Contents

Preface

This is a book about algorithms for computing with modular forms that started as a series of notes for a graduate course at Harvard University in 2004. This book is meant to answer the question "How do *you* compute spaces of modular forms", by both providing a clear description of the specific algorithms that are used and explaining how to apply them using SAGE [SJ05].

I have spent many years trying to find good practical ways to compute with classical modular forms for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$, and have implemented most of these algorithms several times, first in C++ [Ste99], then in MAGMA [BCP97], and most recently as part of SAGE. Much of this work has involved turning formulas and constructions burried in obscure research papers into precise computational recipes, then testing these in many cases and eliminating subtle inaccuracies (published theorems sometimes contain small mistakes that appear magnified when implemented and run on a computer). The goal of this book is to explain some of what I have learned along the way.

The author is aware of no other books on computing with modular forms, the closest work being Cremona's book [Cre97a], which is about computing with elliptic curves, and Cohen's book [Coh93] about algebraic number theory. The field is not yet mature, and there are missing details and potential improvements to many of the algorithms, which you the reader might fill in, and which would be greatly appreciated by other mathematicians.

This book focuses on how best to compute the spaces $M_k(N, \varepsilon)$ of modular forms, where $k \geq 2$ is an integer and ε is a Dirichlet character modulo N. I will spend the most effort explaining the algorithms that appear so far to be the best (in practice!) for such computations. I will not discuss computing half-integral weight forms, weight one forms, forms for non-congruence subgroups or groups other than GL_2, Hilbert and Siegel modular forms, trace formulas, p-adic modular forms, and modular abelian varieties, all of which are topics for another book.

The reader is not assumed to have prior exposure to modular forms, but should be familiar with abstract algebra, basic algebraic number theory, Riemann surfaces, and complex analysis.

Acknowledgement. Kevin Buzzard made many helpful remarks which were helpful in finding the algorithms in Chapter 2. Noam Elkies made many remarks about chapters 1 and 2. The students in the Harvard course made help-

ful remark; in particular, Abhinav Kumar made observations about computing widths of cusps, Thomas James Barnet-Lamb about how to represent Dirichlet characters, and Tseno V. Tselkov, Jennifer Balakrishnan and Jesse Kass made other remarks.

Parts of Chapter 7 follow [Ser73, Ch. VII] closely, though we adjust the notation, definitions, and order of presentation to be consistent with the rest of this book. (For example, Serre writes $2k$ for the weight instead of k.)

Grant Information. This material is based upon work supported by the National Science Foundation under Grant No. 0400386.

Chapter 1

Modular Forms

1.1 Basic Definitions

Modular forms are certain types of functions on the *complex upper half plane*

$$\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}.$$

The group

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : ad - bc = 1, \text{ and } a, b, c, d \in \mathbb{R} \right\}$$

acts on $\mathfrak{h}$ via linear fractional transformations, as follows. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $z \in \mathfrak{h}$, then (see Exercise 7.1)

$$\gamma(z) = \frac{az + b}{cz + d} \in \mathfrak{h}. \tag{1.1.1}$$

Definition 1.1.1 (Modular Group)**.** The *modular group* is the subgroup $\mathrm{SL}_2(\mathbb{Z})$ of $\mathrm{SL}_2(\mathbb{R})$ of matrices with integer entries. Thus is is the group of all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$.

For example, the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{1.1.2}$$

are both elements of $\mathrm{SL}_2(\mathbb{Z})$; the matrix S defines the function $z \mapsto -1/z$, and T the function $z \mapsto z + 1$.

