous relation. Before that, we notice that $\Omega^1_{R/\mathbb{C}} \cong R\,d\tau$ (because the derivative of a holomorphic function is again holomorphic). Since the derivation $\partial_\tau = \frac{d}{d\tau}$ is dual to $d\tau$, it suffices to study $\nabla_\tau = \nabla(\partial_\tau)\colon \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \to \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R)$. Indeed, for a section ψ of $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R)$, we observe that $\nabla(\psi) = \nabla_\tau(\psi)\otimes d\tau$.

All in all, applying ∇_τ to the relation between (φ_1,φ_2) and (ω,η), we obtain that

$$\begin{pmatrix}0\\ 0\end{pmatrix} = \begin{pmatrix}\partial_\tau\eta_1 & -\partial_\tau\omega_1\\ \partial_\tau\eta_2 & -\partial_\tau\omega_2\end{pmatrix}\begin{pmatrix}\omega\\ \eta\end{pmatrix} + \begin{pmatrix}\eta_1 & -\omega_1\\ \eta_2 & -\omega_2\end{pmatrix}\begin{pmatrix}\nabla_\tau(\omega)\\ \nabla_\tau(\eta)\end{pmatrix}.$$

Again by the Legendre relation, we can invert the last matrix and express

$$\begin{pmatrix}\nabla_\tau(\omega)\\ \nabla_\tau(\eta)\end{pmatrix} = -\frac{1}{2\pi i}\begin{pmatrix}\omega_2 & -\omega_1\\ \eta_2 & -\eta_1\end{pmatrix}\begin{pmatrix}\partial_\tau\eta_1 & -\partial_\tau\omega_1\\ \partial_\tau\eta_2 & -\partial_\tau\omega_2\end{pmatrix}\begin{pmatrix}\omega\\ \eta\end{pmatrix}$$

and, plugging in the values of ω_1, ω_2 and η_2 and using the Legendre relation, we conclude that

$$\begin{pmatrix}\nabla_\tau(\omega)\\ \nabla_\tau(\eta)\end{pmatrix} = -\frac{1}{2\pi i}\begin{pmatrix}-\frac{\pi^2}{3}E_2(\tau) & -1\\ \frac{\pi^4}{9}E_2(\tau)^2 + \frac{2\pi^3 i}{3}E_2'(\tau) & \frac{\pi^2}{3}E_2(\tau)\end{pmatrix}\begin{pmatrix}\omega\\ \eta\end{pmatrix}.$$

Now consider the canonical differential ω_{can} on the Tate curve. Recall that, on fibres, $\omega_{\mathrm{can},\tau} = 2\pi i\,dz$ and $\omega_\tau = dz$. Therefore, $\omega_{\mathrm{can}} = 2\pi i\omega$. Set $\eta_{\mathrm{can}} = \frac{1}{2\pi i}\eta$, which is then *dual* to ω_{can}. Consider also the derivation $q\frac{d}{dq} = \frac{1}{2\pi i}\frac{d}{d\tau}$. We can compute $\nabla\big(q\frac{d}{dq}\big)$ as

$$\nabla\Big(q\frac{d}{dq}\Big)\begin{pmatrix}\omega\\ \eta\end{pmatrix} = \frac{1}{2\pi i}\begin{pmatrix}\nabla_\tau(\omega)\\ \nabla_\tau(\eta)\end{pmatrix}$$

and then express it in terms of $(\omega_{\mathrm{can}},\eta_{\mathrm{can}})$. In this way, we have proved the following result.

Theorem 3.23. *The Gauss–Manin connection of* $\mathrm{Tate}(q)/R/\mathbb{C}$ *is defined by*

$$\nabla\Big(q\frac{d}{dq}\Big)\begin{pmatrix}\omega_{\mathrm{can}}\\ \eta_{\mathrm{can}}\end{pmatrix} = \begin{pmatrix}-\frac{1}{12}P & 1\\ -\frac{1}{144}\Big(P^2 - 12q\frac{d}{dq}P\Big) & \frac{1}{12}P\end{pmatrix}\begin{pmatrix}\omega_{\mathrm{can}}\\ \eta_{\mathrm{can}}\end{pmatrix}$$

(where P is the q–expansion of the Eisenstein series E_2; as in section 1.3).

Corollary 3.24. *In the above situation, the image of $\omega_{\mathrm{can}}^{\otimes 2}$ under the Kodaira–Spencer morphism* $\mathrm{KS}\colon \underline{\omega}^{\otimes 2}_{\mathrm{Tate}(q)/R} \to \Omega^1_{R/\mathbb{C}}$ *is the differential $\frac{dq}{q}$. In particular, the map* KS *is an isomorphism of R–modules.*

Proof. The Hodge filtration

$$0 \to \underline{\omega}_{\mathrm{Tate}(q)/R} \to \mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/R) \to \mathrm{R}^1\pi_*(\mathscr{O}_{\mathrm{Tate}(q)}) \to 0$$

allows us to project η_{can} to an element $\widetilde{\eta}_{\mathrm{can}}$ of $\mathrm{R}^1\pi_*(\mathscr{O}_{\mathrm{Tate}(q)})$ which is the Serre–Grothendieck dual of ω_{can}. Thus, the projection of

$$\nabla\Big(q\frac{d}{dq}\Big)(\omega_{\mathrm{can}}) = \frac{-P}{12}\omega_{\mathrm{can}} + \eta_{\mathrm{can}}$$

is also $\widetilde{\eta}_{\mathrm{can}}$. Since the differential $\frac{dq}{q}$ is the dual of the derivation $q\frac{d}{dq}$, we see that the image of ω_{can} under $\kappa\colon \underline{\omega}_{\mathrm{Tate}(q)/R} \to \underline{\omega}^{\otimes -1}_{\mathrm{Tate}(q)/R} \otimes_R \Omega^1_{R/\mathbb{C}}$ is

$$\kappa(\omega_{\mathrm{can}}) = \widetilde{\eta}_{\mathrm{can}} \otimes \frac{dq}{q}\,,$$

whence the corollary follows. □

Remark. Since the canonical differential ω_{can} for the Tate curve is always defined (regardless of the base ring) and the Kodaira–Spencer morphism commutes with base change, the corollary is true for the Tate curve over any base ring, even if we have used the complex analytic structure over $\mathbb{C}$ for the computations.

Chapter 4

Modular forms in characteristic p

Throughout this chapter, let p be a fixed prime number. This chapter explains the structure of the algebra of Katz's modular forms over an algebraically closed field of characteristic p analogously to the exposition of the classical case in section 1.3. In the context of Katz's modular forms, it is not enough to study q–expansions as power series. Instead, most proofs require a geometric approach using the theory introduced in chapters 2 and 3.

In characteristic p, one can define the Hasse invariant, a modular form which describes the action of the Frobenius morphism on the de Rham cohomology of elliptic curves. The Hasse invariant plays a fundamental role in the study of the algebra of modular forms because it is *essentially* the only modular form with q–expansions equal to 1. That is, we can compare modular forms with their q–expansions by means of the Hasse invariant. Moreover, the Hasse invariant induces a filtration on the algebra of modular forms and it is used to define a certain derivation which respects the filtration. The properties of this derivation help us to understand the filtration.

This chapter explains the theory mentioned in Katz's paper [9]. The main result of ibid. is also the main topic of the last section. Before that, the first section presents the several characterizations of the Hasse invariant appearing in section 12.4 of Katz and Mazur's book [12] (some of which are also in section 2.0 of Katz's article [8]) and the second section explains the main result of section 2.7 of Cais's notes [2].

4.1 The Hasse invariant

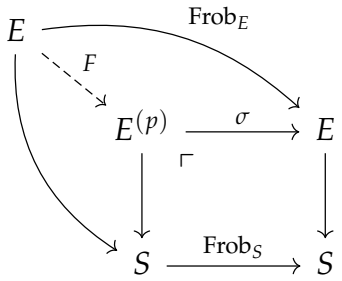
We introduce the Hasse invariant, regarded as a modular form, following section 12.4 of Katz and Mazur's book [12]. Unless otherwise stated, all results appearing in this section are proved in ibid.

Consider a pair $(E/R, \omega)$, where R is an $\mathbb{F}_p$–algebra, E is an elliptic curve over R and ω is a basis of $\mathrm{H}^0(\mathrm{Spec}(R), \underline{\omega}_{E/R}) = \mathrm{H}^0(E, \Omega^1_{E/R})$. By Serre duality, the R–module $\mathrm{H}^1(E, \mathscr{O}_E)$ is dual to $\mathrm{H}^0(E, \Omega^1_{E/R})$. Let η be the basis of $\mathrm{H}^1(E, \mathscr{O}_E)$

dual to ω. Let $\mathrm{Frob}_E\colon E \to E$ be the absolute Frobenius endomorphism (given by the p–th power endomorphism of $\mathscr{O}_E$). We get an induced $\mathbb{F}_p$–linear morphism $\mathrm{Frob}_E^*\colon \mathrm{H}^1(E,\mathscr{O}_E) \to \mathrm{H}^1(E,\mathscr{O}_E)$ and so

$$\mathrm{Frob}_E^*(\eta) = A(E/R,\omega)\eta \quad \text{for some } A(E/R,\omega) \in R\,.$$

Recall that the absolute Frobenius endomorphism fits in a commutative diagram



(where $S = \mathrm{Spec}(R)$), so we can also write $F^*(\eta^{(p)}) = A(E/R,\omega)\eta$.

