

INSTITUTE OF SCIENCE TOKYO
YOUNG SCIENCE AND ENGINEERING RESEARCHERS PROGRAM (YSEP)

Research Notes on Modular Forms and Hecke Operators

MOH HASBI ALFARIZI
YSEP RESEARCH STUDENT, INSTITUTE OF SCIENCE TOKYO

Version October 22, 2026

Preface

These notes are a typeset version of my handwritten notes for my research project on *modular forms*, which I carried out at the Institute of Science Tokyo during the Young Science and Engineering Researchers Program (YSEP), under the supervision of Prof. Yuichiro Taguchi and Specially Appointed Associate Professor Soma Purkait. The final goal of the project is to understand *Hecke operators*: the Hecke algebra of a congruence subgroup and the action of the Hecke operators T_n on spaces of modular forms for $\Gamma_0(N)$, which leads to bases of eigenforms whose Fourier coefficients are multiplicative. To get there, the notes start from the analytic continuation and functional equation of the Riemann zeta function, whose proof (via the Gamma function, the Mellin and Fourier transforms, Poisson summation and the theta function) is the prototype of the analytic arguments used later. They then pass through elliptic curves over $\mathbb{C}$, the modular group and its congruence subgroups, the valence formula, the growth of modular forms and cusp forms, and modular forms of congruence subgroups and their properties. As an application we prove Jacobi’s four square theorem, and we finally develop the Hecke algebra and the Hecke operators for $\Gamma_0(N)$. Much of the material follows the lecture notes of Dr. Soma Purkait from the course on modular forms taught by Dr. Purkait a few years ago; the presentation also follows Koblitz [2], supplemented by [1, 4]. Chapter k follows my handwritten “Note k ”.

Why the Riemann zeta function and elliptic curves? Originally the project was meant to go further in two directions: from modular forms to their L -functions (which generalise the Riemann zeta function), and from elliptic curves to the modularity theorem (every elliptic curve over $\mathbb{Q}$ is modular). Chapters 1 and 2 were written as preparation for these. Since the research period was coming to an end, there was not enough time to reach either goal, and the notes stop at Hecke operators.

In my handwritten notes many statements were only announced (“we can show”, “by the lemma”, “exactly like in the book”), and some proofs were left blank. Most of these proofs were written afterwards to complete the notes. The conversion of my handwritten notes into this L^AT_EX document, as well as several of the proofs and corrections, was produced with the help of Claude Code (Opus 5.5, High-Ultracode), in order to ease the formal write-up. Furthermore:

- boxes labelled **Notes**. make precise notions which were ambiguous or left implicit in my handwritten notes;
- boxes labelled **Correction**. point out places where my handwritten notes contained a slip, and what the correct statement is;
- results whose proofs are far beyond the scope of the project are stated with references; a *Proof sketch* indicates that some standard input is quoted.

Standard facts from a first course in complex analysis (Cauchy’s and Goursat’s theorems, Morera’s theorem, the residue theorem, the identity theorem, ...) and in real analysis (dominated convergence, the Fubini–Tonelli theorem, ...) are used freely; see e.g. [6, 7].

I would also like to give a shout-out to Orlando Ferrari’s lecture notes project, which influenced the way I make and write up these notes; see orlandoferrari.github.io.

The reader is advised to stay vigilant and verify the calculations; any remaining errors are mine. If you find a mistake or have a suggestion, please feel free to reach out to me on GitHub: github.com/Peanut2Butter9.

Contents

Preface	i
1 Analytic Continuation of the Riemann Zeta Function	1
1.1 The Gamma function	1
1.2 The reflection and duplication formulas	6
1.3 The Riemann zeta function for $\text{Re } s > 1$	9
1.4 The Mellin and Fourier transforms	11
1.5 Poisson summation and the theta function	13
1.6 Analytic continuation and functional equation of ζ	16
2 Elliptic Curves and Elliptic Functions	20
2.1 Elliptic curves over a field	20
2.2 Lattices and doubly periodic functions	23
2.3 The Weierstrass $\wp$ -function	28
2.4 The Weierstrass equation and Eisenstein series	33
2.5 The torus $\mathbb{C}/L$ as an elliptic curve	37
3 The Modular Group and Congruence Subgroups	40
3.1 $\text{SL}_2(\mathbb{Z})$ and its action on the upper half-plane	40
3.2 Congruence subgroups	42
3.3 Fundamental domains	48
3.4 Exercises from Koblitz, §III.1	51
4 Modular Forms of Level 1 and the Valence Formula	63
4.1 Modular functions and modular forms	63
4.2 Eisenstein series and their q -expansions	65
4.3 The valence formula	70
4.4 Consequences of the valence formula	72
4.5 Dimension formula and generators	74
4.6 The j -invariant	76
4.7 The Eisenstein series of weight 2	78
4.8 The Dedekind η -function and the product formula for Δ	81
4.9 Derivatives of modular forms	83
5 Growth of Fourier Coefficients and Cusp Forms	85
6 Modular Forms of Congruence Subgroups	88

6.1	The slash operator	88
6.2	Modular forms for congruence subgroups	89
6.3	Stabilisers of cusps and odd weight	92
6.4	The valence formula and a dimension bound	93
6.5	Growth conditions and cusp forms	94
7	Properties of Modular Forms of Congruence Subgroups	98
7.1	Changing the group	98
7.2	Dirichlet characters	100
7.3	Decomposition by the Nebentypus	102
7.4	Oldforms and the Atkin–Lehner involution	103
7.5	Eisenstein series for $\Gamma_0(N)$	105
7.6	Jacobi’s theta function	107
8	Jacobi’s Four Squares Theorem and Hecke Operators of Level 1	110
8.1	Proof of Jacobi’s four squares theorem	110
8.2	Modular forms as functions on lattices	112
8.3	Hecke operators on $M_k(\mathrm{SL}_2(\mathbb{Z}))$	113
8.4	Multiplicativity of Hecke operators	117
8.5	Hecke eigenforms	119
9	The Hecke Algebra	122
9.1	Commensurable subgroups and the commensurator	122
9.2	Double cosets	123
9.3	The Hecke algebra	124
10	Hecke Operators for $\Gamma_0(N)$	127
10.1	The Hecke algebra acts on $M_k(\Gamma, \chi)$	127
10.2	The Hecke algebras of $\Gamma_0(N)$	128
10.3	The Hecke operators T_n and T_n^*	130
10.4	Decomposition into single cosets	132
10.5	Generators of the Hecke algebra and the action on q -expansions	133

Chapter 1

Analytic Continuation of the Riemann Zeta Function

The aim of this chapter is to prove that the Riemann zeta function $\zeta(s) = \sum_{n \geq 1} n^{-s}$, defined for $\operatorname{Re} s > 1$, extends to a meromorphic function on $\mathbb{C}$ whose only pole is a simple pole at $s = 1$ with residue 1, and that it satisfies the functional equation

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{-(1-s)/2} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s).$$

We follow [2, Ch. II]. The proof goes through the Gamma function, the Mellin and Fourier transforms, the Poisson summation formula and the transformation law of the theta function. Throughout, we write $s = \sigma + i\tau$ with $\sigma = \operatorname{Re} s$ and $\tau = \operatorname{Im} s$, and for $t > 0$ we put $t^s = e^{s \log t}$ with the real logarithm. Note that

$$|t^s| = \left| e^{\sigma \log t} \right| \left| e^{i\tau \log t} \right| = t^\sigma, \quad (1.1)$$

since $|e^{i\theta}| = 1$ for real θ .

1.1 The Gamma function

Definition 1.1.1 (Gamma function)

For $\operatorname{Re} s > 0$ we put

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt.$$

Proposition 1.1.2

The integral defining $\Gamma(s)$ converges absolutely for every s with $\operatorname{Re} s > 0$.

Proof. Let $s = \sigma + i\tau$ with $\sigma > 0$. Since an absolutely convergent improper integral is convergent, it suffices to look at the modulus of the integrand. By (1.1), $|e^{-t} t^{s-1}| = e^{-t} t^{\sigma-1}$, so we have to show that

$$\int_0^\infty e^{-t} t^{\sigma-1} dt = \int_0^1 e^{-t} t^{\sigma-1} dt + \int_1^\infty e^{-t} t^{\sigma-1} dt$$

is finite. We split the integral into these two parts.

The part over $(0, 1]$. Since $e^{-t} \leq 1$ for $t \geq 0$, we have $e^{-t} t^{\sigma-1} \leq t^{\sigma-1}$, and since $\sigma > 0$,

$$\int_0^1 t^{\sigma-1} dt = \lim_{\varepsilon \rightarrow 0^+} \left[\frac{t^\sigma}{\sigma} \right]_\varepsilon^1 = \frac{1}{\sigma} - \lim_{\varepsilon \rightarrow 0^+} \frac{\varepsilon^\sigma}{\sigma} = \frac{1}{\sigma}.$$

Hence $\int_0^1 e^{-t} t^{\sigma-1} dt \leq 1/\sigma < \infty$.

The part over $[1, \infty)$. The exponential dominates every power. To make this precise, choose an integer $N \geq \max(\sigma - 1, 0)$. From the power series $e^x = \sum_{j \geq 0} x^j/j!$, all of whose terms are ≥ 0 for $x \geq 0$, we get $e^x \geq x^N/N!$ for $x \geq 0$. With $x = t/2$ and $t \geq 1$,

$$t^{\sigma-1} \leq t^N = 2^N \left(\frac{t}{2}\right)^N \leq 2^N N! e^{t/2} =: M e^{t/2},$$

where the first inequality uses $t \geq 1$ and $\sigma - 1 \leq N$ (for $t \geq 1$ the function $x \mapsto t^x$ is increasing). Hence

$$\begin{aligned} \int_1^\infty e^{-t} t^{\sigma-1} dt &\leq \int_1^\infty e^{-t} \cdot M e^{t/2} dt = M \int_1^\infty e^{-t/2} dt \\ &= M \lim_{b \rightarrow \infty} \left[-2e^{-t/2} \right]_1^b = M \lim_{b \rightarrow \infty} (2e^{-1/2} - 2e^{-b/2}) = 2Me^{-1/2} < \infty. \end{aligned}$$

Both parts are finite, so the integral converges absolutely. ■

To prove that Γ is holomorphic we use two standard lemmas from complex analysis (see [6, Ch. 2, Thms. 5.2 and 5.4]). Both rest on the following two theorems.

Recall (Goursat and Morera). *Goursat's theorem:* if $\Omega \subseteq \mathbb{C}$ is open, $T \subseteq \Omega$ is a closed triangle (with its interior) and f is holomorphic on Ω , then $\int_{\partial T} f(z) dz = 0$. *Morera's theorem:* if f is continuous on an open disc D and $\int_{\partial T} f(z) dz = 0$ for every closed triangle $T \subseteq D$, then f is holomorphic on D .

Lemma 1.1.3

Let $\Omega \subseteq \mathbb{C}$ be open and let $(f_n)_{n \geq 0}$ be a sequence of holomorphic functions on Ω that converges uniformly to a function f on every compact subset of Ω . Then f is holomorphic on Ω .

Proof. Continuity. Let $z_0 \in \Omega$ and choose $r > 0$ with $\overline{D(z_0, r)} \subseteq \Omega$; this closed disc is compact. Given $\varepsilon > 0$ choose n with $|f_n - f| < \varepsilon/3$ on $\overline{D(z_0, r)}$, and then $\delta \in (0, r)$ with $|f_n(z) - f_n(z_0)| < \varepsilon/3$ for $|z - z_0| < \delta$ (f_n is continuous). For such z ,

$$|f(z) - f(z_0)| \leq |f(z) - f_n(z)| + |f_n(z) - f_n(z_0)| + |f_n(z_0) - f(z_0)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

So f is continuous on Ω .