Theorem 1.1.2. *The group* $\mathrm{SL}_2(\mathbb{Z})$ *is generated by* S *and* T*.*

Proof. See e.g. [Ser73, §VII.1], which uses the fundamental domain $\mathcal{F}$ consisting of all elements of $\mathfrak{h}$ that satisfy $|z| \geq 1$ and $\mathrm{Re}(z) \leq 1/2$. $\square$

In SAGE we compute the group $\mathrm{SL}_2(\mathbb{Z})$ and its generators as follows:

```
sage: G = SL(2,Z)
sage: print G
The modular group SL(2,Z)
sage: S, T = G.gens()
sage: S
[ 0 -1]
[ 1  0]
sage: T
[1 1]
[0 1]
```

Definition 1.1.3 (Holomorphic and Meromorphic)**.** A function $f : \mathfrak{h} \to \mathbb{C}$ is *holomorphic* if f is complex differentiable at every point $z \in \mathfrak{h}$, i.e., for each $z \in \mathfrak{h}$ the limit $\lim_{h\to 0}(f(z+h) - f(z))/h$ exists, where h may approach 0 along any path. The function f is *meromorphic* if it is holomorphic except (possibly) at a discrete set of points in $\mathfrak{h}$.

The function $f(z) = e^z$ is a holomorphic function on $\mathfrak{h}$ (in fact on all of $\mathbb{C}$). The function $1/(z - i)$ is meromorphic on $\mathfrak{h}$, and fails to be analytic at i.

Modular forms are holomorphic functions on $\mathfrak{h}$ that transform in a particular way under a subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Before definining general modular forms, we define modular forms of level 1.

1.2 Modular Forms of Level 1

Definition 1.2.1 (Weakly Modular Function)**.** A *weakly modular function* of weight $k \in \mathbb{Z}$ is a meromorphic function f on $\mathfrak{h}$ such that for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and all $z \in \mathfrak{h}$ we have

$$f(z) = (cz + d)^{-k} f(\gamma(z)). \tag{1.2.1}$$

The constant functions are weakly modular of weight 0. There are no nonzero weakly modular functions of odd weight (see Exercise 7.4), and it is by no means obvious that there are any weakly modular functions of even weight $k \geq 2$. The product of two weakly modular functions of weights k_1 and k_2 is a weakly modular function of weight $k_1 + k_2$ (see Exercise 7.3), so once we find some nonconstant weakly modular functions, we'll find many of them.

When k is even (1.2.1) has a possibly more conceptual interpretation; namely (1.2.1) is the same as

$$f(\gamma(z))d(\gamma(z))^{k/2} = f(z)dz^{k/2}.$$

Thus (1.2.1) simply says that the weight k "differential form" $f(z)dz^{k/2}$ is fixed under the action of every element of $\mathrm{SL}_2(\mathbb{Z})$.

Since $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices S and T of (1.1.2), to show that a meromorphic function f on $\mathfrak{h}$ is a weakly modular function all we have to do is show that for all $z \in \mathfrak{h}$ we have

$$f(z+1) = f(z) \qquad \text{and} \qquad f(-1/z) = z^k f(z). \tag{1.2.2}$$

Suppose that f is a weakly modular function of some weight k. Then f might have a *Fourier expansion*, which we try to obtain as follows. Let $q = q(z) = e^{2\pi iz}$, which we view as a holomorphic function $\mathbb{C} \cup \infty \to D$, where D is the closed unit disk. Let D' be the punctured unit disk, i.e., D with the origin removed, and note that $q : \mathbb{C} \to D'$. By (1.2.2) we have $f(z+1) = f(z)$, so there is a set-theoretic map $F : D' \to \mathbb{C}$ such that for every $z \in \mathfrak{h}$ we have $F(q(z)) = f(z)$. This function F is thus a complex-valued function on the open unit disk. It may or may not be well behaved at 0.

Suppose that F is well-behaved at 0, namely that for some $m \in \mathbb{Z}$ and all q in a neighborhood of 0 we have the equality

$$F(q) = \sum_{n=m}^{\infty} a_n q^n.$$

If this is the case, we say that f is *meromorphic at* ∞. If, moreover, $m \geq 0$, then we say that f is *holomorphic at* ∞.

Definition 1.2.2 (Modular Function)**.** A *modular function* of weight k is a weakly modular function of weight k that is meromorphic at ∞.

Definition 1.2.3 (Modular Form)**.** A *modular form* of weight k (and level 1) is a modular function of weight k that is holomorphic on $\mathfrak{h}$ and at ∞.