Moreover, if we replace ω with $\lambda\omega$ for some $\lambda \in R^\times$, we obtain $\lambda^{-1}\eta$ instead of η and then

$$\mathrm{Frob}_E^*(\lambda^{-1}\eta) = \lambda^{-p}\,\mathrm{Frob}_E^*(\eta) = \lambda^{-p}A(E/R,\omega)\eta = \lambda^{1-p}A(E/R,\omega)(\lambda^{-1}\eta)\,,$$

whence $A(E/R,\lambda\omega) = \lambda^{1-p}A(E/R,\omega)$. In this way, we have constructed an element $A \in F(\mathbb{F}_p;\Gamma(1),p-1)$.

Definition 4.1. The *Hasse invariant* is the modular form $A \in F(\mathbb{F}_p;\Gamma(1),p-1)$ which satisfies that $\mathrm{Frob}_E^*(\eta) = A(E/R,\omega)\eta$ for each triple $(E/R,\omega,\eta)$ as above.

Remark. We can regard the Hasse invariant as a modular form for $\Gamma(N)$ for any $N \in \mathbb{N}$ and with coefficients in any $\mathbb{F}_p$–algebra.

Theorem 4.2. *The q–expansion of the Hasse invariant is equal to* 1 *(in $\mathbb{F}_p((q))$).*

Proof. Let R be an $\mathbb{F}_p$–algebra and consider an elliptic curve E over R. By Serre duality, the R–module $\mathrm{H}^1(E,\mathscr{O}_E)$ is dual to $\mathrm{H}^0(E,\Omega^1_{E/R})$ which, in turn, is dual to the restricted Lie algebra (of characteristic p) of left invariant R–derivations of E (regarded as an R–module). Thus, we identify the elements of $\mathrm{H}^1(E,\mathscr{O}_E)$ with left invariant derivations. The action of Frob_E^* on $\mathrm{H}^1(E,\mathscr{O}_E)$ corresponds to taking the p–th iterate of a left invariant derivation.

Consider the Tate curve $\mathrm{Tate}(q)$ over $\mathbb{F}_p((q))$ together with its canonical differential ω_{can}. There is a local parameter X on the completion of $\mathrm{Tate}(q)$ along

its identity section in terms of which ω_{can} is $\frac{dX}{1+X}$. Let D be the left invariant derivation dual to ω_{can}, so that $D(X) = 1 + X$. Therefore, $D(1 + X) = 1 + X$. We deduce that $D^p(1 + X) = 1 + X$ and so $D^p(X) = 1 + X = D(X)$, which implies that $D^p = D$. By the observations in the previous paragraph, we conclude that $A(\text{Tate}(q)/\mathbb{F}_p((q)), \omega_{\text{can}}) = 1$. □

The proof of theorem 4.2 hints at an alternative definition of the Hasse invariant in terms of derivations. Let E be an elliptic curve over $S = \text{Spec}(R)$ for an $\mathbb{F}_p$–algebra R with a basis ω of $\text{H}^0(E, \Omega^1_{E/R})$. Consider the relative Frobenius morphism $F\colon E \to E^{(p)}$ and the Verschiebung morphism $V\colon E^{(p)} \to E$. Also, write $m\colon E \times_S E \to E$ for the group law morphism of the elliptic curve and $e\colon S \to E$ for its identity section. We work locally as in section 3.1 to describe F and V. Let $U = \text{Spec}(B)$ be an affine open subset of E. Let d be the basis of $\text{Der}_R(B, R)$ dual to the differential $\omega\big|_U$. The left invariant R–derivation D corresponding to d is the composition

$$B \xrightarrow{m^\sharp} B \otimes_R B \xrightarrow{\text{id}_B \otimes d} B \otimes_R R \cong B$$

(here, we view R as an R–module through the identity) and we can then recover $d = e^\sharp \circ D$. We want to express the map $D \mapsto D^p = D \circ \overset{(p)}{\cdot} \circ D$ (corresponding to the action of Frob_E^* used to define the Hasse invariant) in terms of the Verschiebung morphism.

In what follows, the notation $\cdot^{(p)}$ denotes the pull-back of $\cdot$ under Frob_S. The tangent map $\text{tg}(V)\colon \text{Der}_R(B^{(p)}, R) \to \text{Der}_R(B, R)$ maps $d^{(p)} \in \text{Der}_R(B^{(p)}, R)$ to the composition

$$B \xrightarrow{V^\sharp} B^{(p)} \xrightarrow{d^{(p)}} R\,.$$

We claim that the left invariant derivation $\text{tg}(V)(D^{(p)})$ (corresponding to the derivation $\text{tg}(V)(d^{(p)})$) is precisely D^p. Indeed, we have a commutative diagram

$$\begin{array}{ccccccccc}
B & \xrightarrow{m^\sharp} & B \otimes_R B & \xrightarrow{\text{id}_B \otimes (m^p)^\sharp} & B \otimes_R T^p(B) & \xrightarrow{\text{id}_B \otimes D^{\otimes p}} & B \otimes_R T^p(B) & \xrightarrow{\text{id}_B \otimes (e^\sharp)^{\otimes p}} & B \\
\| & & & & \uparrow & & \uparrow & & \| \\
 & & & & B \otimes_R S^p(B) & \xrightarrow{\text{id}_B \otimes D^{\otimes p}} & B \otimes_R S^p(B) & & \\
 & & & & \uparrow & & \uparrow & & \\
B & \xrightarrow{m^\sharp} & B \otimes_R B & \xrightarrow{\text{id}_B \otimes V^\sharp} & B \otimes_R B^{(p)} & \xrightarrow{\text{id}_B \otimes D^{(p)}} & B \otimes_R B^{(p)} & \xrightarrow{\text{id}_B \otimes (e^{(p)})^\sharp} & B
\end{array}$$

$$\text{id}_B \otimes d^{(p)}$$

where we used the same notation as in the local description of V in section 3.1 and the downmost vertical arrows are given by the map

$$\begin{aligned} B^{(p)} &\to S^p(B) \\ b \otimes a &\mapsto a(b \otimes \overset{(p)}{\cdots} \otimes b) \end{aligned}$$

which induces the isomorphism $\varphi_{B/R}\colon B^{(p)} \to S^p(B)/J$ (see lemma 3.3). But the first row of the diagram is D^p and the last row is $\mathrm{tg}(V)(D^{(p)})$, which proves the claim. We conclude that $\mathrm{tg}(V)(D^{(p)}) = A\big(U/R, \omega\big|_U\big)D$.

Repeating the previous construction on an affine open covering of E, we obtain that $\mathrm{tg}(V)(D^{(p)}) = A\big(E/R, \omega\big)D$ for the left invariant derivation D dual to ω and its pull-back $D^{(p)}$ under Frob_S.

We can compute the Hasse invariant in yet another way. Take, as above, an elliptic curve E over an $\mathbb{F}_p$–algebra R with a basis ω of $\mathrm{H}^0(E, \Omega^1_{E/R})$. The Cartier isomorphism $\mathcal{C}_1\colon \mathscr{H}^1(F_*(\Omega^\bullet_{E/R})) \to \Omega^1_{E^{(p)}/R}$ induces an R–linear map

$$\begin{array}{ccc} \mathrm{H}^0(E, \Omega^1_{E/R}) \cong \mathrm{H}^0(E^{(p)}, F_*(\Omega^1_{E/R})) & \longrightarrow & \mathrm{H}^0(E^{(p)}, \mathscr{H}^1(F_*(\Omega^\bullet_{E/R}))) \\ & \overset{C}{\searrow} & \wr\big\| \downarrow \mathcal{C}_1 \\ & & \mathrm{H}^0(E^{(p)}, \Omega^1_{E^{(p)}/R}) \end{array}$$

called the *Cartier operator*. Locally, F^* corresponds to taking p–th powers of the generators of $\mathscr{O}_E$ over R, whereas C corresponds to taking p–th roots. Thus, we see that C is obtained from F^* by Serre duality. Since F^* maps $\eta^{(p)}$ (whose dual is $\omega^{(p)}$) to $A(E/R, \omega)\eta$ (whose dual is $A(E/R, \omega)^{-1}\omega$), we conclude that

$$C(\omega) = A(E/R, \omega)\omega^{(p)}\,.$$

The next result is lemma 3.6.1 of Katz's article [8].

Lemma 4.3. *Let R be an $\mathbb{F}_p$–algebra and let E be an elliptic curve over R with a basis ω of $\underline{\omega}_{E/R}$. If X is a parameter for the formal group of E/R for which*

$$\omega = \Big(\sum_{n\geq 0} a_n X^n\Big)\, dX$$

with $a_0 = 1$, then

$$a_{p^n-1} = A(E/R, \omega)^{(p^n-1)/(p-1)} \quad \textit{for every } n \in \mathbb{N}\,.$$

Proof. On the one hand, we know that $C(\omega) = A(E/R, \omega)\omega^{(p)}$. On the other hand, we can compute *locally*

$$C(a_n X^n\, dX) = \begin{cases} 0 & \text{if } p \nmid n+1\,, \\ a_n (X^{(p)})^{-1+(n+1)/p}\, dX^{(p)} & \text{if } p \mid n+1\,, \end{cases}$$

using the properties defining $\mathcal{C}_1$ (see theorem 3.12). Therefore,

$$A(E/R,\omega) \sum_{m\geq 0} a_m^p (X^{(p)})^m\, dX^{(p)} = C(\omega) = \sum_{m\geq 0} a_{p(m+1)-1} (X^{(p)})^m\, dX^{(p)}$$

and, for $m = p^n - 1$, we find that

$$a_{p^{n+1}-1} = A(E/R,\omega) a_{p^n-1}^p\,.$$

The lemma follows from this equality by induction on n (as $a_0 = 1$). □

Having seen several interpretations of the Hasse invariant, we prove a fundamental result which we use later to determine the structure of the algebra of modular forms over an algebraically closed field of characteristic p. It is theorem 12.4.3 of Katz and Mazur's book [12].