Holomorphy. Let $D \subseteq \Omega$ be an open disc and $T \subseteq D$ a closed triangle. By Goursat's theorem $\int_{\partial T} f_n(z) dz = 0$ for every n . Since ∂T is compact, $\sup_{\partial T} |f_n - f| \rightarrow 0$, and with $\ell(\partial T)$ the length of ∂T ,

$$\left| \int_{\partial T} f(z) dz - \int_{\partial T} f_n(z) dz \right| \leq \ell(\partial T) \cdot \sup_{z \in \partial T} |f(z) - f_n(z)| \xrightarrow{n \rightarrow \infty} 0,$$

hence

$$\int_{\partial T} f(z) dz = \lim_{n \rightarrow \infty} \int_{\partial T} f_n(z) dz = \lim_{n \rightarrow \infty} 0 = 0.$$

By Morera's theorem f is holomorphic on D ; as D was arbitrary, f is holomorphic on Ω . ■

Lemma 1.1.4

Let $\Omega \subseteq \mathbb{C}$ be open, $a < b$ real, and let $F(z, s)$ be a function on $\Omega \times [a, b]$ such that

- (i) $F(z, s)$ is holomorphic in z for each fixed $s \in [a, b]$;
- (ii) F is continuous on $\Omega \times [a, b]$.

Then the function f defined on Ω by

$$f(z) = \int_a^b F(z, s) ds$$

is holomorphic on Ω .

Proof. Let D be an open disc whose closure $\overline{D}$ lies in Ω .

Continuity. F is continuous on the compact set $\overline{D} \times [a, b]$, hence uniformly continuous there. So, given $\varepsilon > 0$, there is $\delta > 0$ such that $|F(z, s) - F(z_0, s)| < \varepsilon$ for all $s \in [a, b]$ whenever $z, z_0 \in \overline{D}$ and $|z - z_0| < \delta$. For such z, z_0 ,

$$|f(z) - f(z_0)| = \left| \int_a^b (F(z, s) - F(z_0, s)) ds \right| \leq \int_a^b |F(z, s) - F(z_0, s)| ds \leq (b - a) \varepsilon.$$

So f is continuous on D .

Holomorphy. Let $T \subseteq D$ be a closed triangle. The integrand $F(z, s)$ is continuous on the compact set $\partial T \times [a, b]$, hence bounded there, so it is integrable for the product of arc length and ds , and Fubini's theorem allows us to swap the integrals. By hypothesis (i) and Goursat's theorem, $\int_{\partial T} F(z, s) dz = 0$ for each fixed s . Hence

$$\int_{\partial T} f(z) dz = \int_{\partial T} \left(\int_a^b F(z, s) ds \right) dz = \int_a^b \left(\int_{\partial T} F(z, s) dz \right) ds = \int_a^b 0 ds = 0.$$

By Morera's theorem f is holomorphic on D . Every point of Ω lies in such a disc, so f is holomorphic on Ω . ■

Correction. In my handwritten notes the lemma was stated for $(z, s) \in \Omega \times [0, 1]$, but condition (ii) and the integral were over $[a, b]$. The interval is arbitrary; it must of course be the same everywhere. We will apply the lemma with $[a, b] = [\varepsilon, 1/\varepsilon]$.

Proposition 1.1.5

The Gamma function is holomorphic on the half-plane $\{s \in \mathbb{C} : \operatorname{Re} s > 0\}$.

Proof. For $0 < a < b$ consider the closed vertical strip

$$K_{a,b} = \{s \in \mathbb{C} : a \leq \operatorname{Re} s \leq b\}.$$

Every compact subset K of $\{\operatorname{Re} s > 0\}$ is contained in such a strip: since Re is continuous and K is compact, $\operatorname{Re} s$ attains a minimum $a > 0$ and a maximum b on K .

For $0 < \varepsilon < 1$ put

$$F_\varepsilon(s) = \int_\varepsilon^{1/\varepsilon} e^{-t} t^{s-1} dt.$$

The function $I(s, t) = e^{-t} t^{s-1} = e^{-t} e^{(s-1) \log t}$ is holomorphic in s for each fixed $t > 0$ and continuous on $\{\operatorname{Re} s > 0\} \times [\varepsilon, 1/\varepsilon]$, so F_ε is holomorphic on $\{\operatorname{Re} s > 0\}$ by Lemma 1.1.4.

Next we show that $F_\varepsilon \rightarrow \Gamma$ uniformly on $K_{a,b}$ as $\varepsilon \rightarrow 0^+$. For $s = \sigma + i\tau \in K_{a,b}$,

$$|\Gamma(s) - F_\varepsilon(s)| = \left| \int_0^\varepsilon e^{-t} t^{s-1} dt + \int_{1/\varepsilon}^\infty e^{-t} t^{s-1} dt \right| \leq \underbrace{\int_0^\varepsilon e^{-t} t^{\sigma-1} dt}_{(1)} + \underbrace{\int_{1/\varepsilon}^\infty e^{-t} t^{\sigma-1} dt}_{(2)}.$$

Bound for (1). Here $0 < t < \varepsilon < 1$, so $e^{-t} \leq 1$, and since $\sigma \geq a$ we have $t^{\sigma-1} \leq t^{a-1}$ (for $0 < t < 1$ the function $x \mapsto t^x$ is decreasing). Hence

$$\int_0^\varepsilon e^{-t} t^{\sigma-1} dt \leq \int_0^\varepsilon t^{a-1} dt = \left[\frac{t^a}{a} \right]_0^\varepsilon = \frac{\varepsilon^a}{a}.$$

Bound for (2). Here $t > 1/\varepsilon > 1$, and since $\sigma \leq b$ we have $t^{\sigma-1} \leq t^{b-1}$. Arguing as in the proof of Proposition 1.1.2, there is $M_b > 0$ with $t^{b-1} \leq M_b e^{t/2}$ for all $t \geq 1$, so

$$\int_{1/\varepsilon}^{\infty} e^{-t} t^{\sigma-1} dt \leq \int_{1/\varepsilon}^{\infty} M_b e^{-t/2} dt = 2M_b e^{-1/(2\varepsilon)}.$$

Altogether,

$$\sup_{s \in K_{a,b}} |\Gamma(s) - F_\varepsilon(s)| \leq \frac{\varepsilon^a}{a} + 2M_b e^{-1/(2\varepsilon)} \xrightarrow{\varepsilon \rightarrow 0^+} 0.$$

Taking $\varepsilon = \varepsilon_n = 1/(n+2)$, the holomorphic functions F_{ε_n} converge to Γ uniformly on every compact subset of $\{\operatorname{Re} s > 0\}$. By Lemma 1.1.3, Γ is holomorphic on $\{\operatorname{Re} s > 0\}$. ■

Notes. The upper bounds for (1) and (2) are deliberately chosen to depend only on a , b and ε , and not on s : integrating $t^{\sigma-1}$ directly would give a bound depending on σ , which only proves pointwise convergence. Uniformity is exactly what Lemma 1.1.3 requires. Note also that the strip $K_{a,b}$ is closed but not compact (it is unbounded in the imaginary direction); this does not matter, since uniform convergence on $K_{a,b}$ implies uniform convergence on each compact subset of it.

Proposition 1.1.6

For $\operatorname{Re} s > 0$ we have $\Gamma(s+1) = s\Gamma(s)$. In particular $\Gamma(n) = (n-1)!$ for all integers $n \geq 1$.

Proof. Integrate by parts with $u = t^s$ and $dv = e^{-t} dt$, so that $du = s t^{s-1} dt$ and $v = -e^{-t}$:

$$\Gamma(s+1) = \lim_{\varepsilon \rightarrow 0^+} \int_{\varepsilon}^{1/\varepsilon} t^s e^{-t} dt = \lim_{\varepsilon \rightarrow 0^+} \left([-t^s e^{-t}]_{\varepsilon}^{1/\varepsilon} + s \int_{\varepsilon}^{1/\varepsilon} e^{-t} t^{s-1} dt \right).$$

Here $\frac{d}{dt} t^s = \frac{d}{dt} e^{s \log t} = \frac{s}{t} e^{s \log t} = s t^{s-1}$, which justifies du . By (1.1), $|t^s e^{-t}| = t^\sigma e^{-t}$. We check the two ends of the boundary term:

- at $t = 1/\varepsilon \rightarrow \infty$: with $N \geq \sigma$ an integer, $e^t \geq t^{N+1}/(N+1)!$, so $t^\sigma e^{-t} \leq t^N e^{-t} \leq (N+1)!/t \rightarrow 0$ (using $t \geq 1$);
- at $t = \varepsilon \rightarrow 0^+$: $t^\sigma e^{-t} \leq t^\sigma = \varepsilon^\sigma \rightarrow 0$, since $\sigma > 0$.

Hence $[-t^s e^{-t}]_{\varepsilon}^{1/\varepsilon} = -(1/\varepsilon)^s e^{-1/\varepsilon} + \varepsilon^s e^{-\varepsilon} \rightarrow 0 - 0 = 0$, and

$$\Gamma(s+1) = 0 + s \lim_{\varepsilon \rightarrow 0^+} \int_{\varepsilon}^{1/\varepsilon} e^{-t} t^{s-1} dt = s\Gamma(s).$$

Finally

$$\Gamma(1) = \int_0^{\infty} e^{-t} t^0 dt = \lim_{b \rightarrow \infty} [-e^{-t}]_0^b = \lim_{b \rightarrow \infty} (1 - e^{-b}) = 1 = 0!,$$

and by induction on n : if $\Gamma(n) = (n-1)!$, then $\Gamma(n+1) = n\Gamma(n) = n \cdot (n-1)! = n!$. ■

Recall (Identity theorem). If $\Omega \subseteq \mathbb{C}$ is open and connected, f, g are holomorphic on Ω , and $f = g$ on a subset of Ω having a limit point in Ω (e.g. a nonempty open subset, or a segment of positive length), then $f = g$ on all of Ω . In particular an analytic continuation to a connected domain is unique, and identities between holomorphic functions persist under continuation.

Proposition 1.1.7

The Gamma function has an analytic continuation to a meromorphic function on $\mathbb{C}$ whose only singularities are simple poles at $s = 0, -1, -2, \dots$. The identity $\Gamma(s+1) = s\Gamma(s)$ holds for all $s \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$.

Proof. The strategy is to push the domain of definition to the left, one unit at a time, using Proposition 1.1.6 in the form

$$\Gamma(s) = \frac{\Gamma(s+1)}{s} \quad (\operatorname{Re} s > 0).$$

For an integer $m \geq 1$ define

$$F_m(s) = \frac{\Gamma(s+m)}{(s+m-1)(s+m-2)\cdots(s+1)s} \quad (\operatorname{Re} s > -m).$$

This makes sense, because $\operatorname{Re}(s+m) > 0$. Since Γ is holomorphic on the right half-plane, F_m is holomorphic on $\{\operatorname{Re} s > -m\}$ except possibly at the zeros $s = 0, -1, \dots, -(m-1)$ of the denominator. At $s = -n$ with $0 \leq n \leq m-1$ the numerator equals $\Gamma(m-n) = (m-n-1)! \neq 0$ and the denominator has a simple zero, so F_m has a simple pole there.

$F_m = \Gamma$ on $\{\operatorname{Re} s > 0\}$. For $\operatorname{Re} s > 0$ all of $s, s+1, \dots, s+m-1$ have positive real part, so Proposition 1.1.6 applies to each of them:

$$\begin{aligned} \Gamma(s+m) &= (s+m-1)\Gamma(s+m-1) = (s+m-1)(s+m-2)\Gamma(s+m-2) \\ &= \cdots = (s+m-1)(s+m-2)\cdots(s+1)s\Gamma(s). \end{aligned}$$

Dividing by the product (which is nonzero, as $\operatorname{Re} s > 0$) gives $F_m(s) = \Gamma(s)$.