If f is a modular form, then there are complex numbers a_n such that for all $z \in \mathfrak{h}$,

$$f(z) = \sum_{n=0}^{\infty} a_n q^n = \sum_{n=0}^{\infty} a_n e^{2\pi inz}.$$

Proposition 1.2.4. *The above series converges for all* $z \in \mathfrak{h}$.

Proof. n The function $f(q)$ is holomorphic on D, so its Taylor series converges absolutely in D. See also [Ser73, §VII.4] for an explicit bound on the $|a_n|$. $\square$

We set $f(\infty) = a_0$, since $q^{2\pi iz} \to 0$ as $z \to i\infty$, and the value of f at ∞ should be the value of F at 0, which is a_0 from the power series.

Definition 1.2.5 (Cusp Form)**.** A *cusp form* of weight k (and level 1) is a modular form of weight k such that $f(\infty) = 0$, i.e., $a_0 = 0$.

1.3 Modular Forms of Any Level

We next define spaces of modular forms of level possibly bigger than 1. When $k = 2$ these are closely related to elliptic curves and abelian varieties.

For each positive integer N, define a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ as follows:

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}.$$

This is a "congruence subgroup", since it is given by congruence conditions. We are now in a position to define $M_k(\Gamma_1(N))$.

Definition 1.3.1 (Modular Forms)**.** Let $M_k(\Gamma_1(N))$ be the complex vector space of holomorphic functions $f : \mathfrak{h}^* \to \mathbb{C}$ such that $f|[\gamma]_k = f$ for all $\gamma \in \Gamma_1(N)$.

What it means for f to be holomorphic at the elements of $\mathbb{Q} \cup \{i\infty\}$ is subtle. We say f is *holomorphic* at $i\infty$ if its q-expansion $\sum a_n q^n$ has no nonzero coefficient a_n for $n < 0$. To make sense of holomorphicity of f at $\alpha \in \mathbb{Q}$, let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ be such that $\gamma(\infty) = \alpha$. We say f is holomorphic at α if $f|[\gamma]_k$ is holomorphic at infinity. Note that formally

$$f|[\gamma]_k(\infty) = (cz + d)^{-k} f(\alpha),$$

where (c, d) is the bottom row of γ and the factor $(cz + d)^{-k}$ does not affect holomorphicity at α.

Another subtlety hidden in this definition is that $f|[\gamma]_k$ is a modular form for the conjugate group $G = \gamma^{-1}\Gamma_1(N)\gamma$, which need not equal $\Gamma_1(N)$. In particular, the matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ need not be in G, so $f|[\gamma]_k$ need not even have a power series expansion $\sum_{n\in\mathbb{Z}} b_n q^n$ at infinity! Fortunately (see Exercise 3.1) there is some positive integer h such that $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in G$, so $f|[\gamma]_k$ has a power series expansion $\sum_{n\in\mathbb{Z}} b_{n/h} q^{n/h}$ in powers of $q^{1/h}$, and we again say $f|[\gamma]_k$ is holomorphic at infinity if $b_{n/h} = 0$ for all $n < 0$. (The reason we obtain a power series in $q^{1/h}$ is that $f|[\gamma]_k(hz)$ is invariant under $z \mapsto z + 1$, so $f|[\gamma]_k(hz)$ has an expansion in powers of q.)

A *congruence subgroup* is a subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ that contains the kernel $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$ for some N. The smallest such N is the *level* of Γ.

Definition 1.3.2 (Width of Cusp)**.** The minimal h such that $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \gamma^{-1}\Gamma\gamma$ is called the *width of the cusp* $\gamma(\infty)$ for the group Γ.

Algorithm 1.3.3 (Width of Cusp)**.**
Given a congruence subgroup Γ of level N and a cusp α for Γ, this algorithm computes the width h of α. We assume that Γ is given by congruence conditions, e.g., $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$.

1. [Find γ] Using the extended Euclidean algorithm, find $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(\infty) = \alpha$. If $\alpha = \infty$ set $\gamma \leftarrow 1$; otherwise, write $\alpha = a/b$, find c, d such that $ad - bc = 1$, and set $\gamma \leftarrow \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

2. [Generic Conjugate Matrix] Compute the following matrix in $M_2(\mathbb{Z}[x])$:

$$\delta(x) \leftarrow \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1}.$$

   Note that $\delta(x)$ matrix whose entries are constant or linear in x.