Theorem 4.4 (Igusa). *Let K be a perfect field of characteristic p and let $(R, \mathfrak{m})$ be an artinian local K–algebra with residue field K. Let E be an elliptic curve over R and consider the Verschiebung morphism $V_{E/R}\colon E^{(p)} \to E$. If $\operatorname{tg}(V) = 0$, then there exist a supersingular elliptic curve E_0 over K and an R–isomorphism $E_0 \otimes_K R \cong E$.*

Proof. We use the formal groups of the elliptic curves E and $E^{(p)}$. Let X be a parameter for the formal group of E such that, for each $(p-1)$–th root of unity $\zeta \in \mathbb{Z}_p^\times$, $[\zeta](X) = \zeta X$ and let $X^{(p)}$ be the induced parameter for the formal group of $E^{(p)}$. Using these parameters, we can express

$$V_{E/R}(X) = \sum_{n\geq 1} a_n (X^{(p)})^n$$

with $a_1 = \operatorname{tg}(V) = 0$. On the other hand, since $V_{E/R}$ is a homomorphism of group schemes, $V_{E/R}([\zeta](X)) = [\zeta](V_{E/R}(X))$ for every $(p-1)$–th root of unity $\zeta \in \mathbb{Z}_p^\times$. That is to say, $V_{E/R}(\zeta X) = \zeta V_{E/R}(X)$ or, comparing coefficients of the power series expansions, $a_n \zeta^n = a_n \zeta$ for all $n \in \mathbb{N}$. But, if ζ is a primitive $(p-1)$–th root of unity in $\mathbb{Z}_p^\times$, $\zeta - \zeta^n$ is invertible in $\mathbb{Z}_p$ for all $n \in \mathbb{N}$ such that

$n \not\equiv 1 \bmod p-1$. We conclude that $a_n = 0$ unless $n \equiv 1 \bmod p-1$. All in all,

$$V_{E/R}(X) = \sum_{m\geq 1} a_{m(p-1)+1}(X^{(p)})^{m(p-1)+1} = a_p(X^{(p)})^p + \cdots .$$

From this last equation and using that $F_{E/R}(X^{(p)}) = X^p$, we see that

$$[p](X) = (V_{E/R} \circ F_{E/R})(X) = \sum_{m\geq 1} a_{m(p-1)+1} X^{mp(p-1)+p} = a_p X^{p^2} + \cdots .$$

Reducing coefficients modulo $\mathfrak{m}$, we observe that the formal group of the elliptic curve $\overline{E} = E \otimes_R K$ over K has height ≥ 2, so $\overline{E}$ must be supersingular and the height is 2. That is, $a_p \notin \mathfrak{m}$. Thus, $V_{E/R}(X) = (X^{(p)})^p u(X^{(p)})$ for some $u \in R[\![X^{(p)}]\!]^\times$. In particular, in terms of the formal group, $\operatorname{Ker}(V_{E/R})$ is defined by the equation $(X^{(p)})^p = 0$. But this equation defines $\operatorname{Ker}(F_{E^{(p)}/R})$ as well. In this way, we obtain a commutative diagram

$$\begin{array}{ccc} E^{(p)}/\operatorname{Ker}(V_{E/R}) & = & E^{(p)}/\operatorname{Ker}(F_{E^{(p)}/R}) \\ {\scriptstyle V_{E/R}}\downarrow \wr & & \wr \downarrow {\scriptstyle F_{E^{(p)}/R}} \\ E & \dashrightarrow & E^{(p^2)} \end{array}$$

and the last row must be the isomorphism given by taking p^2–th powers on R. Iterating this construction, we obtain a sequence of isomorphisms

$$E \xrightarrow{\cong} E^{(p^2)} \xrightarrow{\cong} E^{(p^4)} \xrightarrow{\cong} \cdots \xrightarrow{\cong} E^{(p^{2n})} \xrightarrow{\cong} \cdots .$$

As $(R, \mathfrak{m})$ is an artinian local ring, there exists some $n \gg 0$ such that $\mathfrak{m}^{p^{2n}} = 0$. Thus, for every $a \in R$ and every $b \in \mathfrak{m}$, $(a+b)^{p^{2n}} = a^{p^{2n}} + b^{p^{2n}} = a^{p^{2n}}$. This shows that the morphism $\cdot^{p^{2n}} : R \to R$ given by $a \mapsto a^{p^{2n}}$ factors through the residue field K or, equivalently, that the diagram

$$\begin{array}{ccc} R & \xrightarrow{\cdot^{p^{2n}}} & R \\ \downarrow & & \uparrow \\ K & \xrightarrow{\cdot^{p^{2n}}} & K \end{array}$$

is commutative. Therefore, taking pull-backs of E, we obtain R–isomorphisms $E \cong E^{(p^{2n})} \cong ((E \otimes_R K)^{(p^{2n})}) \otimes_K R$ and we can take $E_0 = (E \otimes_R K)^{(p^{2n})} = \overline{E}^{(p^{2n})}$. Note that, as K is a perfect field, $\cdot^{p^{2n}} : K \to K$ is an isomorphism and so $\overline{E}^{(p^{2n})} \cong \overline{E}$

(and we have seen that $\overline{E}/K$ is supersingular). □

Corollary 4.5. *Let K be an algebraically closed field of characteristic p and let $N \in \mathbb{N}$ such that $N \geq 3$ and $p \nmid N$. The Hasse invariant $A \in M(K; \Gamma(N), p-1)$ has only simple zeros (cf. lemma 1.40).*

Proof. Since the q–expansions of A are all equal to 1, the zeros of A must be points of $Y(N)_K$ (regarded as a subscheme of $X(N)_K$). Write $Y = Y(N)_K$ and $\mathbb{E} = \mathbb{E}_K$ for the universal elliptic curve over Y.

Let y be a closed point of Y at which A has a zero. Since Y is a smooth curve over the algebraically closed field K, the local ring $\mathscr{O}_{Y,y}$ is a discrete valuation ring with residue field K and also a K–algebra. Set $S = \operatorname{Spec}(\mathscr{O}_{Y,y})$. We have a cartesian diagram

$$\begin{array}{ccc} E & \longrightarrow & \mathbb{E} \\ \downarrow & \ulcorner & \downarrow \\ S & \xrightarrow{g} & Y \end{array}$$

(where g is the canonical morphism). We want to see that $g^*(A) \in \mathrm{H}^0(S, \underline{\omega}_{E/S}^{\otimes p-1})$ has a simple zero at the closed point of S.

Let $\mathfrak{m}_y$ be the maximal ideal of $\mathscr{O}_{Y,y}$. Take a basis ω of $\mathrm{H}^0(S, \underline{\omega}_{E/S})$ and consider the Hasse invariant $A(E/S, \omega) \in \mathfrak{m}_y \subset \mathscr{O}_{Y,y}$. We have to prove that $A(E/S, \omega) \notin \mathfrak{m}_y^2$. Suppose, for the sake of contradiction, that the Hasse invariant has a zero of order ≥ 2 at the closed point of S. Consider $R = \mathscr{O}_{Y,y}/\mathfrak{m}_y^2$ and set $E_R \cong E \otimes_{\mathscr{O}_{Y,y}} R$. We see that $A(E_R/R, \omega) = 0$. But, by definition, R is an artinian local K–algebra with residue field K, so we are in the situation of theorem 4.4 and we can express $E_R = E_0 \otimes_K R$ for a supersingular elliptic curve E_0/K. In fact, we can take $E_0 = E_R \otimes_R K = E \otimes_{\mathscr{O}_{Y,y}} K$.

We get two different morphisms $\operatorname{Spec}(R) \to Y$ defining E_R by pull-back from $\mathbb{E}$: the morphisms induced by the canonical projection $\mathscr{O}_{Y,y} \twoheadrightarrow R$ and by the composition $\mathscr{O}_{Y,y} \twoheadrightarrow R \twoheadrightarrow K \hookrightarrow R$. By the universal property of $\mathbb{E}/Y$, these two morphisms are associated with two level $\Gamma(N)$–structures α_1 and α_2 on E_R. Let $\varphi\colon R \to R$ denote the composition of the canonical projection $R \twoheadrightarrow K$ and the canonical inclusion $K \hookrightarrow R$. This morphism induces a cartesian diagram

$$\begin{array}{ccc} (E_R, \alpha_2) & \xrightarrow{\cong} & (E_R, \alpha_1) \\ \downarrow & \ulcorner & \downarrow \\ \operatorname{Spec}(R) & \xrightarrow{\operatorname{Spec}(\varphi)} & \operatorname{Spec}(R) \end{array}$$

and, as a matter of fact, we can define inductively α_{n+1} to be the pull-back of α_n for each $n \in \mathbb{N}$. The isomorphism $E_R \to E_R$ in the top row of the previous diagram must permute the level $\Gamma(N)$–structures of E_R, so at some point we obtain a repetition. That is to say, $\alpha_n = \alpha_1$ for some $n > 1$. This contradicts the universal property of $\mathbb{E}/Y$ because the corresponding morphisms $\mathrm{Spec}(R) \to Y$ are different (observe that $\varphi \circ \varphi = \varphi$). □

4.2 The structure theorem

This section describes the structure of the algebra of modular forms in positive characteristic in terms of their q–expansions and using the Hasse invariant. The main result is analogous to theorem 1.41 in the classical setting. The exposition in this section is based on section 2.7 of Cais's notes [2].