Compatibility. For $\operatorname{Re} s > -m$ we have $\operatorname{Re}(s+m) > 0$, so $\Gamma(s+m+1) = (s+m)\Gamma(s+m)$ and

$$F_{m+1}(s) = \frac{\Gamma(s+m+1)}{(s+m)(s+m-1)\cdots s} = \frac{(s+m)\Gamma(s+m)}{(s+m)(s+m-1)\cdots s} = \frac{\Gamma(s+m)}{(s+m-1)\cdots s} = F_m(s)$$

(away from the poles $0, -1, \dots, -(m-1)$). Hence the F_m glue together to a meromorphic function on $\mathbb{C} = \bigcup_m \{\operatorname{Re} s > -m\}$, which extends Γ and whose only singularities are simple poles at $0, -1, -2, \dots$. By the identity theorem this continuation is unique, and the identity $\Gamma(s+1) = s\Gamma(s)$, valid for $\operatorname{Re} s > 0$, persists on the connected open set $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$. We denote the continuation again by Γ . ■

Correction. In my handwritten notes the statement says that the poles are at $-2\mathbb{Z}$, and that F_m is defined “for all $m \in \mathbb{Z}$ ” with simple poles at $s = 0, -1, \dots, -m$. The correct statements are: the poles of Γ are exactly the non-positive integers $0, -1, -2, \dots$ (not only the even ones), m runs over the positive integers, and F_m has poles at $s = 0, -1, \dots, -(m-1)$ (the point $-m$ lies on the boundary of the half-plane $\operatorname{Re} s > -m$).

Proposition 1.1.8

For every integer $n \geq 0$,

$$\operatorname{Res}_{s=-n} \Gamma(s) = \frac{(-1)^n}{n!}.$$

Proof. Choose $m \geq n+1$, so that $\Gamma = F_m$ near $s = -n$. As $s \rightarrow -n$,

- $\Gamma(s+m) \rightarrow \Gamma(m-n) = (m-n-1)!;$

- the factors $s, s+1, \dots, s+n-1$ tend to $-n, -(n-1), \dots, -1$, whose product is $(-1)^n n!$;
- the factor $s+n$ has a simple zero (this is where the pole comes from);
- the factors $s+n+1, \dots, s+m-1$ tend to $1, 2, \dots, m-n-1$, whose product is $(m-n-1)!$.

Since $s = -n$ is a simple pole,

$$\begin{aligned} \operatorname{Res}_{s=-n} \Gamma(s) &= \lim_{s \rightarrow -n} (s+n) F_m(s) = \lim_{s \rightarrow -n} \frac{\Gamma(s+m)}{(s+m-1) \cdots (s+n+1)(s+n-1) \cdots s} \\ &= \frac{(m-n-1)!}{(m-n-1)!(-1)^n n!} = \frac{(-1)^n}{n!}. \end{aligned} \quad \blacksquare$$

Example 1.1.9. For $n = 0$ and $n = 1$ (taking $m = 2$, so $\Gamma(s) = \Gamma(s+2)/((s+1)s)$ for $\operatorname{Re} s > -2$):

$$\operatorname{Res}_{s=0} \Gamma(s) = \lim_{s \rightarrow 0} \frac{\Gamma(s+2)}{s+1} = \frac{\Gamma(2)}{1} = 1, \quad \operatorname{Res}_{s=-1} \Gamma(s) = \lim_{s \rightarrow -1} \frac{\Gamma(s+2)}{s} = \frac{\Gamma(1)}{-1} = -1,$$

in agreement with $(-1)^0/0! = 1$ and $(-1)^1/1! = -1$.

1.2 The reflection and duplication formulas

The following two classical identities were stated in my handwritten notes, but the proof was left blank; we supply one here. We will use them to show that Γ has no zeros, to compute $\Gamma(\frac{1}{2})$, and later in the study of ζ .

Proposition 1.2.1

For $s \in \mathbb{C} \setminus \mathbb{Z}$,

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}, \quad (1.2)$$

and for $s \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$,

$$\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right) = \sqrt{\pi} 2^{1-s} \Gamma(s). \quad (1.3)$$

We first need two lemmas.

Lemma 1.2.2 (Beta integral)

For $\operatorname{Re} x > 0$ and $\operatorname{Re} y > 0$ put $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1} du$. Then

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Proof. All integrals below converge absolutely (the absolute values of the integrands are the same integrands with x, y replaced by $\operatorname{Re} x, \operatorname{Re} y$, and these are finite by Proposition 1.1.2), so Fubini's theorem applies. We have

$$\Gamma(x)\Gamma(y) = \int_0^\infty e^{-a} a^{x-1} da \int_0^\infty e^{-b} b^{y-1} db = \int_0^\infty \int_0^\infty e^{-(a+b)} a^{x-1} b^{y-1} da db.$$

Consider $\Phi : (0, \infty) \times (0, 1) \rightarrow (0, \infty)^2$, $\Phi(r, u) = (a, b) = (ru, r(1-u))$. It is a bijection, with inverse $(a, b) \mapsto (r, u) = (a+b, \frac{a}{a+b})$: indeed $ru + r(1-u) = r$ and $\frac{ru}{r} = u$. Its Jacobian matrix and determinant are

$$\frac{\partial(a, b)}{\partial(r, u)} = \begin{pmatrix} \partial a / \partial r & \partial a / \partial u \\ \partial b / \partial r & \partial b / \partial u \end{pmatrix} = \begin{pmatrix} u & r \\ 1-u & -r \end{pmatrix}, \quad \det = u \cdot (-r) - r(1-u) = -ur - r + ur = -r,$$

so $da db = |-r| dr du = r dr du$. Substituting, $a + b = r$, $a^{x-1} = r^{x-1}u^{x-1}$ and $b^{y-1} = r^{y-1}(1-u)^{y-1}$, so

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_0^\infty \int_0^1 e^{-r} r^{x-1} u^{x-1} r^{y-1} (1-u)^{y-1} r du dr \\ &= \int_0^\infty e^{-r} r^{x+y-1} dr \int_0^1 u^{x-1} (1-u)^{y-1} du \\ &= \Gamma(x+y) B(x, y).\end{aligned}$$

To divide by $\Gamma(x+y)$ we need $\Gamma(x+y) \neq 0$. We only use the lemma for real $x, y > 0$, where $\Gamma(x+y) = \int_0^\infty e^{-t} t^{x+y-1} dt > 0$ (integral of a positive continuous function); in general it follows from Corollary 1.2.4 below, whose proof only uses the real case. ■

Lemma 1.2.3

For $0 < a < 1$,

$$\int_{-\infty}^\infty \frac{e^{ax}}{1+e^x} dx = \frac{\pi}{\sin \pi a}.$$

Proof. Convergence. For $x \geq 0$: $\frac{e^{ax}}{1+e^x} \leq \frac{e^{ax}}{e^x} = e^{-(1-a)x}$, and $1-a > 0$. For $x \leq 0$: $\frac{e^{ax}}{1+e^x} \leq e^{ax}$, and $a > 0$. Both bounds are integrable on the respective half-lines, so the integral converges absolutely; call it I .

The contour. Let $f(z) = e^{az}/(1+e^z)$ and let ∂R_R be the boundary of the rectangle with vertices $-R, R, R+2\pi i, -R+2\pi i$, oriented counterclockwise.

Poles. $1+e^z = 0 \iff e^z = -1 \iff z \in \pi i + 2\pi i\mathbb{Z}$. The only such point with $0 < \text{Im } z < 2\pi$ is $z = \pi i$. There $1+e^z$ has a simple zero, since its derivative $e^{\pi i} = -1 \neq 0$; so f has a simple pole with residue

$$\text{Res}_{z=\pi i} f = \frac{e^{a\pi i}}{\frac{d}{dz}(1+e^z)|_{z=\pi i}} = \frac{e^{a\pi i}}{e^{\pi i}} = -e^{\pi ia}.$$

The four sides.

- Bottom ($z = x$, x from $-R$ to R): $\int_{-R}^R f(x) dx$.
- Top ($z = x + 2\pi i$, x from R to $-R$): since $e^{x+2\pi i} = e^x$, $f(x+2\pi i) = \frac{e^{ax} e^{2\pi ia}}{1+e^x} = e^{2\pi ia} f(x)$, so the contribution is $-e^{2\pi ia} \int_{-R}^R f(x) dx$.
- Right ($z = R + iy$, $0 \leq y \leq 2\pi$): $|e^{az}| = e^{aR}$ and $|1+e^z| \geq |e^z| - 1 = e^R - 1$, so $|f(z)| \leq \frac{e^{aR}}{e^R - 1}$; the side has length 2π , so its integral is at most $\frac{2\pi e^{aR}}{e^R - 1} \rightarrow 0$ (as $a < 1$).
- Left ($z = -R + iy$): $|e^{az}| = e^{-aR}$ and $|1+e^z| \geq 1 - e^{-R}$, so the integral is at most $\frac{2\pi e^{-aR}}{1 - e^{-R}} \rightarrow 0$ (as $a > 0$).

Residue theorem. For every $R > 0$, $\int_{\partial R_R} f = 2\pi i \text{Res}_{z=\pi i} f = -2\pi i e^{\pi ia}$. Letting $R \rightarrow \infty$,

$$(1 - e^{2\pi ia}) I = -2\pi i e^{\pi ia}.$$

Since $0 < a < 1$, $e^{2\pi ia} \neq 1$, and dividing and then multiplying numerator and denominator by $-e^{-\pi ia}$,

$$I = \frac{-2\pi i e^{\pi ia}}{1 - e^{2\pi ia}} = \frac{2\pi i}{e^{\pi ia} - e^{-\pi ia}} = \frac{2\pi i}{2i \sin \pi a} = \frac{\pi}{\sin \pi a}. \quad \blacksquare$$

Proof of Proposition 1.2.1. Reflection formula. Let $0 < a < 1$. By Lemma 1.2.2 with $x = a$, $y = 1 - a$ and $\Gamma(1) = 1$,

$$\Gamma(a)\Gamma(1-a) = B(a, 1-a) = \int_0^1 u^{a-1}(1-u)^{-a} du.$$

Substitute $u = e^x/(1+e^x)$, which maps $\mathbb{R}$ increasingly onto $(0, 1)$ ($u \rightarrow 0$ as $x \rightarrow -\infty$, $u \rightarrow 1$ as $x \rightarrow \infty$). Then

$$1-u = \frac{1+e^x-e^x}{1+e^x} = \frac{1}{1+e^x}, \quad du = \frac{e^x(1+e^x) - e^x \cdot e^x}{(1+e^x)^2} dx = \frac{e^x}{(1+e^x)^2} dx,$$

and

$$u^{a-1} = e^{(a-1)x}(1+e^x)^{1-a}, \quad (1-u)^{-a} = (1+e^x)^a.$$

Multiplying, the powers of e^x combine to $e^{(a-1)x+x} = e^{ax}$ and the powers of $1+e^x$ to $(1+e^x)^{(1-a)+a-2} = (1+e^x)^{-1}$. Hence

$$\Gamma(a)\Gamma(1-a) = \int_{-\infty}^{\infty} e^{(a-1)x}(1+e^x)^{1-a} (1+e^x)^a \frac{e^x}{(1+e^x)^2} dx = \int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x} dx = \frac{\pi}{\sin \pi a}$$

by Lemma 1.2.3. Both sides of (1.2) are meromorphic on $\mathbb{C}$ and holomorphic on the connected open set $\mathbb{C} \setminus \mathbb{Z}$, and they agree on the interval $(0, 1)$; by the identity theorem they agree on $\mathbb{C} \setminus \mathbb{Z}$.