3. [Solve] The congruence conditions that define Γ give rise to four linear congruence conditions on x. Use techniques from elementary number theory to find the smallest simultaneous positive solution h to these four equations.

Example 1.3.4.

1. Suppose $\alpha = 0$ and $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$. Then $\gamma = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ has the property that $\gamma(\infty) = \alpha$. Next, the congruence condition is

$$\delta(x) = \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1} = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N}.$$

   Thus the smallest positive solution is $h = N$, so the width of 0 is N.

2. Suppose $N = pq$ where p, q are distinct primes, and let $\alpha = 1/p$. Then $\gamma = \left(\begin{smallmatrix} 1 & 0 \\ p & 1 \end{smallmatrix}\right)$ sends ∞ to α. The congruence condition for $\Gamma_0(pq)$ is

$$\delta(x) = \gamma \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \gamma^{-1} = \begin{pmatrix} 1 - px & x \\ -p^2x & px + 1 \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{pq}.$$

   Since $p^2x \equiv 0 \pmod{pq}$, we see that $x = q$ is the smallest solution. Thus $1/p$ has width q, and likewise $1/q$ has width p.

Remark 1.3.5. For $\Gamma_0(N)$, once we enforce that the bottom left entry is 0 $\pmod N$, and use that the determinant is 1, the coprimeness that one gets from the other two congruences is automatic. So there is one congruence to solve for $\Gamma_0(N)$. There are 2 congruences in the $\Gamma_1(N)$ case (the bottom left entry and top left entry).

1.4 Eisenstein Series and Delta

For an even integer $k \geq 4$, define the (non-normalized) *weight k Eisenstein series* to be

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

where the sum is over all $m, n \in \mathbb{Z}$ such that $mz + n \neq 0$.

Proposition 1.4.1. *The function $G_k(z)$ is a modular form of weight k.*

be the Δ-function.

Conjecture 1.8.1 (Edixhoven, et al.)**.** *There is an algorithm to compute $\tau(p)$, for prime p, that is polynomial-time in* $\log(p)$*. More generally, suppose $f = \sum a_n q^n$ is an eigenform in some space $M_k(N, \varepsilon)$, where $k \geq 2$. Then there is an algoirthm to compute a_p, for p prime, in time polynomial in* $\log(p)$*.*

Bas Edixhoven and his students have been working for years to apply sophisticated techniques from arithmetic geometry (e.g., étale cohomology, motives, Arakelov theory) in order to prove that such an algorithm exists (among other things), and he believes they are almost there. There is evidently a significant gap between proving *existence* of an algorithm that shold be polynomial time, and actually writing down such an algorithm with explicitly bounded running times. The ideas Edixhoven uses are very similar to the ones used for counting points on elliptic curves in polynomial time (the algorithm of Schoof, with refinements by Atkins and Elkies).

1.9 Exercises

1.1 Suppose $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a matrix with real entries and positive determinant. Prove that if $z \in \mathbb{C}$ is a complex number with positive imaginary part, then the imaginary part of $\gamma(z) = (az + b)/(cz + d)$ is also positive.

1.2 (a) Prove that a polynomial is an analytic function on $\mathbb{C}$.

(b) Prove that a rational function (quotient of two polynomials) is a meromorphic function on $\mathbb{C}$.

1.3 Suppose f and g are weakly modular functions with $f \neq 0$.

(a) Prove that the product fg is a weakly modular function.

(b) Prove that $1/f$ is a weakly modular function.

(c) If f and g are modular functions, show that fg is a modular function.

(d) If f and g are modular forms, show that fg is a modular form.

1.4 Suppose f is a weakly modular function of odd weight k. Show that $f = 0$.

1.5 (a) Prove that $\Gamma_1(N)$ is a group.

(b) Prove that $\Gamma_1(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$ (Hint: it contains the kernel of the homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$.)