First of all, we establish some notation for the rest of this chapter. Fix $N \in \mathbb{N}$ such that $N \geq 3$ and $p \nmid N$. Let K be an algebraically closed field of characteristic p. The moduli problem $\Gamma(N)_K$ giving elliptic curves E over K–algebras with level $\Gamma(N)$–structures α_N is represented by

$$\begin{array}{c} (\mathbb{E}_K, \alpha_{K,\mathrm{univ}}) \\ \big\downarrow \pi \\ Y(N)_K \end{array}$$

where the modular curve $Y(N)_K$ is a smooth affine curve over K (see definition 2.19). Adding the cusps, we obtain the modular curve $X(N)_K$ (see definition 2.21), which is a proper smooth curve over K (see theorem 2.23). To simplify the notation, from here on we write $Y = Y(N)_K$, $X = X(N)_K$ and $\mathbb{E} = \mathbb{E}_K$.

For each primitive N–th root of unity $\zeta_N \in K$, the modular curve Y (resp. X) has a connected component Y_{ζ_N} (resp. X_{ζ_N}) formed of the points corresponding to elliptic curves with level $\Gamma(N)$–structures of determinant ζ_N (see theorem 2.24). Observe that an element $f \in \Gamma(Y_{\zeta_N}, \underline{\omega}_{\mathbb{E}/Y}^{\otimes k})$ (resp. $f \in \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes k})$) for some $k \in \mathbb{Z}$, which is the restriction of a modular form for $\Gamma(N)$ of weight k to one connected component of the modular curve, is uniquely determined by any one of its q–expansions at that component and its weight k by corollaries 2.30 and 2.31.

Consider the Hasse invariant $A \in M(K; \Gamma(N), p-1)$. Recall that all of its q–expansions are 1. Thus, if we multiply any element of $F(K; \Gamma(N))$ by A, its

q–expansions remain unchanged. The main theorem of this section shows that this is essentially the only relation between the q–expansions of the elements of the algebra of modular forms for $\Gamma(N)$ with coefficients in K.

The proof of the following result is based on the formulae explained in section A1.5 of Katz's article [8].

Proposition 4.6. *The line bundle $\underline{\omega}$ on the curve X is ample.*

Proof. Each cusp corresponds to a triple $(\mathrm{Tate}(q)/K((q^{1/N})), \omega_{\mathrm{can}}, \alpha_N)$, where the level $\Gamma(N)$–structure α_N is defined up to automorphism of $\mathrm{Tate}(q)/K((q^{1/N}))$ (see theorem 2.26). Also, the formal completion of X along the cusp corresponds to the formal group of $\mathrm{Tate}(q)$. Thus, since the Kodaira–Spencer isomorphism for the Tate curve maps $\omega_{\mathrm{can}}^{\otimes 2}$ to the logarithmic differential $\frac{dq}{q}$ (by corollary 3.24), the Kodaira–Spencer isomorphism $\mathrm{KS}\colon \underline{\omega}_{\mathbb{E}/Y}^{\otimes 2} \to \Omega^1_{Y/K}$ extends to an isomorphism $\underline{\omega}^{\otimes 2} \to \Omega^1_{X/K}(\log(C))$ over X, where C is the closed subscheme of cusps of X.

Next, we study each connected component of X separately. That is, take a primitive N–th root of unity $\zeta_N \in K$, so that X_{ζ_N} is a proper smooth connected curve over the algebraically closed field K. We have that

$$\deg\left(\underline{\omega}^{\otimes 2}\big|_{X_{\zeta_N}}\right) = \deg(\Omega^1_{X_{\zeta_N}/K}(\log(C_{\zeta_N}))) = 2g - 2 + |C_{\zeta_N}|\,,$$

where g is the genus of X_{ζ_N} and $|C_{\zeta_N}|$ denotes the number of cusps in the connected component X_{ζ_N}. In order to prove that $\underline{\omega}\big|_{X_{\zeta_N}}$ is ample, it suffices to show that the previous degree is positive.

Consider the morphism $j\colon X_{\zeta_N} \to \mathbb{P}^1_K$ given on points by the j–invariant. A closed point of $\mathbb{P}^1_K$ other than ∞ corresponds to an elliptic curve E/K defined up to isomorphism and the closed points of X_{ζ_N} lying over it correspond to the level $\Gamma(N)$–structures of determinant ζ_N on E modulo the automorphisms of E/K. Thus, the fibre over this point of $\mathbb{P}^1_K$ has $|\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|/|\mathrm{Aut}(E/K)|$ points. In particular, if $j(E) \notin \{0, 1728, \infty\}$, then $\mathrm{Aut}(E/K) = \{[1], [-1]\}$. We deduce that j is a covering of degree $\frac{|\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|}{2}$. On the other hand, the cusps correspond to $\mathrm{Tate}(q)/K((q^{1/N}))$ and the formal completion of X along each cusp is isomorphic to $K[\![q^{1/N}]\!]$, while the formal completion of $\mathbb{P}^1_K$ along ∞ is isomorphic to $K[\![q]\!]$ (via the map $j^{-1} \mapsto q(1 - 744q + \cdots)$). Hence, each cusp has ramification index N over ∞. We conclude that

$$|C_{\zeta_N}| = \frac{|\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|}{2N}$$

and we observe that this quantity is greater than 2 (for $N \geq 3$). Since $g \geq 0$, this concludes the proof. □

Lemma 4.7. *Let T be a proper normal connected locally noetherian scheme over a base ring R and let $\mathscr{L}$ be an ample line bundle on T. Consider the graded ring*

$$B = \bigoplus_{n \geq 0} \mathrm{H}^0(T, \mathscr{L}^{\otimes n})$$

and choose $f \in \mathrm{H}^0(T, \mathscr{L}^{\otimes k})$ for some $k \in \mathbb{N}$ such that $k \in R^\times$. If f has at least one simple zero, then $f - 1$ generates a prime ideal in B.

Proof. Since $\mathscr{L}$ is ample, we have an isomorphism

$$T \cong \mathrm{Proj}\Big(\bigoplus_{n \geq 0} \mathscr{L}^{\otimes n}\Big)$$

and the open immersion $T_f \hookrightarrow T$ is an affine morphism (here, T_f is the open subscheme of points at which f is invertible). Also, T is integral: it is irreducible because it is connected and locally noetherian and it is reduced because it is normal.

The natural morphism of graded R–algebras

$$B_f \to \widetilde{B} = \bigoplus_{n \in \mathbb{Z}} \Gamma(T_f, \mathscr{L}^{\otimes n})$$

is an isomorphism. It is clear that the closed subscheme $V(f-1)$ of $\mathrm{Spec}(B)$ lies in $\mathrm{Spec}(B_f)$, so we consider the closed subscheme Z of $\mathrm{Spec}(\widetilde{B})$ corresponding to $V(f-1)$ through the previous isomorphism.

Take an affine open subset $U = \mathrm{Spec}(C)$ of T such that $\mathscr{L}\big|_U$ is trivial. Since $T_f \hookrightarrow T$ is affine, we see that $U \cap T_f = U_f = \mathrm{Spec}(\widetilde{C})$. Choose a generator t of $\Gamma(U, \mathscr{L})$, so that $\mathscr{L}\big|_U = t\mathscr{O}_U$. In particular, $\Gamma(U_f, \mathscr{L}) = t\widetilde{C}$ and we identify

$$\bigoplus_{n \in \mathbb{Z}} \Gamma(U_f, \mathscr{L}^{\otimes n}) = \widetilde{C}[t, t^{-1}]\,.$$

With this identification, the restriction of f to U_f is of the form ut^k for some $u \in \widetilde{C}^\times$. Hence, $f - 1$ is given by $ut^k - 1$. As $k \in R^\times$, we deduce that Z is étale over U_f. Using this argument on an affine open covering, we get that Z is étale over T_f and so over T too. Therefore, Z is normal (because T is).

Furthermore, we claim that Z is connected. As T_f is irreducible, any two open subschemes of T_f have non-trivial intersection. Therefore, it suffices to prove it over $U_f = \mathrm{Spec}(\widetilde{C})$ irreducible. That is to say, we have to prove that $\mathrm{Spec}(\widetilde{C}[t,t^{-1}]/(ut^k-1)) \cong \mathrm{Spec}(\widetilde{C}[z]/(z^k-u))$ (where $z=t^{-1}$) is connected. As $\mathrm{Spec}(\widetilde{C}[z]/(z^k-u))$ is étale over $\mathrm{Spec}(\widetilde{C})$, each of its connected components has at least one point in the generic fibre. Hence, we only need to prove that the generic fibre is connected, so we consider the fraction field F of the normal domain $\widetilde{C}$. Next, we want to show that z^k-u is an irreducible polynomial over F (which implies the claim).

Consider the closed subscheme $V(f)$ of T (i.e., the zeros of f). By Krull's principal ideal theorem, each irreducible component of $V(f)$ has codimension 1 in T. We deduce that the local rings at the maximal points of $V(f)$ are discrete valuation rings because T is normal. By hypothesis, f has a simple zero. We choose a maximal point x of $V(f)$ where f has a simple zero, so that the image of f in $\mathscr{O}_{T,x}$ is a uniformizer. But we can also regard F as the fraction field of $\mathscr{O}_{T,x}$ and we deduce that the element $u \in F$ must be a uniformizer of $\mathscr{O}_{T,x}$. Eisenstein's criterion shows that the polynomial z^k-u is irreducible over F, as desired.