The value $\Gamma(\frac{1}{2})$. Putting $s = \frac{1}{2}$ in (1.2) gives $\Gamma(\frac{1}{2})^2 = \pi$. Since $\Gamma(\frac{1}{2}) = \int_0^{\infty} e^{-t} t^{-1/2} dt > 0$, we get

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \quad (1.4)$$

Duplication formula. Let $z > 0$ be real. The integrand of $B(z, z)$ is $(u(1-u))^{z-1}$, which is symmetric under $u \mapsto 1-u$; so the integrals over $(0, \frac{1}{2})$ and $(\frac{1}{2}, 1)$ are equal and

$$B(z, z) = 2 \int_0^{1/2} (u(1-u))^{z-1} du.$$

Substitute $v = 4u(1-u) = 4u - 4u^2$. Then $\frac{dv}{du} = 4 - 8u = 4(1-2u) > 0$ on $(0, \frac{1}{2})$, so v increases from $v(0) = 0$ to $v(\frac{1}{2}) = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1$. Moreover

$$(1-2u)^2 = 1 - 4u + 4u^2 = 1 - v, \quad \text{so} \quad 1-2u = \sqrt{1-v} \quad (\text{as } 1-2u > 0),$$

and $u(1-u) = v/4$, $du = \frac{dv}{4(1-2u)} = \frac{dv}{4\sqrt{1-v}}$. Therefore

$$B(z, z) = 2 \int_0^1 \left(\frac{v}{4}\right)^{z-1} \frac{dv}{4\sqrt{1-v}} = 2 \cdot 4^{-(z-1)} \cdot 4^{-1} \int_0^1 v^{z-1} (1-v)^{-1/2} dv = 2^{1-2z} B\left(z, \frac{1}{2}\right),$$

using $2 \cdot 4^{-z} = 2^{1-2z}$. By Lemma 1.2.2 (with $(x, y) = (z, z)$ and $(z, \frac{1}{2})$) and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ from (1.4), this reads

$$\frac{\Gamma(z)\Gamma(z)}{\Gamma(2z)} = 2^{1-2z} \frac{\Gamma(z)\sqrt{\pi}}{\Gamma(z+\frac{1}{2})}.$$

Multiplying both sides by $\Gamma(2z)\Gamma(z+\frac{1}{2})/\Gamma(z)$ (all three are positive for $z > 0$),

$$\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = \sqrt{\pi} 2^{1-2z} \Gamma(2z).$$

Putting $z = s/2$ gives (1.3) for real $s > 0$, and the identity theorem extends it to $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ (where both sides are holomorphic). ■

Corollary 1.2.4

$\Gamma(s) \neq 0$ for all $s \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$. Consequently $1/\Gamma(s)$ is an entire function whose zeros are exactly the points $0, -1, -2, \dots$, and they are all simple.

Proof. For $s = n \geq 1$ we have $\Gamma(n) = (n-1)! \neq 0$. If $s \notin \mathbb{Z}$, then $1-s \notin \mathbb{Z}_{\leq 0}$, so $\Gamma(1-s)$ is finite, and (1.2) shows $\Gamma(s)\Gamma(1-s) = \pi/\sin \pi s \neq 0$; hence $\Gamma(s) \neq 0$. Thus $1/\Gamma$ is holomorphic on $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$, and at each simple pole $-n$ of Γ it has a removable singularity with a simple zero. ■

Notes. On the same page of my handwritten notes there is a side computation showing that the theta series $\theta(t) = \sum_{n \in \mathbb{Z}} e^{-\pi t n^2}$ (introduced in Section 1.5) also converges for complex t with $\operatorname{Re} t > 0$: writing $t = \sigma + i\tau$,

$$\left| e^{-\pi t n^2} \right| = \left| e^{-\pi \sigma n^2} \right| \left| e^{-i\pi \tau n^2} \right| = e^{-\pi \sigma n^2}.$$

In Chapter 1 we only need $t > 0$, but the complex version is the starting point for theta functions as modular forms (with $t = -iz$, $z \in \mathfrak{H}$).

1.3 The Riemann zeta function for $\operatorname{Re} s > 1$

Definition 1.3.1 (Riemann zeta function)

For $\operatorname{Re} s > 1$ we put

$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}.$$

We recall the Weierstrass M -test, which holds verbatim for complex-valued functions.

Lemma 1.3.2 (Weierstrass M -test)

Let $f_1, f_2, \dots$ be complex-valued functions on a set X . Suppose there are constants M_k with

- (i) $|f_k(x)| \leq M_k$ for all $x \in X$ and $k \geq 1$, and
- (ii) $\sum_{k \geq 1} M_k < \infty$.

Then $\sum_{k \geq 1} f_k(x)$ converges absolutely and uniformly on X .

Proof. For each x , $\sum_k |f_k(x)| \leq \sum_k M_k < \infty$, so the series converges absolutely; let $f(x)$ be its sum. For every K and every $x \in X$,

$$\left| f(x) - \sum_{k=1}^K f_k(x) \right| = \left| \sum_{k > K} f_k(x) \right| \leq \sum_{k > K} |f_k(x)| \leq \sum_{k > K} M_k,$$

and the right-hand side is independent of x and tends to 0 as $K \rightarrow \infty$ (tail of a convergent series). Hence the convergence is uniform on X . ■

Proposition 1.3.3

The series defining $\zeta(s)$ converges absolutely for $\operatorname{Re} s > 1$, and ζ is holomorphic on the half-plane $\Omega = \{s \in \mathbb{C} : \operatorname{Re} s > 1\}$.

Proof. By (1.1), for $s = \sigma + i\tau$ we have $|n^{-s}| = |e^{-\sigma \log n}| |e^{-i\tau \log n}| = n^{-\sigma}$. Fix $\delta > 0$ and let $\Omega_\delta = \{s : \operatorname{Re} s \geq 1 + \delta\}$. For $s \in \Omega_\delta$,

$$|n^{-s}| = n^{-\sigma} \leq \frac{1}{n^{1+\delta}} =: M_n,$$

and $\sum_{n \geq 1} n^{-1-\delta}$ converges since $1 + \delta > 1$. By Lemma 1.3.2, the series $\sum_{n \geq 1} n^{-s}$ converges absolutely and uniformly on Ω_δ . Every compact subset of Ω lies in some Ω_δ (the minimum of $\operatorname{Re} s$ on it is > 1), so the partial sums, which are entire functions, converge uniformly on compact subsets of Ω . By Lemma 1.1.3, ζ is holomorphic on Ω . ■

Correction. In my handwritten notes $\Omega_\delta = \{\operatorname{Re} z \geq 1 + \delta\}$ is called compact. It is closed but not compact (it is unbounded). What the argument really uses is that every compact subset of Ω lies in some Ω_δ , on which the convergence is uniform.

An integral representation

Our next goal is to show that ζ extends analytically to $\mathbb{C}$, except for a simple pole at $s = 1$ with residue 1. A first link between ζ and Γ is the following formula.

Proposition 1.3.4

For $\operatorname{Re} s > 1$,

$$\Gamma(s) \zeta(s) = \int_0^\infty \frac{x^{s-1}}{e^x - 1} dx.$$

Proof. For $x > 0$ we have $0 < e^{-x} < 1$, so by the geometric series $\frac{1}{1-y} = \sum_{n \geq 0} y^n$ ($|y| < 1$),

$$\frac{x^{s-1}}{e^x - 1} = \frac{x^{s-1} e^{-x}}{1 - e^{-x}} = \sum_{n \geq 0} e^{-(n+1)x} x^{s-1}.$$

We want to integrate term by term. By Tonelli's theorem (applied to the nonnegative functions $e^{-(n+1)x} x^{s-1}$) and the substitution $t = (n+1)x$,

$$\sum_{n \geq 0} \int_0^\infty e^{-(n+1)x} x^{s-1} dx = \sum_{n \geq 0} \frac{\Gamma(s)}{(n+1)^s} = \Gamma(s) \zeta(s) < \infty$$

for $\sigma > 1$; here $|e^{-(n+1)x} x^{s-1}| = e^{-(n+1)x} x^{s-1}$ by (1.1). Hence the interchange of sum and integral is justified by Fubini's theorem (for the counting measure on $\mathbb{Z}_{\geq 0}$ times Lebesgue measure on $(0, \infty)$). The same substitution $t = (n+1)x$, $dx = dt/(n+1)$, together with

$$\left(\frac{t}{n+1}\right)^{s-1} \frac{1}{n+1} = t^{s-1} (n+1)^{-(s-1)} (n+1)^{-1} = t^{s-1} (n+1)^{-s}$$

(valid since $t, n+1 > 0$ and we use the real logarithm), gives

$$\int_0^\infty \frac{x^{s-1}}{e^x - 1} dx = \sum_{n \geq 0} \int_0^\infty \left(\frac{t}{n+1}\right)^{s-1} e^{-t} \frac{dt}{n+1} = \sum_{n \geq 0} \frac{\Gamma(s)}{(n+1)^s} = \Gamma(s) \sum_{n \geq 1} \frac{1}{n^s} = \Gamma(s) \zeta(s). \quad \blacksquare$$

Remark 1.3.5. Thus $\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1}}{e^x - 1} dx$ for $\operatorname{Re} s > 1$. Riemann's original proof of the analytic continuation starts from this formula and deforms the path of integration into a Hankel contour. We follow instead the approach via the theta function, as in [2], which generalises directly to L -functions of modular forms.

1.4 The Mellin and Fourier transforms

Definition 1.4.1 (Mellin transform)

Let f be a function on $(0, \infty)$. The *Mellin transform* of f is the function

$$g(s) = \int_0^\infty f(t) t^s \frac{dt}{t} = \int_0^\infty f(t) t^{s-1} dt,$$

defined for those s for which the integral converges.

Example 1.4.2. The Mellin transform of $f(t) = e^{-t}$ is $\Gamma(s)$, for $\operatorname{Re} s > 0$ (Proposition 1.1.2).

Example 1.4.3. Let $c > 0$ be a constant. The Mellin transform of $f(t) = e^{-ct}$ is $c^{-s}\Gamma(s)$ for $\operatorname{Re} s > 0$. Indeed, substituting $u = ct$, $dt = du/c$ (and $u \rightarrow 0$, resp. $u \rightarrow \infty$, as $t \rightarrow 0$, resp. $t \rightarrow \infty$),

$$\int_0^\infty e^{-ct} t^{s-1} dt = \int_0^\infty e^{-u} \left(\frac{u}{c}\right)^{s-1} \frac{du}{c} = \frac{1}{c^s} \int_0^\infty e^{-u} u^{s-1} du = c^{-s}\Gamma(s).$$

Definition 1.4.4 (Schwartz space)

Let $\mathcal{S}$ be the vector space of infinitely differentiable functions $f : \mathbb{R} \rightarrow \mathbb{C}$ such that

$$|x|^N f^{(k)}(x) \longrightarrow 0 \quad \text{as } x \rightarrow \pm\infty, \quad \text{for all integers } N, k \geq 0.$$

An example of such a function is $f(x) = e^{-\pi x^2}$: each derivative $f^{(k)}(x)$ is a polynomial times $e^{-\pi x^2}$, and the Gaussian decays faster than any power of $|x|$.

Correction. In my handwritten notes the decay condition $|x|^N f(x) \rightarrow 0$ was only required for f itself. For the theory of the Fourier transform one needs it for all derivatives $f^{(k)}$ as well (this is the standard definition of the Schwartz space); otherwise $\hat{f}$ need not be smooth and rapidly decreasing, and the proof of Proposition 1.4.6 breaks down.