In conclusion, the closed subscheme $V(f-1)$ of $\mathrm{Spec}(B)$ is normal and connected and so the ideal $(f-1)B$ is prime. □

Theorem 4.8. *Let μ be the set of primitive N–th roots of unity in K. For each $\zeta_N \in \mu$, choose a level $\Gamma(N)$–structure $\alpha(\zeta_N)$ of determinant ζ_N on $\mathrm{Tate}(q)/K((q^{1/N}))$. The kernel of the K–algebra homomorphism*

$$\begin{aligned} M(K;\Gamma(N)) &\to \prod_{\zeta_N \in \mu} K[\![q^{1/N}]\!] \\ f &\mapsto (\widehat{f}_{\alpha(\zeta_N)}(q))_{\zeta_N \in \mu} \end{aligned}$$

(taking one q–expansion of each determinant) is the ideal generated by $A-1$, where $A \in M(K;\Gamma(N),p-1)$ is the Hasse invariant (cf. theorem 1.41).

Proof. We prove a stronger result, namely that for each $\zeta_N \in \mu$ the kernel of the morphism

$$\Phi\colon B = \bigoplus_{n \in \mathbb{Z}} \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes n}) \to K[\![q^{1/N}]\!]$$

given by evaluation at $(\mathrm{Tate}(q)/K((q^{1/N})), \omega_{\mathrm{can}}, \alpha(\zeta_N))$ is the ideal generated by $A-1$ (in fact, by its restriction to X_{ζ_N}, but we omit it from the notation). That is, we prove the theorem separately on each connected component of X.

First, observe that $\Gamma(X_{\zeta_N}, \mathscr{O}_X) = K$ because K is an algebraically closed field and X_{ζ_N} is connected and reduced. Next, choose $f \in \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes k})$ for some $k < 0$. We can consider the modular discriminant $\Delta \in \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes 12})$ (as in the proof of corollary 2.31). We see that $f^{12}\Delta^{-k} \in \Gamma(X_{\zeta_N}, \mathscr{O}_X)$, which means that $f^{12}\Delta^{-k}$ is a constant. But it is also (the restriction of) a cusp form, so $f^{12}\Delta^{-k} = 0$, which is only possible if $f = 0$. In conclusion, there are no non-zero modular forms of negative weight which are holomorphic at ∞.

All in all, we can express

$$B = \bigoplus_{n \in \mathbb{Z}} \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes n}) = \bigoplus_{n \geq 0} \Gamma(X_{\zeta_N}, \underline{\omega}^{\otimes n})$$

and so lemma 4.7 implies that $A - 1$ generates a prime ideal $\mathfrak{a} = (A - 1)$ in B. It is clear that $\mathfrak{a} \subseteq \mathrm{Ker}(\Phi)$, as the q–expansions of the Hasse invariant are all equal to 1, so it remains to prove that the inclusion is not proper. But X_{ζ_N} is a proper smooth curve over K, so $X_{\zeta_N} \cong \mathrm{Proj}(B)$ and this implies that B has Krull dimension 2. Since $\Phi(\Delta) = q(1 + \cdots)$, we see that the image of Φ has dimension ≥ 1 and so $\mathrm{Ker}(\Phi)$ cannot be a maximal ideal of B. Therefore, $\mathfrak{a} = \mathrm{Ker}(\Phi)$. □

Corollary 4.9. *Theorem 4.8 holds also for modular forms which are not holomorphic at ∞ (that is, replacing $M(K;\Gamma(N))$ with $F(K;\Gamma(N))$ and $K[\![q^{1/N}]\!]$ with $K((q^{1/N}))$).*

Proof. Let $f \in F(K;\Gamma(N))$ such that $\widehat{f}_{\alpha(\zeta_N)}(q) = 0$ for every $\zeta_N \in \mu$. As in the proof of corollary 2.31, we can choose $r \gg 0$ such that $f\Delta^r$ is holomorphic at ∞ (because the modular discriminant Δ is a cusp form) and so theorem 4.8 implies that $A - 1$ divides $f\Delta^r$ in $M(K;\Gamma(N)) \subset F(K;\Gamma(N))$. But Δ is invertible in $F(K;\Gamma(N))$, which means that $A - 1$ divides f in $F(K;\Gamma(N))$. □

We are basically in the same situation as in the classical case: we can use A to define a filtration on the space of modular forms of weight in a congruence class modulo $p - 1$ (cf. definition 1.43).

Definition 4.10. Let $k \in \mathbb{Z}$. We say that $f \in F(K;\Gamma(N), k)$ is of *exact filtration* k if f is not divisible by A in $F(K;\Gamma(N))$ or, equivalently, if there exists no $g \in F(K;\Gamma(N), k')$ for $k' < k$ having the same q–expansions as f.

Proposition 4.11. *If f is a non-zero element of $M(K;\Gamma(N), k)$ for some $k < p - 1$ (resp. for $k = p - 1$ which vanishes at one cusp in each connected component of X), then f has exact filtration k.*

Proof. Suppose that $f = Ag$ for some $g \in F(K;\Gamma(N), k-(p-1))$. In this case, the q–expansions of g must coincide with the q–expansions of f. But there are no non-zero modular forms holomorphic at ∞ of weight < 0 and those of weight 0 are constant on each connected component of X, as we saw in the proof of theorem 4.8. Thus, there cannot be one such g. □

4.3 The operator $A\theta$

The exposition in this section follows closely Katz's article [9]. We keep the notation from the previous section except that we set $\underline{\omega} = \underline{\omega}_{\mathbb{E}/Y}$ to further simplify it. (We work exclusively on the modular curve Y without the cusps, so there is no possible confusion with the extended sheaf $\underline{\omega}$ on X.)

The main result we prove in this section is the following (cf. proposition 1.44):

Theorem 4.12. *There exists a derivation*

$$A\theta \colon F(K;\Gamma(N), \bullet) \to F(K;\Gamma(N), \bullet + p + 1)$$

whose effect on q–expansions is $q\dfrac{d}{dq}$ *and such that*

(1) *if* $f \in F(K;\Gamma(N),k)$ *has exact filtration* k *and* $p \nmid k$*, then* $A\theta(f)$ *has exact filtration* $k+p+1$ *(in particular,* $A\theta(f) \neq 0$*), and*
(2) *if* $f \in F(K;\Gamma(N), pk)$ *and* $A\theta(f) = 0$*, then there is a unique* $g \in F(K;\Gamma(N),k)$ *such that* $f = g^p$*.*

Before moving to the proof of the theorem, we state some consequences.

Corollary 4.13.

(1) *$A\theta$ maps modular forms which are holomorphic at ∞ to cusp forms.*
(2) *The restriction of $A\theta$ to $M(K;\Gamma(N),k)$ is injective if $1 \leq k < p-1$.*
(3) *If $f \in F(K;\Gamma(N),k)$ satisfies that $A\theta(f) = 0$, then we can write uniquely $f = A^r g^p$ with $0 \leq r \leq p-1$, $r + k \equiv 0 \bmod p$, and $g \in F(K;\Gamma(N),l)$, $pl + r(p-1) = k$. If, in addition, f is holomorphic at ∞ (resp. a cusp form), so is g.*

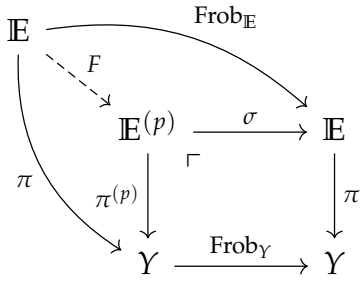
Proof.

(1) $A\theta$ acts as $q\frac{d}{dq}$ on q–expansions.
(2) Let $f \in M(K;\Gamma(N),k) \setminus \{0\}$. By proposition 4.11, f has exact filtration k. Now the theorem says that $A\theta(f) \neq 0$.

(3) Take $r \in \{0, 1, \ldots, p-1\}$ such that $r \equiv -k \bmod p$. We argue by induction on r. The case $r = 0$ is given by the theorem. Now suppose that $r > 0$ and that the result holds for $r - 1$. Since $p \nmid k$ and $A\theta(f) = 0$, f cannot have exact filtration k; that is, $f = Af'$ for a unique $f' \in F(K; \Gamma(N), k - p + 1)$. But $-(k - p + 1) \equiv r - 1 \bmod p$ and $A\theta(f') = 0$ by the q–expansion principle (because it has the same q–expansions as $A\theta(f) = 0$). Therefore, we can express $f' = A^{r-1}g^p$ for a unique $g \in F(K; \Gamma(N), l)$. □

The operator $A\theta$ is roughly the composition of an operator θ and multiplication by the Hasse invariant A. However, in the construction of θ, we get an A in the denominator, so θ is only defined over the locus where A is invertible. Next, we investigate this locus.

Consider the commutative diagram



where Frob_Y and $\mathrm{Frob}_{\mathbb{E}}$ are the absolute Frobenius endomorphisms and F is the relative Frobenius morphism. We are in the situation of section 3.1. The relative Frobenius morphism induces a morphism $F^*\colon \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}^{(p)}/Y) \to \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)$ on de Rham cohomology. Let $\mathscr{F}$ be the image of F^*.

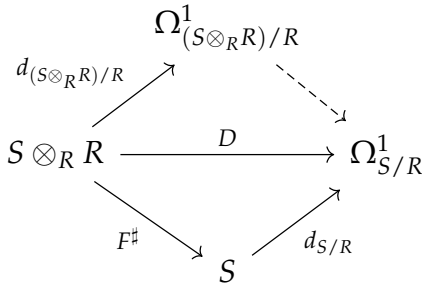
Lemma 4.14. *The image $\mathscr{F}$ and the cokernel $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)/\mathscr{F}$ of F^* are two locally free $\mathscr{O}_Y$–modules of rank* 1.

Proof. The Hodge–de Rham spectral sequences for $\mathbb{E}^{(p)}/Y$ and $\mathbb{E}/Y$ degenerate at the first page by theorem 3.15 and, by functoriality, yield a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathrm{R}^0\pi^{(p)}_*\left(\Omega^1_{\mathbb{E}^{(p)}/Y}\right) & \longrightarrow & \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}^{(p)}/Y) & \longrightarrow & \mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right) & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle F^*} & & \downarrow{\scriptstyle F^*} & & \downarrow{\scriptstyle F^*} & & \\
0 & \longrightarrow & \mathrm{R}^0\pi_*\left(\Omega^1_{\mathbb{E}/Y}\right) & \longrightarrow & \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) & \longrightarrow & \mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right) & \longrightarrow & 0
\end{array}$$

with exact rows.

We claim that the leftmost vertical arrow is 0. By functoriality, this map is the push-forward by $\pi^{(p)}$ of the canonical map $\Omega^1_{\mathbb{E}^{(p)}/Y} \to F_*(\Omega^1_{\mathbb{E}/Y})$. We work locally to prove that this last map vanishes. Restrict $\pi\colon \mathbb{E} \to Y$ to $\pi\colon \operatorname{Spec}(S) \to \operatorname{Spec}(R)$ for some affine open subsets $\operatorname{Spec}(S)$ and $\operatorname{Spec}(R)$ of $\mathbb{E}$ and Y, respectively. In what follows, the tensor product $\cdot \otimes_R R$ is taken regarding R as an R–module via $\operatorname{Frob}_Y^\sharp$. Then, the map $\Omega^1_{\mathbb{E}^{(p)}/Y} \to F_*(\Omega^1_{\mathbb{E}/Y})$ corresponds to the $(S \otimes_R R)$–linear map $\Omega^1_{(S\otimes_R R)/R} \to \Omega^1_{S/R}$ induced by the derivation D making



a commutative diagram (here, $d_{(S\otimes_R R)/R}$ and $d_{S/R}$ are the universal derivations). Then,

$$D(b \otimes a) = d_{S/R}(F^\sharp(b \otimes a)) = d_{S/R}(ab^p) = pab^{p-1}d_{S/R}(b) = 0$$

for all $b \in S$ and all $a \in R$, so $D = 0$ and the claim follows.

The Hodge–de Rham spectral sequence for $\mathbb{E}^{(p)}/Y$ (which degenerates at the first page by theorem 3.15) and the conjugate spectral sequence for $\mathbb{E}/Y$ (which degenerates at the second page by theorem 3.14) yield two short exact sequences of $\mathscr{O}_Y$–modules

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathrm{R}^0\pi^{(p)}_*\left(\Omega^1_{\mathbb{E}^{(p)}/Y}\right) & \longrightarrow & \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}^{(p)}/Y) & \longrightarrow & \mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right) & \longrightarrow & 0 \\
 & & & \searrow^{0} & \downarrow{\scriptstyle F^*} & \swarrow & & & \\
0 & \longrightarrow & \mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right) & \longrightarrow & \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) & \longrightarrow & \mathrm{R}^0\pi^{(p)}_*\left(\Omega^1_{\mathbb{E}^{(p)}/Y}\right) & \longrightarrow & 0
\end{array}$$

and we see that F^* factors through $\mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right)$. In lemma 1 of his article [9], Katz claims without proof that the induced map coincides with the inclusion $\mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right) \to \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)$ in the bottom row. (Unfortunately, the author of this work has been unable to understand why this is true.)

Therefore, $\mathscr{F} \cong \mathrm{R}^1\pi^{(p)}_*\left(\mathscr{O}_{\mathbb{E}^{(p)}}\right)$ and $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)/\mathscr{F} \cong \mathrm{R}^0\pi^{(p)}_*\left(\Omega^1_{\mathbb{E}^{(p)}/Y}\right)$ and both are locally free $\mathscr{O}_Y$–modules of rank 1. □

Lemma 4.15. *The open subset Y^{H} of Y where A is invertible is the largest open subset over which the map $\underline{\omega} \oplus \mathscr{F} \to \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)$ induced by the inclusions is an isomorphism.*

Proof. We work locally on Y. We choose an open subset U of Y with a basis (ω, η) of $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_U$ such that ω is the image of a basis of $\mathrm{R}^0\pi_*\left(\Omega^1_{\mathbb{E}/Y}\right)\big|_U$ whose Serre–Grothendieck dual is the projection $\widetilde{\eta}$ of η on $\mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right)\big|_U$. (In particular, (ω, η) is a basis adapted to the Hodge filtration.) The basis $(\omega^{(p)}, \eta^{(p)})$ of $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}^{(p)}/Y)\big|_U$ obtained after base change by Frob_Y has exactly the same properties.

The proof of lemma 4.14 shows that $F^*(\omega^{(p)}) = 0$. Moreover, $F^* \circ \sigma^* = \mathrm{Frob}^*_{\mathbb{E}}$ and so the projection of $F^*(\eta^{(p)})$ on $\mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right)\big|_U$ is precisely $A(\pi^{-1}(U)/U, \omega)\widetilde{\eta}$, by the definition of the Hasse invariant. That is, F^* can be described in matrix form as

$$\begin{pmatrix} F^*(\omega^{(p)}) & F^*(\eta^{(p)}) \end{pmatrix} = \begin{pmatrix} \omega & \eta \end{pmatrix} \begin{pmatrix} 0 & B_1 \\ 0 & A_1 \end{pmatrix},$$

where we write A_1 for the value $A(\pi^{-1}(U)/U, \omega)$ of the Hasse invariant. Therefore, $\mathscr{F}\big|_U$ is generated by $B_1\omega + A_1\eta$. In conclusion, $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_U$ is isomorphic to $\underline{\omega}\big|_U \oplus \mathscr{F}\big|_U$ if and only if the sections ω and $B_1\omega + A_1\eta$ generate $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_U$ or, equivalently, A_1 is invertible. □

Remark. The proof shows that $\mathscr{F}\big|_U$ is generated by $B_1\omega + A_1\eta$. Since U is chosen so that $\mathscr{F}\big|_U$ is isomorphic to $\mathscr{O}_U$, we see that A_1 and B_1 have no common zeros (cf. lemma 1.40).

We can now define a derivation θ of $F(K; \Gamma(N))\left[\frac{1}{A}\right]$ (i.e., over the locus where the Hasse invariant A is invertible). By lemma 4.15, we have an isomorphism $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_{Y^{\mathrm{H}}} \cong \underline{\omega}\big|_{Y^{\mathrm{H}}} \oplus \mathscr{F}\big|_{Y^{\mathrm{H}}}$. Since both $\underline{\omega}$ and $\mathscr{F}$ are invertible $\mathscr{O}_Y$–modules, for each $k \in \mathbb{N}$ we obtain a decomposition

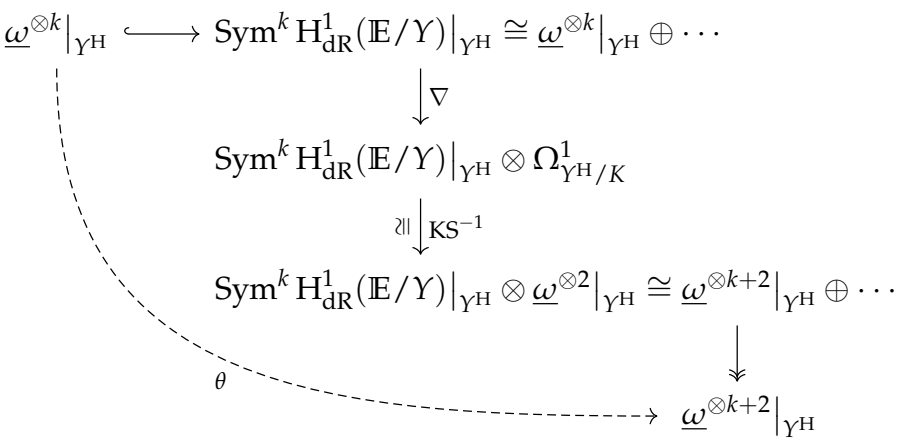
$$\mathrm{Sym}^k \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_{Y^{\mathrm{H}}} \cong \underline{\omega}^{\otimes k}\big|_{Y^{\mathrm{H}}} \oplus \left(\underline{\omega}^{\otimes k-1} \otimes_{\mathscr{O}_Y} \mathscr{F}\right)\big|_{Y^{\mathrm{H}}} \oplus \cdots \oplus \mathscr{F}^{\otimes k}\big|_{Y^{\mathrm{H}}}.$$

The Gauss–Manin connection $\nabla\colon \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) \to \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) \otimes_{\mathscr{O}_Y} \Omega^1_{Y/K}$ induces for each $k \in \mathbb{N}$ a connection

$$\nabla\colon \mathrm{Sym}^k \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) \to \mathrm{Sym}^k \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) \otimes_{\mathscr{O}_Y} \Omega^1_{Y/K}$$

and we have the Kodaira–Spencer isomorphism $\mathrm{KS}\colon \underline{\omega}^{\otimes 2} \to \Omega^1_{Y/K}$.