Notes. Two elementary consequences of the definition will be used repeatedly. First, if $f \in \mathcal{S}$ then $f' \in \mathcal{S}$ and $x^k f(x) \in \mathcal{S}$ for every $k \geq 0$ (by the Leibniz rule). Second, every $f \in \mathcal{S}$ is integrable: since $(1+x^2)f(x)$ is continuous and tends to 0 at $\pm\infty$, it is bounded by some C , so $|f(x)| \leq C/(1+x^2)$.

Definition 1.4.5 (Fourier transform)

For $f \in \mathcal{S}$, the *Fourier transform* of f is the function $\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$ given by

$$\hat{f}(y) = \int_{-\infty}^\infty e^{-2\pi ixy} f(x) dx.$$

Proposition 1.4.6

For $f \in \mathcal{S}$ the integral defining $\hat{f}(y)$ converges absolutely for every $y \in \mathbb{R}$. Moreover, $\hat{f} \in \mathcal{S}$.

Proof. Since $|e^{-2\pi ixy} f(x)| = |f(x)|$ and f is integrable, the integral converges absolutely.

Smoothness. For each $k \geq 0$ the function $x^k f(x)$ is integrable. Differentiating under the integral sign (justified by dominated convergence, since $|\partial_y^k (e^{-2\pi ixy} f(x))| = (2\pi |x|)^k |f(x)|$ is integrable

and independent of y) gives

$$\hat{f}^{(k)}(y) = \int_{-\infty}^{\infty} (-2\pi ix)^k f(x) e^{-2\pi ixy} dx = \widehat{(-2\pi ix)^k f}(y). \quad (1.5)$$

So $\hat{f}$ is infinitely differentiable.

Decay. For $g \in \mathcal{S}$, integration by parts gives

$$\hat{g}'(y) = \left[g(x) e^{-2\pi ixy} \right]_{-\infty}^{\infty} + 2\pi iy \int_{-\infty}^{\infty} g(x) e^{-2\pi ixy} dx = 2\pi iy \hat{g}(y),$$

since $g(x) \rightarrow 0$ as $x \rightarrow \pm\infty$. By induction, $(2\pi iy)^N \hat{g}(y) = \widehat{g^{(N)}}(y)$. Applying this to $g = (-2\pi ix)^k f \in \mathcal{S}$ and using (1.5),

$$\left| (2\pi y)^N \hat{f}^{(k)}(y) \right| = \left| \widehat{g^{(N)}}(y) \right| \leq \int_{-\infty}^{\infty} |g^{(N)}(x)| dx < \infty.$$

Thus $y^N \hat{f}^{(k)}(y)$ is bounded for all $N, k \geq 0$. Applying this with $N+1$ in place of N shows $|y|^N |\hat{f}^{(k)}(y)| \leq C/|y| \rightarrow 0$ as $y \rightarrow \pm\infty$. Hence $\hat{f} \in \mathcal{S}$. ■

Proposition 1.4.7

Let $f \in \mathcal{S}$.

- (1) If $a \in \mathbb{R}$ and $g(x) = f(x+a)$, then $\hat{g}(y) = e^{2\pi iay} \hat{f}(y)$.
- (2) If $a \in \mathbb{R}$ and $g(x) = e^{2\pi iax} f(x)$, then $\hat{g}(y) = \hat{f}(y-a)$.
- (3) If $b > 0$ and $g(x) = f(bx)$, then $\hat{g}(y) = \frac{1}{b} \hat{f}\left(\frac{y}{b}\right)$.

Proof. (1) Substituting $t = x+a$, so $x = t-a$ and $dx = dt$,

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-2\pi ixy} f(x+a) dx = \int_{-\infty}^{\infty} e^{-2\pi i(t-a)y} f(t) dt = e^{2\pi iay} \int_{-\infty}^{\infty} e^{-2\pi ity} f(t) dt = e^{2\pi iay} \hat{f}(y).$$

(2) Directly,

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-2\pi ixy} e^{2\pi iax} f(x) dx = \int_{-\infty}^{\infty} e^{-2\pi ix(y-a)} f(x) dx = \hat{f}(y-a).$$

(3) Substituting $t = bx$, so $dx = dt/b$ (and the limits are preserved since $b > 0$),

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-2\pi ixy} f(bx) dx = \frac{1}{b} \int_{-\infty}^{\infty} e^{-2\pi it(y/b)} f(t) dt = \frac{1}{b} \hat{f}\left(\frac{y}{b}\right). \quad \blacksquare$$

Notes. In my handwritten notes the proof of (3) was omitted with a reference to [2, p. 71]; we have written it out. In the last line of the proof of (1) the factor y was missing from the exponent: the correct factor is $e^{2\pi iay}$, as in the statement.

Lemma 1.4.8 (Gaussian integral)

$$\int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1.$$

Proof. Let C denote the integral; clearly $C > 0$. By Fubini's theorem and polar coordinates $(x, y) = (r \cos \varphi, r \sin \varphi)$,

$$C^2 = \int_{-\infty}^{\infty} e^{-\pi x^2} dx \int_{-\infty}^{\infty} e^{-\pi y^2} dy = \iint_{\mathbb{R}^2} e^{-\pi(x^2+y^2)} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\pi r^2} r dr d\varphi.$$

Here $x^2 + y^2 = r^2(\cos^2 \varphi + \sin^2 \varphi) = r^2$ and the Jacobian of polar coordinates is

$$\det \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} = r \cos^2 \varphi + r \sin^2 \varphi = r.$$

With $u = \pi r^2$, $du = 2\pi r dr$, the inner integral is $\frac{1}{2\pi} \int_0^{\infty} e^{-u} du = \frac{1}{2\pi} \Gamma(1) = \frac{1}{2\pi}$. Hence $C^2 = \int_0^{2\pi} \frac{1}{2\pi} d\varphi = 1$, and since $C > 0$, $C = 1$. ■

Proposition 1.4.9

If $f(x) = e^{-\pi x^2}$, then $\hat{f} = f$.

Proof. For all $y \in \mathbb{R}$,

$$\hat{f}(y) = \int_{-\infty}^{\infty} e^{-2\pi ixy} e^{-\pi x^2} dx.$$

Differentiating under the integral sign (justified as in the proof of Proposition 1.4.6),

$$\frac{d}{dy} \hat{f}(y) = \int_{-\infty}^{\infty} (-2\pi i x) e^{-2\pi ixy} e^{-\pi x^2} dx = -2\pi i \int_{-\infty}^{\infty} e^{-2\pi ixy} x e^{-\pi x^2} dx.$$

We compute the last integral by parts with $u = e^{-2\pi ixy}$ and $dv = x e^{-\pi x^2} dx$, so that $du = -2\pi i y e^{-2\pi ixy} dx$ and $v = -\frac{1}{2\pi} e^{-\pi x^2}$:

$$\int_{-\infty}^{\infty} e^{-2\pi ixy} x e^{-\pi x^2} dx = \left[-\frac{e^{-2\pi ixy} e^{-\pi x^2}}{2\pi} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-\frac{1}{2\pi} e^{-\pi x^2} \right) (-2\pi i y) e^{-2\pi ixy} dx = 0 - i y \hat{f}(y).$$

Here the boundary term vanishes because $|e^{-2\pi ixy} e^{-\pi x^2}| = e^{-\pi x^2} \rightarrow 0$ as $x \rightarrow \pm\infty$, and in the remaining integral $(-\frac{1}{2\pi})(-2\pi i y) = i y$, so it equals $-i y \int e^{-2\pi ixy} e^{-\pi x^2} dx = -i y \hat{f}(y)$. Hence $\hat{f}$ satisfies the differential equation

$$\frac{d}{dy} \hat{f}(y) = -2\pi i (-i y \hat{f}(y)) = -2\pi y \hat{f}(y).$$

Then $\frac{d}{dy} (e^{\pi y^2} \hat{f}(y)) = e^{\pi y^2} (2\pi y \hat{f}(y) + \hat{f}'(y)) = 0$, so $\hat{f}(y) = C e^{-\pi y^2}$ for a constant C . Putting $y = 0$ and using Lemma 1.4.8, $C = \hat{f}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1$. Therefore $\hat{f}(y) = e^{-\pi y^2} = f(y)$. ■

Notes. In my handwritten notes the ODE was solved by separation of variables, $d\hat{f}/\hat{f} = -2\pi y dy$, which tacitly divides by $\hat{f}(y)$. The argument with $e^{\pi y^2} \hat{f}(y)$ above avoids this. Also, several integrals in the integration by parts were written with dy instead of dx .

1.5 Poisson summation and the theta function

Theorem 1.5.1 (Poisson summation formula)

If $g \in \mathcal{S}$, then

$$\sum_{m=-\infty}^{\infty} g(m) = \sum_{m=-\infty}^{\infty} \hat{g}(m),$$

and both series converge absolutely.

Proof. Both series converge absolutely, since $g, \hat{g} \in \mathcal{S}$ (Proposition 1.4.6) are bounded by $C/(1+x^2)$. Put

$$h(x) = \sum_{k=-\infty}^{\infty} g(x+k) \quad (x \in \mathbb{R}).$$

For each $j \geq 0$ we have $|g^{(j)}(y)| \leq C_j/(1+y^2)$, so the series $\sum_k g^{(j)}(x+k)$ converges uniformly for $x \in [0, 1]$ (and then on every bounded interval) by the M -test. Hence h is infinitely differentiable. It is periodic of period 1, since

$$h(x+1) = \sum_{k=-\infty}^{\infty} g(x+1+k) = \sum_{j=-\infty}^{\infty} g(x+j) = h(x) \quad (j = k+1).$$

A 1-periodic C^2 function is the sum of its Fourier series, which converges absolutely and uniformly (see [7, Ch. 2–3]):

$$h(x) = \sum_{m=-\infty}^{\infty} c_m e^{2\pi i m x}, \quad c_m = \int_0^1 h(x) e^{-2\pi i m x} dx.$$

Since the series defining h converges uniformly on $[0, 1]$, we may integrate term by term:

$$c_m = \sum_{k=-\infty}^{\infty} \int_0^1 g(x+k) e^{-2\pi i m x} dx.$$

Substitute $u = x+k$, so $du = dx$, $x = u-k$, and u runs from k to $k+1$. Since $e^{2\pi i m k} = 1$,

$$c_m = \sum_{k \in \mathbb{Z}} \int_k^{k+1} g(u) e^{-2\pi i m (u-k)} du = \sum_{k \in \mathbb{Z}} \int_k^{k+1} g(u) e^{-2\pi i m u} du = \int_{-\infty}^{\infty} g(u) e^{-2\pi i m u} du = \hat{g}(m).$$

Evaluating the Fourier series at $x = 0$ we obtain

$$\sum_{k=-\infty}^{\infty} g(k) = h(0) = \sum_{m=-\infty}^{\infty} c_m = \sum_{m=-\infty}^{\infty} \hat{g}(m). \quad \blacksquare$$

Notes. The input quoted from Fourier analysis is the following: if h is 1-periodic and of class C^2 , then integrating by parts twice gives $|c_m| \leq \frac{1}{4\pi^2 m^2} \int_0^1 |h''|$ for $m \neq 0$, so the Fourier series converges absolutely and uniformly; and a continuous periodic function all of whose Fourier coefficients vanish is identically zero (by Fejér's theorem), so the sum of the Fourier series is h .