Definition 4.16. Let $k \in \mathbb{N}$. The map $\theta\colon \underline{\omega}^{\otimes k}\big|_{Y^{\mathrm{H}}} \to \underline{\omega}^{\otimes k+2}\big|_{Y^{\mathrm{H}}}$ is defined to be the composition



and we write again θ for the induced map $\mathrm{H}^0(Y^{\mathrm{H}}, \underline{\omega}^{\otimes k}) \to \mathrm{H}^0(Y^{\mathrm{H}}, \underline{\omega}^{\otimes k+2})$ (cf. corollary 1.37).

Lemma 4.17. *The effect of θ on q–expansions is $q\dfrac{d}{dq}$.*

Proof. Consider $(\mathrm{Tate}(q)/K((q^{1/N})), \omega_{\mathrm{can}}, \alpha_N)$ for some level $\Gamma(N)$–structure α_N. By corollary 3.24, $\mathrm{KS}(\omega_{\mathrm{can}}^{\otimes 2}) = \frac{dq}{q}$, the dual derivation to which is $q\frac{d}{dq}$. Set $\eta = \nabla(q\frac{d}{dq})(\omega_{\mathrm{can}})$. In section A2.2 of his article [8], Katz computes the effect of the Frobenius endomorphism on the de Rham cohomology of the Tate curve using complex analytic tools (the same kind of techniques as in section 3.4) and shows that

$$\begin{pmatrix}\mathrm{Frob}^*_{\mathrm{Tate}(q)}(\omega_{\mathrm{can}}) & \mathrm{Frob}^*_{\mathrm{Tate}(q)}(\eta)\end{pmatrix} = \begin{pmatrix}\omega_{\mathrm{can}} & \eta\end{pmatrix}\begin{pmatrix}p & 0\\ 0 & 1\end{pmatrix}.$$

From this expression, since $p = 0$ in K, we deduce that the image of the Frobenius morphism in $\mathrm{H}^1_{\mathrm{dR}}(\mathrm{Tate}(q)/K((q^{1/N})))$ is spanned by η.

Given $f \in \mathrm{H}^0(Y^{\mathrm{H}}, \underline{\omega}^{\otimes k})$, consider $\widehat{f}_{\alpha_N}(q) = f(\mathrm{Tate}(q)/K((q^{1/N})), \omega_{\mathrm{can}}, \alpha_N)$ and the local expression $\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k}$ obtained by pull-back from f (observe that $\mathrm{Tate}(q)/K((q^{1/N}))$ can be obtained by pull-back from $\pi^{-1}(Y^{\mathrm{H}})/Y^{\mathrm{H}}$ because the q–expansions of A are equal to 1). We compute the action of θ on this local expression via the chain of maps appearing in definition 4.16, all of which commute with base change. By duality, we can express

$$\nabla(\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k}) = \nabla\Big(q\frac{d}{dq}\Big)(\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k}) \otimes \frac{dq}{q}$$

and, after applying the Kodaira–Spencer isomorphism, we obtain

$$\begin{aligned}\nabla\Big(q\frac{d}{dq}\Big)(\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k})\otimes\omega_{\mathrm{can}}^{\otimes 2} &= q\frac{d}{dq}(\widehat{f}_{\alpha_N}(q))\omega_{\mathrm{can}}^{\otimes k+2}+k\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k+1}\otimes\eta\\ &\equiv q\frac{d}{dq}(\widehat{f}_{\alpha_N}(q))\omega_{\mathrm{can}}^{\otimes k+2} \mod \eta\,.\end{aligned}$$

Therefore,

$$\theta(\widehat{f}_{\alpha_N}(q)\omega_{\mathrm{can}}^{\otimes k}) = q\frac{d}{dq}(\widehat{f}_{\alpha_N}(q))\omega_{\mathrm{can}}^{\otimes k+2}\,,$$

whence

$$(\widehat{\theta f})_{\alpha_N}(q) = q\frac{d}{dq}(\widehat{f}_{\alpha_N}(q))\,,$$

as required. □

Lemma 4.18. *For each $k\in\mathbb{N}$, there is a unique map*

$$A\theta\colon F(K;\Gamma(N),k)\to F(K;\Gamma(N),k+p+1)$$

making the diagram

$$\begin{array}{ccccc}\mathrm{H}^0(Y^{\mathrm{H}},\underline{\omega}^{\otimes k}) & \xrightarrow{\theta} & \mathrm{H}^0(Y^{\mathrm{H}},\underline{\omega}^{\otimes k+2}) & \xrightarrow{\cdot A} & \mathrm{H}^0(Y^{\mathrm{H}},\underline{\omega}^{\otimes k+p+1})\\ \uparrow & & & & \uparrow\\ \mathrm{H}^0(Y,\underline{\omega}^{\otimes k}) & & \xrightarrow{\qquad\qquad A\theta\qquad\qquad} & & \mathrm{H}^0(Y,\underline{\omega}^{\otimes k+p+1})\end{array}$$

commute.

Proof. First, observe that the vertical arrows in the previous diagram are injective by the q–expansion principle, as all q–expansions are defined over Y^{H}.

We work locally on Y. We can choose an open subset U of Y such that $\mathrm{R}^0\pi_*(\Omega^1_{\mathbb{E}/Y})\big|_U = \underline{\omega}\big|_U$, $\mathrm{R}^1\pi_*(\mathscr{O}_{\mathbb{E}})\big|_U \cong \underline{\omega}^{\otimes -1}\big|_U$ and $\mathscr{F}\big|_U$ are isomorphic to $\mathscr{O}_U$. Let ω be a basis of $\underline{\omega}\big|_U$. Set $\xi = \mathrm{KS}(\omega^{\otimes 2})$, which is a basis of $\Omega^1_{U/K}$, and consider the basis D of $\mathscr{D}\mathit{er}_K(\mathscr{O}_U,\mathscr{O}_U)$ dual to ξ. Define $\eta = \nabla(D)(\omega)$. Observe that the image of ω under the composition of maps

$$\begin{array}{ccccccc}\underline{\omega}\big|_U \hookrightarrow \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_U & \xrightarrow{\nabla} & (\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\otimes\Omega^1_{Y/K})\big|_U & \twoheadrightarrow & (\underline{\omega}^{\otimes -1}\otimes\Omega^1_{Y/K})\big|_U & \xrightarrow{D} & \underline{\omega}^{\otimes -1}\big|_U\\ \omega & \longmapsto & \omega^{\otimes -1}\otimes\xi & \longmapsto & \omega^{\otimes -1} & & \end{array}$$

with κ is the Kodaira–Spencer isomorphism.

is its Serre–Grothendieck dual. (The unlabelled arrows above come from the short exact sequence

$$0 \to \underline{\omega} = \mathrm{R}^0\pi_*\left(\Omega^1_{\mathbb{E}/Y}\right) \to \mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y) \to \mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right) \cong \underline{\omega}^{\otimes -1} \to 0$$

induced by the Hodge filtration.) Hence, the projection of η on $\mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right)\big|_U$ is the Serre–Grothendieck dual of ω (i.e., (ω, η) is a basis of $\mathrm{H}^1_{\mathrm{dR}}(\mathbb{E}/Y)\big|_U$ adapted to the Hodge filtration). We are in the situation of the proof of lemma 4.15, so

$$\begin{pmatrix} F^*(\omega^{(p)}) & F^*(\eta^{(p)}) \end{pmatrix} = \begin{pmatrix} \omega & \eta \end{pmatrix} \begin{pmatrix} 0 & B_1 \\ 0 & A_1 \end{pmatrix}.$$

Now set $V = U \cap Y^{\mathrm{H}}$ and write again ω, η, A_1 and B_1 for the restrictions of ω, η, A_1 and B_1 to V. Since A_1 is invertible over V, we have a basis

$$\varphi = \frac{B_1}{A_1}\omega + \eta$$

of $\mathscr{F}\big|_V$ whose projection on $\mathrm{R}^1\pi_*\left(\mathscr{O}_{\mathbb{E}}\right)\big|_V$ is the Serre–Grothendieck dual of ω. Let $f \in \mathrm{H}^0(Y, \underline{\omega}^{\otimes k})$ and express it locally on U as $f_1\omega^{\otimes k}$ for some $f_1 \in \mathscr{O}_Y(U)$. Abusing notation we write again $f_1\omega^{\otimes k}$ for its restriction to V. Next, we compute $\theta(f_1\omega^{\otimes k})$ via the chain of maps appearing in definition 4.16. By duality, we can write

$$\nabla(f_1\omega^{\otimes k}) = \nabla(D)(f_1\omega^{\otimes k}) \otimes \xi\,,$$

which corresponds by the Kodaira–Spencer isomorphism to

$$\begin{aligned} \nabla(D)(f_1\omega^{\otimes k}) \otimes \omega^{\otimes 2} &= D(f_1)\omega^{\otimes k+2} + kf_1\omega^{\otimes k+1} \otimes \eta \\ &= D(f_1)\omega^{\otimes k+2} + kf_1\omega^{\otimes k+1} \otimes \varphi - kf_1\frac{B_1}{A_1}\omega^{\otimes k+2} \\ &\equiv \Big(D(f_1) - kf_1\frac{B_1}{A_1}\Big)\omega^{\otimes k+2} \mod \varphi\,. \end{aligned}$$

Therefore,

$$\theta(f_1\omega^{\otimes k}) = \Big(D(f_1) - kf_1\frac{B_1}{A_1}\Big)\omega^{\otimes k+2}$$

and, multiplying by the Hasse invariant $A_1\omega^{\otimes p-1}$, we see that the formula

$$A\theta(f_1\omega^{\otimes k}) = (D(f_1)A_1 - kf_1B_1)\omega^{\otimes k+p+1}\,.$$

makes sense over U. □

Corollary 4.19. *The local formula obtained at the end of the proof of lemma 4.18 defines a map* $A\theta\colon F(K;\Gamma(N),k) \to F(K;\Gamma(N),k+p+1)$ *for every* $k \in \mathbb{Z}$.