Definition 1.5.2 (Theta function)

For $t > 0$ we put

$$\theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi t n^2} = 1 + 2 \sum_{n \geq 1} e^{-\pi t n^2}.$$

Proposition 1.5.3

For all $t > 0$,

$$\theta(t) = \frac{1}{\sqrt{t}} \theta\left(\frac{1}{t}\right).$$

Proof. Fix $t > 0$ and apply Theorem 1.5.1 to $g(x) = e^{-\pi t x^2} \in \mathcal{S}$:

$$\theta(t) = \sum_{m=-\infty}^{\infty} e^{-\pi t m^2} = \sum_{m=-\infty}^{\infty} g(m) = \sum_{m=-\infty}^{\infty} \hat{g}(m).$$

(That $g \in \mathcal{S}$: $g(x) = f(\sqrt{t}x)$ with $f \in \mathcal{S}$, and $g^{(k)}(x) = t^{k/2} f^{(k)}(\sqrt{t}x)$, so the decay conditions for g follow from those for f .) Let $f(x) = e^{-\pi x^2}$, so that $\hat{f} = f$ by Proposition 1.4.9. Since $(\sqrt{t}x)^2 = tx^2$, we have $g(x) = e^{-\pi(\sqrt{t}x)^2} = f(\sqrt{t}x)$, and Proposition 1.4.7(3) with $b = \sqrt{t} > 0$ gives

$$\hat{g}(y) = \frac{1}{\sqrt{t}} \hat{f}\left(\frac{y}{\sqrt{t}}\right) = \frac{1}{\sqrt{t}} e^{-\pi y^2/t}.$$

Therefore

$$\theta(t) = \frac{1}{\sqrt{t}} \sum_{m=-\infty}^{\infty} e^{-\pi m^2/t} = \frac{1}{\sqrt{t}} \theta\left(\frac{1}{t}\right). \quad \blacksquare$$

Lemma 1.5.4 (i) For $t \geq 1$ we have $0 < \theta(t) - 1 \leq 4e^{-\pi t}$.

(ii) For $0 < t \leq 1$ we have $0 < \theta(t) - t^{-1/2} \leq 4t^{-1/2}e^{-\pi/t}$.

(iii) There is $C > 0$ such that $|\theta(t) - t^{-1/2}| < e^{-C/t}$ for all sufficiently small $t > 0$; one can take $C = \pi - 1$.

Proof. (i) The sum $\theta(t) - 1 = 2 \sum_{n \geq 1} e^{-\pi t n^2}$ has positive terms, so it is > 0 . For the upper bound, for $n \geq 1$ we have $n^2 \geq n$, hence $e^{-\pi t n^2} \leq e^{-\pi t n} = r^n$ with $r = e^{-\pi t} \in (0, 1)$. By the geometric series $\sum_{n \geq 1} r^n = \frac{r}{1-r}$,

$$\theta(t) - 1 = 2 \sum_{n \geq 1} e^{-\pi t n^2} \leq 2 \sum_{n \geq 1} e^{-\pi t n} = \frac{2e^{-\pi t}}{1 - e^{-\pi t}} \leq 4e^{-\pi t}.$$

The last step uses $1 - e^{-\pi t} \geq \frac{1}{2}$, i.e. $e^{-\pi t} \leq \frac{1}{2}$: for $t \geq 1$, $e^{-\pi t} \leq e^{-\pi} \approx 0.0432 < \frac{1}{2}$.

(ii) For $0 < t \leq 1$ we have $1/t \geq 1$. By Proposition 1.5.3, $\theta(t) = t^{-1/2} \theta(1/t)$, so

$$\theta(t) - t^{-1/2} = t^{-1/2} \theta(1/t) - t^{-1/2} = t^{-1/2} (\theta(1/t) - 1),$$

and (i) applied at $1/t \geq 1$ gives $0 < \theta(1/t) - 1 \leq 4e^{-\pi/t}$; multiplying by $t^{-1/2} > 0$,

$$0 < \theta(t) - t^{-1/2} \leq 4t^{-1/2} e^{-\pi/t}.$$

(iii) As $t \rightarrow 0^+$ we have $t^{1/2} e^{1/t} \rightarrow \infty$ (the exponential grows faster), so $4t^{-1/2} < e^{1/t}$ for all sufficiently small $t > 0$. For such t , (ii) gives $|\theta(t) - t^{-1/2}| < e^{1/t} e^{-\pi/t} = e^{-(\pi-1)/t}$. $\blacksquare$

Notes. Part (iii) is the statement in my handwritten notes, where it was proved by expanding $2t^{-1/2} \sum_{n \geq 1} e^{-\pi n^2/t}$ and comparing with a geometric series, using $t^{-1/2} < \frac{1}{4} e^{1/t}$ and $e^{-2\pi/t} < \frac{1}{2}$ for small t . The bounds (i) and (ii), which hold on the whole ranges $t \geq 1$ and $0 < t \leq 1$,

are what is really needed below: they show that $\theta(t) - 1$ decays exponentially as $t \rightarrow \infty$ (in my handwritten notes it was only observed that $\theta(t) - 1 \rightarrow 0$, which by itself does not imply convergence of the integrals) and that $\theta(t) - t^{-1/2}$ decays like $e^{-\pi/t}$ as $t \rightarrow 0^+$.

1.6 Analytic continuation and functional equation of ζ

The idea

We want to take the Mellin transform of θ , i.e. to study $\int_0^\infty \theta(t) t^s \frac{dt}{t}$. This integral diverges for every s : as $t \rightarrow \infty$ we have $\theta(t) \rightarrow 1$, so the part over $[1, \infty)$ diverges when $\operatorname{Re} s \geq 0$; and as $t \rightarrow 0^+$ we have $\theta(t) \approx t^{-1/2}$, so the part over $(0, 1]$ diverges when $\operatorname{Re} s \leq \frac{1}{2}$. We therefore subtract the leading behaviour at both ends. We also replace s by $s/2$, since otherwise we would end up with $\zeta(2s)$ instead of $\zeta(s)$.

Definition 1.6.1

For $s \in \mathbb{C}$ put

$$\phi(s) = \underbrace{\int_0^1 (\theta(t) - t^{-1/2}) t^{s/2} \frac{dt}{t}}_{(1)} + \underbrace{\int_1^\infty (\theta(t) - 1) t^{s/2} \frac{dt}{t}}_{(2)}.$$

Lemma 1.6.2

The integrals (1) and (2) converge absolutely for every $s \in \mathbb{C}$, and ϕ is an entire function.

Proof. Let $B > 0$ and suppose $|\operatorname{Re} s| \leq B$. Then $|t^{s/2-1}| = t^{\sigma/2-1}$, which is $\leq t^{-B/2-1}$ for $0 < t \leq 1$ and $\leq t^{B/2-1}$ for $t \geq 1$. By Lemma 1.5.4(i),(ii) the integrands of (1) and (2) are bounded by

$$4t^{-1/2} e^{-\pi/t} \cdot t^{-B/2-1} = 4t^{-B/2-3/2} e^{-\pi/t} \quad (0 < t \leq 1) \quad \text{and} \quad 4e^{-\pi t} \cdot t^{B/2-1} \quad (t \geq 1),$$

which are independent of s . They are integrable: the second because $\int_1^\infty e^{-\pi t} t^{B/2-1} dt \leq \pi^{-B/2} \Gamma(B/2)$ (Example 1.4.3); the first because the substitution $t = 1/u$, $dt = -du/u^2$, turns it into

$$\begin{aligned} \int_0^1 t^{-B/2-3/2} e^{-\pi/t} dt &= \int_1^\infty u^{B/2+3/2} e^{-\pi u} \frac{du}{u^2} \\ &= \int_1^\infty u^{B/2-1/2} e^{-\pi u} du \leq \pi^{-(B+1)/2} \Gamma\left(\frac{B+1}{2}\right) < \infty. \end{aligned}$$

As in the proof of Proposition 1.1.5, the truncated integrals $\int_\varepsilon^1$ and $\int_1^{1/\varepsilon}$ are holomorphic in s by Lemma 1.1.4 (the integrands are continuous in (s, t) since θ is continuous on $(0, \infty)$, being a locally uniformly convergent series of continuous functions), and these bounds show that they converge uniformly on $\{|\operatorname{Re} s| \leq B\}$ as $\varepsilon \rightarrow 0^+$. By Lemma 1.1.3, ϕ is holomorphic on $\{|\operatorname{Re} s| < B\}$ for every B , i.e. entire. ■

Proposition 1.6.3

For $\operatorname{Re} s > 1$,

$$\frac{1}{2} \phi(s) = \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) + \frac{1}{s} + \frac{1}{1-s}.$$

Proof. Let $\sigma = \operatorname{Re} s > 1$.

Two elementary integrals. For $\operatorname{Re} w > 0$, $|t^w| = t^{\operatorname{Re} w} \rightarrow 0$ as $t \rightarrow 0^+$, so

$$\int_0^1 t^w \frac{dt}{t} = \int_0^1 t^{w-1} dt = \left[\frac{t^w}{w} \right]_0^1 = \frac{1}{w}.$$

With $w = s/2$ ($\operatorname{Re} w = \sigma/2 > 0$) and $w = (s-1)/2$ ($\operatorname{Re} w = (\sigma-1)/2 > 0$, since $\sigma > 1$) we get

$$\int_0^1 t^{s/2} \frac{dt}{t} = \frac{2}{s}, \quad \int_0^1 t^{-1/2} t^{s/2} \frac{dt}{t} = \int_0^1 t^{(s-1)/2} \frac{dt}{t} = \frac{2}{s-1}.$$

Integral (1). Both $\int_0^1 t^{-1/2} t^{s/2} \frac{dt}{t}$ (just computed) and $\int_0^1 \theta(t) t^{s/2} \frac{dt}{t}$ (it differs from the former by the convergent integral (1), Lemma 1.6.2) converge, so we may split (1). Writing $\theta(t) = 1 + 2 \sum_{n \geq 1} e^{-\pi n^2 t}$,

$$(1) = \int_0^1 \theta(t) t^{s/2} \frac{dt}{t} - \frac{2}{s-1} = \frac{2}{s} + 2 \sum_{n \geq 1} \int_0^1 e^{-\pi n^2 t} t^{s/2} \frac{dt}{t} - \frac{2}{s-1}.$$

Integral (2). Since $\theta(t) - 1 = 2 \sum_{n \geq 1} e^{-\pi n^2 t}$,

$$(2) = 2 \sum_{n \geq 1} \int_1^\infty e^{-\pi n^2 t} t^{s/2} \frac{dt}{t}.$$

Sum. Adding (1) and (2), the two integrals over $(0, 1]$ and $[1, \infty)$ of the same function combine to one integral over $(0, \infty)$, and $-\frac{2}{s-1} = \frac{2}{1-s}$:

$$\phi(s) = 2 \sum_{n \geq 1} \left(\int_0^1 + \int_1^\infty \right) e^{-\pi n^2 t} t^{s/2} \frac{dt}{t} + \frac{2}{s} + \frac{2}{1-s} = 2 \sum_{n \geq 1} \int_0^\infty e^{-\pi n^2 t} t^{s/2-1} dt + \frac{2}{s} + \frac{2}{1-s}.$$

By Example 1.4.3 with $c = \pi n^2 > 0$ and $s/2$ in place of s (allowed, as $\operatorname{Re} \frac{s}{2} > 0$), $\int_0^\infty e^{-\pi n^2 t} t^{s/2-1} dt = (\pi n^2)^{-s/2} \Gamma(\frac{s}{2})$, so

$$\phi(s) = 2 \sum_{n \geq 1} (\pi n^2)^{-s/2} \Gamma\left(\frac{s}{2}\right) + \frac{2}{s} + \frac{2}{1-s}.$$

The interchanges of $\sum$ and $\int$ are justified by Fubini–Tonelli: the sum of the integrals of the absolute values is $\sum_{n \geq 1} (\pi n^2)^{-\sigma/2} \Gamma(\sigma/2) = \pi^{-\sigma/2} \Gamma(\sigma/2) \zeta(\sigma) < \infty$ for $\sigma > 1$. Dividing by 2 and using $(\pi n^2)^{-s/2} = e^{-\frac{s}{2}(\log \pi + 2 \log n)} = \pi^{-s/2} n^{-s}$ (real logarithms of positive numbers),

$$\frac{1}{2} \phi(s) = \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \sum_{n \geq 1} n^{-s} + \frac{1}{s} + \frac{1}{1-s} = \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) + \frac{1}{s} + \frac{1}{1-s}. \quad \blacksquare$$

Solving for ζ , we obtain for $\operatorname{Re} s > 1$:

$$\zeta(s) = \frac{\pi^{s/2}}{\Gamma(s/2)} \left(\frac{1}{2} \phi(s) - \frac{1}{s} - \frac{1}{1-s} \right). \quad (1.6)$$

Theorem 1.6.4

The Riemann zeta function $\zeta(s)$, $\operatorname{Re} s > 1$, extends analytically to the whole complex plane except for a simple pole at $s = 1$, with residue 1.

Proof. The functions $\pi^{s/2} = e^{(s/2)\log \pi}$, $1/\Gamma(s/2)$ (Corollary 1.2.4) and $\phi(s)$ (Lemma 1.6.2) are entire. Hence the right-hand side of (1.6) is meromorphic on $\mathbb{C}$, and it gives the analytic continuation of ζ . Its only possible poles are at $s = 0$ and $s = 1$.

The point $s = 0$. Write the right-hand side of (1.6) as

$$\frac{1}{2} \frac{\pi^{s/2}}{\Gamma(s/2)} \phi(s) - \frac{\pi^{s/2}}{s \Gamma(s/2)} - \frac{\pi^{s/2}}{(1-s) \Gamma(s/2)}.$$

The first and third terms are holomorphic at $s = 0$. By Corollary 1.2.4, $g(s) = 1/\Gamma(s/2)$ has a simple zero at $s = 0$, so $g(s) = s h(s)$ with h holomorphic near 0 and $h(0) \neq 0$. Then the middle term equals $-\pi^{s/2} h(s)$, which is holomorphic at $s = 0$. Thus $s = 0$ is a removable singularity.

The point $s = 1$. The terms $\frac{\pi^{s/2}}{\Gamma(s/2)} (\frac{1}{2}\phi(s) - \frac{1}{s})$ are holomorphic at $s = 1$, while the last term $-\frac{\pi^{s/2}}{\Gamma(s/2)} \cdot \frac{1}{1-s}$ has at most a simple pole there. Using $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ from (1.4),

$$\text{Res}_{s=1} \zeta(s) = \lim_{s \rightarrow 1} (s-1) \zeta(s) = \lim_{s \rightarrow 1} \frac{\pi^{s/2}}{\Gamma(s/2)} = \frac{\pi^{1/2}}{\Gamma(1/2)} = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1.$$

Since the residue is nonzero, $s = 1$ is indeed a simple pole. ■

Correction. In my handwritten notes it is stated that $1/\Gamma(s/2)$ has simple zeros at $s = 0, -1, -2, \dots$. Since $1/\Gamma(w)$ vanishes exactly at $w \in \mathbb{Z}_{\leq 0}$, the zeros of $1/\Gamma(s/2)$ are at $s = 0, -2, -4, \dots$ (the even non-positive integers). Only the zero at $s = 0$ is used in the proof.

Remark 1.6.5 (Some special values). The proof gives more information.

- (a) *The value $\zeta(0)$.* By Proposition 1.1.6 (continued to $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$), $\Gamma(\frac{s}{2}) = \Gamma(\frac{s}{2} + 1)/\frac{s}{2}$, so $s \Gamma(\frac{s}{2}) = 2\Gamma(1 + \frac{s}{2}) \rightarrow 2\Gamma(1) = 2$ as $s \rightarrow 0$. Hence $h(0) = \lim_{s \rightarrow 0} \frac{1}{s \Gamma(s/2)} = \frac{1}{2}$. In the decomposition above, the first term $\frac{1}{2} \pi^{s/2} \phi(s)/\Gamma(s/2)$ and the third term $-\pi^{s/2}/((1-s)\Gamma(s/2))$ both tend to 0 at $s = 0$, because $1/\Gamma(s/2) \rightarrow 0$ while $\pi^{s/2} \phi(s)$ and $\pi^{s/2}/(1-s)$ stay bounded. Hence $\zeta(0) = 0 - \pi^0 h(0) - 0 = -\frac{1}{2}$.
- (b) *Trivial zeros.* For $k \geq 1$, $1/\Gamma(s/2)$ vanishes at $s = -2k$, while the bracket in (1.6) is finite there. Hence $\zeta(-2k) = 0$ for $k = 1, 2, \dots$.

Theorem 1.6.6 (Functional equation)

For all $s \in \mathbb{C} \setminus \{0, 1\}$,

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{-(1-s)/2} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s).$$

Proof. Put

$$\Lambda(s) = \frac{1}{2} \phi(s) - \frac{1}{s} - \frac{1}{1-s}.$$

By Proposition 1.6.3, $\Lambda(s) = \pi^{-s/2} \Gamma(s/2) \zeta(s)$ for $\text{Re } s > 1$. Both sides are meromorphic on $\mathbb{C}$, so by the identity theorem this is an identity of meromorphic functions on all of $\mathbb{C}$. Replacing s by $1-s$ in $-\frac{1}{s} - \frac{1}{1-s}$ gives $-\frac{1}{1-s} - \frac{1}{1-(1-s)} = -\frac{1}{1-s} - \frac{1}{s}$, the same expression; so this part is invariant under $s \mapsto 1-s$, and

$$\Lambda(1-s) = \frac{1}{2} \phi(1-s) - \frac{1}{s} - \frac{1}{1-s}.$$

Hence $\Lambda(s) = \Lambda(1-s)$ is equivalent to

$$\phi(s) = \phi(1-s).$$

Recall

$$\phi(s) = \int_0^1 (\theta(t) - t^{-1/2}) t^{s/2} \frac{dt}{t} + \int_1^\infty (\theta(t) - 1) t^{s/2} \frac{dt}{t}.$$

Substitute $t = 1/u$. Then $\frac{dt}{t} = -\frac{du}{u}$; as $t \rightarrow 0^+$ we have $u \rightarrow \infty$, as $t \rightarrow \infty$ we have $u \rightarrow 0^+$, and $t = 1$ corresponds to $u = 1$. Hence

$$\phi(s) = \int_1^\infty (\theta(1/u) - u^{1/2}) u^{-s/2} \frac{du}{u} + \int_0^1 (\theta(1/u) - 1) u^{-s/2} \frac{du}{u}.$$

(the minus sign from $dt/t = -du/u$ is absorbed by swapping the limits of integration, and $(1/u)^{-1/2} = u^{1/2}$, $(1/u)^{s/2} = u^{-s/2}$). By Proposition 1.5.3 applied at $t = u$, $\theta(u) = u^{-1/2}\theta(1/u)$, i.e. $\theta(1/u) = \sqrt{u}\theta(u)$. In the computation below we use $\sqrt{u} \cdot u^{-s/2} = u^{(1-s)/2}$ and $1 \cdot u^{-s/2} = u^{-1/2} \cdot u^{(1-s)/2}$. Therefore

$$\begin{aligned} \phi(s) &= \int_1^\infty (\sqrt{u}\theta(u) - \sqrt{u}) u^{-s/2} \frac{du}{u} + \int_0^1 (\sqrt{u}\theta(u) - 1) u^{-s/2} \frac{du}{u} \\ &= \int_1^\infty (\theta(u) - 1) u^{(1-s)/2} \frac{du}{u} + \int_0^1 (\theta(u) - u^{-1/2}) u^{(1-s)/2} \frac{du}{u} = \phi(1-s). \end{aligned}$$

(in the last step the two integrals are exactly (2) and (1) of the definition of ϕ , with s replaced by $1-s$). Hence $\Lambda(s) = \Lambda(1-s)$, and since $\Lambda(1-s) = \pi^{-(1-s)/2} \Gamma(\frac{1-s}{2}) \zeta(1-s)$ (the identity $\Lambda(w) = \pi^{-w/2} \Gamma(w/2) \zeta(w)$ at $w = 1-s$), this is the claimed functional equation. ■

Correction. In the last line of my handwritten notes the functional equation was written as $\pi^{-1/2} \Gamma(\frac{s}{2}) \zeta(s) = \pi^{-1/2} \Gamma(\frac{1-s}{2}) \zeta(s-1)$. The correct form, as in the statement of the theorem, has the factors $\pi^{-s/2}$ and $\pi^{-(1-s)/2}$, and $\zeta(1-s)$ instead of $\zeta(s-1)$. Similarly, in the definition of Λ the factor $\pi^{-1/2}$ should read $\pi^{-s/2}$.

Notes. At the end of my handwritten Note 1 there is a short reading plan: Koblitz [2], Chapter I up to the addition law on elliptic curves, and Chapter III, Section 1. These are the topics of the next two chapters: elliptic curves, and the modular group and its congruence subgroups.

Chapter 2

Elliptic Curves and Elliptic Functions

In this chapter we introduce elliptic curves over a field, and then study elliptic curves over $\mathbb{C}$ through *elliptic functions*, i.e. meromorphic functions on $\mathbb{C}$ which are periodic with respect to a lattice. The main result is that the Weierstrass $\wp$ -function of a lattice L gives an analytic bijection from the torus $\mathbb{C}/L$ onto the elliptic curve $y^2 = 4x^3 - g_2(L)x - g_3(L)$ in $\mathbb{P}^2(\mathbb{C})$. The coefficients $g_2(L) = 60G_4(L)$ and $g_3(L) = 140G_6(L)$ are Eisenstein series, which will be our first examples of modular forms. We follow [2, Ch. I, §§1–6].

2.1 Elliptic curves over a field

To motivate the definition: the *locus* of a polynomial equation is the set of all points satisfying it. For example, the congruent number problem leads to the locus of points $P = (x, y)$ satisfying $y^2 = x^3 - n^2x$.

Definition 2.1.1 (Elliptic curve, first version)

Let $K \subseteq K'$ be fields with $\text{char}(K) \neq 2$, and let

$$f(x) = x^3 + ax^2 + bx + c \in K[x]$$

be a cubic polynomial with coefficients in K which has distinct roots (in an algebraic closure of K). The solutions $(x, y) \in K' \times K'$ of

$$y^2 = f(x) \tag{2.1}$$

are called the K' -points of the elliptic curve defined by (2.1).

Notes. Recall that $\text{char}(K) = p > 0$ means that p is the smallest positive integer with $\underbrace{1 + \cdots + 1}_p = 0$ in K , and $\text{char}(K) = 0$ if no such p exists. A polynomial f has a *multiple root* x_0 if $f(x_0) = 0$ and $f'(x_0) = 0$; “distinct roots” means that f has no multiple root. The condition $\text{char}(K) \neq 2$ is needed because we will divide by 2 (e.g. in the proof of Proposition 2.1.4). In my handwritten notes the cubic was written with a leading coefficient; for $\text{char } K \neq 2, 3$ one can always reduce to the form above, and later we will also allow the leading coefficient 4, as in $y^2 = 4x^3 - g_2x - g_3$.

Example 2.1.2. Let p be an odd prime and $K = K' = \mathbb{F}_p$, so $\text{char}(\mathbb{F}_p) = p \neq 2$. Consider

$$f(x) = x^3 - n^2x \in \mathbb{F}_p[x], \quad n \in \mathbb{Z}.$$

Since $f(x) = x(x - n)(x + n)$, the roots of f are $x = 0, n, -n$. These are distinct in $\mathbb{F}_p$ if and only if $p \nmid 2n$, i.e. (as p is odd) if and only if $p \nmid n$. Indeed:

- if $p = 2$ (excluded here, but instructive), then $2 \equiv 0$, so $n \equiv -n$ and n is a multiple root;
- if $p \mid n$, then $n \equiv 0 \pmod{p}$, so all three roots are 0, and 0 is a (triple) root.