Proof. Consider $f \in F(K;\Gamma(N),k)$ for some $k \in \mathbb{Z}$. We use that the modular discriminant $\Delta \in F(K;\Gamma(N),12)$ is invertible in $F(K;\Gamma(N))$. Choose an integer $r \gg 0$ such that $k+12pr > 0$ and define

$$A\theta(f) = \frac{A\theta(f\Delta^{pr})}{\Delta^{pr}}$$

(here, $A\theta(f\Delta^{pr})$ is defined as in lemma 4.18, which makes sense because $f\Delta^{pr}$ has positive weight).

Using the same notation as in the proof of lemma 4.18, we can compute locally on Y

$$\begin{aligned} A\theta(f\Delta^{pr}) &= (D(f_1\Delta_1^{pr})A_1 - (k+12pr)f_1\Delta_1^{pr}B_1)\omega^{\otimes k+12pr+p+1} \\ &= (D(f_1)\Delta_1^{pr}A_1 + prf_1\Delta_1^{pr-1}D(\Delta_1) - kf_1\Delta_1^{pr}B_1)\omega^{\otimes k+12pr+p+1} \\ &= (D(f_1)\Delta_1^{pr}A_1 - kf_1\Delta_1^{pr}B_1)\omega^{\otimes k+12pr+p+1} \end{aligned}$$

(where we used that $p = 0$ in K), whence

$$A\theta(f) = \frac{A\theta(f\Delta^{pr})}{\Delta^{pr}} = (D(f_1)A_1 - kf_1B_1)\omega^{\otimes k+p+1}\,.$$

This is the same formula as for modular forms of positive weight. □

Proof of theorem 4.12. We have constructed a derivation $A\theta$ on $F(K;\Gamma(N))$ acting on q–expansions as $q\frac{d}{dq}$, as desired. It remains to show its properties with respect to the filtration given by A.

(1) Let $f \in F(K;\Gamma(N),k)$ of exact filtration k. Since A does not divide f, there is a point y of Y at which A has a zero of larger order than f. As in the proof of lemma 4.18, we express f locally in an open neighbourhood of y as $f_1\omega^{\otimes k}$ and obtain that

$$A\theta(f_1\omega^{\otimes k}) = (D(f_1)A_1 - kf_1B_1)\omega^{\otimes k+p+1}\,.$$

Suppose that $p \nmid k$, so that $k \neq 0$ in K. Since A_1 and B_1 have no common zeros (see the remark after lemma 4.15),

$$\mathrm{ord}_y(D(f_1)A_1 - kf_1B_1) = \mathrm{ord}_y(f_1) < \mathrm{ord}_y(A_1)\,,$$

which implies that A does not divide $A\theta(f)$. That is, $A\theta(f)$ has exact filtration $k+p+1$.

(2) Let $f \in F(K;\Gamma(N),pk)$ such that $A\theta(f)=0$. Again working locally on an open subset U of Y and with the same notation, we have that

$$0 = A\theta(f_1\omega^{\otimes pk}) = (D(f_1)A_1 - pkf_1B_1)\omega^{\otimes pk+p+1} = D(f_1)A_1\omega^{\otimes pk+p+1},$$

which is only possible if $D(f_1)=0$. But Y is a smooth curve over the perfect (in fact, algebraically closed) field K of characteristic p and D is a basis of $\mathscr{D}er_K(\mathscr{O}_U,\mathscr{O}_U)$. Therefore, $f_1 = g_1^p$ for a necessarily unique $g_1 \in \mathscr{O}_Y(U)$. We obtain that $f_1\omega^{\otimes pk} = (g_1\omega^{\otimes k})^p$. By uniqueness, these local sections $g_1\omega^{\otimes k}$ can be glued together to a unique $g \in F(K;\Gamma(N),k)$ satisfying that $g^p = f$. □

Bibliography

[1] Borevich, Z. I. and Shafarevich, I. R. *Number theory*. Pure and applied mathematics 20. New York, NY, USA: Academic press, 1966. 435 pp.

[2] Cais, B. *Serre's conjectures*. Expository notes for a seminar. 2009. 21 pp. URL: http://math.arizona.edu/~cais/Papers/Expos/Serre05.pdf (visited on 08/06/2018).

[3] Deligne, P. 'Théorème de Lefschetz et critères de dégénérescence de suites spectrales'. In: *Publ. Math. IHÉS* 35.1 (1968), pp. 259–278.

[4] Deligne, P. and Illusie, L. 'Relèvements modulo p^2 et décomposition du complexe de de Rham'. In: *Invent. Math.* 89.2 (1987), pp. 247–270.

[5] Deligne, P. and Rapoport, M. 'Les schémas de modules de courbes elliptiques'. In: *Modular functions of one variable II*. Ed. by Deligne, P. and Kuijk, W. Lecture notes in mathematics 349. Berlin, Germany: Springer-Verlag, 1973, pp. 143–316.

[6] Diamond, F. and Taylor, R. 'Non-optimal levels of mod l modular representations'. In: *Invent. Math.* 115.3 (1994), pp. 435–462.

[7] Geer, G. van der and Moonen, B. *Abelian varieties*. Preliminary chapters for a book. 2011. Chap. V.2, pp. 76–83. URL: https://www.math.ru.nl/~bmoonen/BookAV/Isogs.pdf (visited on 03/06/2018).

[8] Katz, N. M. 'p-adic properties of modular schemes and modular forms'. In: *Modular functions of one variable III*. Ed. by Serre, J.-P. and Kuijk, W. Lecture notes in mathematics 350. Berlin, Germany: Springer-Verlag, 1973, pp. 69–190.

[9] Katz, N. M. 'A result on modular forms in characteristic p'. In: *Modular functions of one variable V*. Ed. by Serre, J.-P. and Zagier, D. B. Lecture notes in mathematics 601. Berlin, Germany: Springer-Verlag, 1977, pp. 53–61.

[10] Katz, N. M. 'Algebraic solutions of differential equations (p-curvature and the Hodge filtration)'. In: *Invent. Math.* 18.1-2 (1972), pp. 1–118.

[11] Katz, N. M. 'Nilpotent connections and the monodromy theorem: applications of a result of Turrittin'. In: *Publ. Math. IHÉS* 39.1 (1970), pp. 175–232.

[12] Katz, N. M. and Mazur, B. *Arithmetic moduli of elliptic curves*. Annals of mathematics studies 108. Princeton, NJ, USA: Princeton University Press, 1985. 514 pp.

[13] Katz, N. M. and Oda, T. 'On the differentiation of de Rham cohomology classes with respect to parameters'. In: *J. Math. Kyoto Univ.* 8.2 (1968), pp. 199–213.

[14] Kedlaya, K. S. 'p-adic cohomology: from theory to practice'. In: *p-adic geometry. Lectures from the 2007 Arizona Winter School*. Ed. by Savitt, D. and Thakur, D. S. University lecture series 45. Providence, RI, USA: American Mathematical Society, 2008, pp. 175–203. URL: http://swc.math.arizona.edu/aws/2007/KedlayaNotes11Mar.pdf (visited on 07/06/2018).

[15] Koblitz, N. *Introduction to elliptic curves and modular forms*. 2nd ed. Graduate texts in mathematics 97. New York, NY, USA: Springer-Verlag, 1993. 248 pp.

[16] Serre, J.-P. *A course in arithmetic*. Graduate texts in mathematics 7. New York, NY, USA: Springer-Verlag, 1973. 115 pp.

[17] Serre, J.-P. 'Congruences et formes modulaires'. In: *Séminaire Bourbaki. Vol. 1971/72. Exposés 400–417*. Ed. by Dold, A. and Eckmann, B. Lecture notes in mathematics 317. Berlin, Germany: Springer-Verlag, 1973, pp. 319–338.

[18] Serre, J.-P. 'Formes modulaires et fonctions zêta p-adiques'. In: *Modular functions of one variable III*. Ed. by Serre, J.-P. and Kuijk, W. Lecture notes in mathematics 350. Berlin, Germany: Sprinver-Verlag, 1973, pp. 191–268.

[19] Silverman, J. H. *Advanced topics in the arithmetic of elliptic curves*. Graduate texts in mathematics 151. New York, NY, USA: Springer-Verlag, 1994. 525 pp.

[20] Stein, W. A. *Modular forms, a computational approach*. Graduate studies in mathematics 79. Providence, RI, USA: American Mathematical Society, 2007. 268 pp. URL: http://wstein.org/books/modform/ (visited on 18/04/2018).

[21] Swinnerton-Dyer, H. P. F. 'On ℓ-adic representations and congruences for coefficients of modular forms'. In: *Modular functions of one variable III*. Ed. by Serre, J.-P. and Kuijk, W. Lecture notes in mathematics 350. Berlin, Germany: Springer-Verlag, 1973, pp. 1–55.

[22] Wedhorn, T. 'De Rham cohomology of varieties over fields of positive characteristic'. In: *Higher-dimensional geometry over finite fields*. Ed. by Kaledin, D. and Tschinkel, Y. NATO science for peace and security series. Sub-series D: information and communication security 16. Amsterdam, Netherlands: IOS Press, 2008, pp. 269–314. URL: http://www.math.nyu.edu/~tschinke/books/finite-fields/final/09_wedhorn.pdf (visited on 08/06/2018).

Indices

General index

Index of symbols

Symbols