Hence $y^2 = x^3 - n^2x$ defines an elliptic curve over $\mathbb{F}_p$ whenever $p \nmid 2n$.

Definition 2.1.3 (Smooth point)

Let $\text{char}(K') \neq 2$, and let C be a curve defined by $F(x, y) = 0$, $F \in K'[x, y]$. A point $(x_0, y_0) \in K' \times K'$ of C is called *smooth* if the partial derivatives

$$\frac{\partial F}{\partial x}(x_0, y_0) \quad \text{and} \quad \frac{\partial F}{\partial y}(x_0, y_0)$$

are not both zero. Otherwise (x_0, y_0) is called *non-smooth* (or *singular*).

For the curves of Definition 2.1.1 we take $F(x, y) = y^2 - f(x)$, so that $\frac{\partial F}{\partial x} = -f'(x)$ and $\frac{\partial F}{\partial y} = 2y$.

Proposition 2.1.4

Let $\text{char}(K') \neq 2$, $f \in K'[x]$ a cubic polynomial, and $C : F(x, y) = y^2 - f(x) = 0$. Then C has a non-smooth point $(x_0, y_0) \in K' \times K'$ if and only if f has a multiple root $x_0 \in K'$.

Proof. ($\Rightarrow$) Suppose $(x_0, y_0) \in K' \times K'$ is a non-smooth point of C . Then

$$\frac{\partial F}{\partial x} \Big|_{(x_0, y_0)} = -f'(x_0) = 0 \quad \text{and} \quad \frac{\partial F}{\partial y} \Big|_{(x_0, y_0)} = 2y_0 = 0.$$

Since $\text{char}(K') \neq 2$, the element 2 is invertible in K' , so $2y_0 = 0$ implies $y_0 = 0$. As (x_0, y_0) lies on C , we get

$$0 = y_0^2 - f(x_0) = 0 - f(x_0) \implies f(x_0) = 0.$$

Thus $f(x_0) = 0$ and $f'(x_0) = 0$, i.e. x_0 is a multiple root of f .

($\Leftarrow$) Suppose $x_0 \in K'$ is a multiple root of f , so $f(x_0) = 0$ and $f'(x_0) = 0$. Put $y_0 = 0$. Then $y_0^2 - f(x_0) = 0 - 0 = 0$, so $(x_0, y_0) \in C$. Moreover

$$\frac{\partial F}{\partial y} \Big|_{(x_0, y_0)} = 2y_0 = 0 \quad \text{and} \quad \frac{\partial F}{\partial x} \Big|_{(x_0, y_0)} = -f'(x_0) = 0.$$

Hence $(x_0, 0)$ is a non-smooth point of C . ■

Definition 2.1.5 (Elliptic curve, second version)

Let $\text{char}(K) \neq 2$ and let $f \in K[x]$ be a cubic polynomial. The curve $C : F(x, y) = y^2 - f(x) = 0$ is called an *elliptic curve* if it is smooth at all of its points (over an algebraic closure of K), i.e. if $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ never vanish simultaneously on C . By Proposition 2.1.4, this is the same as requiring that f has distinct roots.

Homogenisation and the point at infinity

The *total degree* of a monomial $x^i y^j$ is $i + j$, and the total degree of a polynomial is the maximum of the total degrees of its monomials (with nonzero coefficient).

Definition 2.1.6 (Homogenisation)

Let $F(x, y) = \sum_{i,j} c_{ij} x^i y^j$ have total degree n . The *homogenisation* of F is the polynomial

$$\tilde{F}(x, y, z) = \sum_{i,j} c_{ij} x^i y^j z^{n-i-j},$$

obtained by multiplying each monomial $x^i y^j$ of F by z^{n-i-j} . Every monomial of $\tilde{F}$ has total degree exactly n .

Example 2.1.7. Let $F(x, y) = y^2 - (x^3 - n^2x) = y^2 - x^3 + n^2x$. The monomials and their total degrees are

$$\underbrace{y^2}_{\text{degree 2}}, \quad \underbrace{x^3}_{\text{degree 3}}, \quad \underbrace{n^2x}_{\text{degree 1}},$$

multiply by $z^{3-0-2}=z$ multiply by $z^{3-3-0}=1$ multiply by $z^{3-1-0}=z^2$

so the total degree is $n = 3$, and

$$\tilde{F}(x, y, z) = y^2z - x^3 + n^2xz^2.$$

Note that $F(x, y) = \tilde{F}(x, y, 1)$.

We are interested in the points (x, y, z) with $\tilde{F}(x, y, z) = 0$.

Proposition 2.1.8

Let K be a field, $F \in K[x, y]$ of total degree n , and $\tilde{F}$ its homogenisation.

- (1) For all $\lambda \in K$: $\tilde{F}(\lambda x, \lambda y, \lambda z) = \lambda^n \tilde{F}(x, y, z)$.
- (2) For all $\lambda \in K$ with $\lambda \neq 0$: $\tilde{F}(\lambda x, \lambda y, \lambda z) = 0 \iff \tilde{F}(x, y, z) = 0$.
- (3) For $z \neq 0$: $\tilde{F}(x, y, z) = 0 \iff F\left(\frac{x}{z}, \frac{y}{z}\right) = 0$.

Proof. Write $\tilde{F}(x, y, z) = \sum c_{ij} x^i y^j z^{n-i-j}$.

(1) Let $\lambda \in K$ be arbitrary. Then

$$\begin{aligned} \lambda^n \tilde{F}(x, y, z) &= \lambda^n \sum c_{ij} x^i y^j z^{n-i-j} = \sum c_{ij} \lambda^{i+j+(n-i-j)} x^i y^j z^{n-i-j} \\ &= \sum c_{ij} (\lambda^i x^i) (\lambda^j y^j) (\lambda^{n-i-j} z^{n-i-j}) = \sum c_{ij} (\lambda x)^i (\lambda y)^j (\lambda z)^{n-i-j} = \tilde{F}(\lambda x, \lambda y, \lambda z). \end{aligned}$$

(2) Let $\lambda \neq 0$. Then $\lambda^n \neq 0$, since a field has no zero divisors. By (1), $\tilde{F}(\lambda x, \lambda y, \lambda z) = \lambda^n \tilde{F}(x, y, z)$, and a product $\lambda^n \cdot w$ with $\lambda^n \neq 0$ vanishes if and only if $w = 0$.

(3) Let $z \neq 0$. Then

$$z^n F\left(\frac{x}{z}, \frac{y}{z}\right) = z^n \sum c_{ij} \frac{x^i}{z^i} \frac{y^j}{z^j} = \sum c_{ij} x^i y^j z^{n-i-j} = \tilde{F}(x, y, z).$$

Since $z^n \neq 0$, we get $\tilde{F}(x, y, z) = 0$ if and only if $F\left(\frac{x}{z}, \frac{y}{z}\right) = 0$. ■

Correction. In my handwritten notes the proof of (3) was left blank, and (2) was marked “similar”. Note also that the assumption $\text{char}(K) \neq 2$ in the statement is not needed for this proposition.

Definition 2.1.9 (Projective plane)

Let K be a field. The *projective plane* $\mathbb{P}^2(K)$ is the set of all one-dimensional vector subspaces of K^3 . Equivalently, it is the set of equivalence classes of $K^3 \setminus \{(0, 0, 0)\}$ under

$$(x, y, z) \sim (x', y', z') \iff (x', y', z') = (\lambda x, \lambda y, \lambda z) \text{ for some } \lambda \in K^\times.$$

The class of (x, y, z) is written $[x : y : z]$. More generally, for a vector space V of dimension $n+1$, the projective space $\mathbb{P}(V)$ is the set of one-dimensional subspaces of V ; it has dimension n .

Correction. In my handwritten notes $\mathbb{P}_{\mathbb{C}}^2$ was described as the set of “projective lines” in $\mathbb{C}^3$ with $(x, y, z) = (\lambda x, \lambda y, \lambda z)$ for some $\lambda \in \mathbb{C}$. The correct description is as above: the points of $\mathbb{P}^2$ are the lines *through the origin* in K^3 ; the point $(0, 0, 0)$ is excluded, and λ must be nonzero.

By Proposition 2.1.8(2), the condition $\tilde{F}(x, y, z) = 0$ depends only on the class $[x : y : z] \in \mathbb{P}^2(K)$, so $\tilde{F} = 0$ defines a subset of $\mathbb{P}^2(K)$. By Proposition 2.1.8(3), its points with $z \neq 0$, i.e. $[x : y : 1]$ after rescaling, correspond exactly to the solutions (x, y) of $F(x, y) = 0$. For an elliptic curve $\tilde{F} = y^2z - f_3(x, z)$, the points with $z = 0$ satisfy $x^3 = 0$, hence $x = 0$, and we obtain the single point at infinity $O = [0 : 1 : 0]$. Geometrically, it is the common point at which all the vertical lines $x = \text{const}$ meet.

2.2 Lattices and doubly periodic functions

Definition 2.2.1 (Lattice)

Let $\omega_1, \omega_2 \in \mathbb{C}$ be linearly independent over $\mathbb{R}$, i.e. neither is a real multiple of the other (they do not lie on the same line through the origin). The *lattice* generated by ω_1, ω_2 is

$$L = \{m\omega_1 + n\omega_2 \mid m, n \in \mathbb{Z}\}.$$

It is an additive subgroup of $\mathbb{C}$.

Example 2.2.2. For $\omega_1 = 1$ and $\omega_2 = i$ we obtain $L = \{n + mi \mid m, n \in \mathbb{Z}\} = \mathbb{Z}[i]$, the square grid in the complex plane, known as the *Gaussian integers*.

Definition 2.2.3 (Fundamental parallelogram)

For a basis ω_1, ω_2 of L , the *fundamental parallelogram* is

$$\Pi = \{a\omega_1 + b\omega_2 \mid 0 \leq a \leq 1, 0 \leq b \leq 1\}.$$

We also use the half-open parallelogram $\Pi' = \{a\omega_1 + b\omega_2 \mid 0 \leq a < 1, 0 \leq b < 1\}$, and for $\alpha \in \mathbb{C}$ the translates $\alpha + \Pi = \{\alpha + z \mid z \in \Pi\}$ and $\alpha + \Pi'$.

Proposition 2.2.4

Every $x \in \mathbb{C}$ can be written as $x = \ell + p$ with $\ell \in L$ and $p \in \Pi$. If we require $p \in \Pi'$, this representation is unique. In particular, the representation with $p \in \Pi$ is unique when $p \notin \partial\Pi$.

Proof. Existence. Since ω_1, ω_2 are linearly independent over $\mathbb{R}$ and $\dim_{\mathbb{R}} \mathbb{C} = 2$, they form an $\mathbb{R}$ -basis of $\mathbb{C}$. Hence $x = u\omega_1 + v\omega_2$ for some $u, v \in \mathbb{R}$. Choose $m = \lfloor u \rfloor$ and $n = \lfloor v \rfloor \in \mathbb{Z}$. Then

$$x = u\omega_1 + v\omega_2 + (m\omega_1 + n\omega_2) - (m\omega_1 + n\omega_2) = \underbrace{m\omega_1 + n\omega_2}_{=\ell \in L} + \underbrace{(u-m)\omega_1 + (v-n)\omega_2}_{=p}.$$

Since every real number t satisfies $t - 1 < \lfloor t \rfloor \leq t$, we get $0 \leq u - m < 1$ and $0 \leq v - n < 1$. Thus $p \in \Pi' \subseteq \Pi$.

Uniqueness. Suppose $x = \ell_1 + p_1 = \ell_2 + p_2$ with $\ell_1, \ell_2 \in L$ and $p_1, p_2 \in \Pi'$. Then $\ell_1 - \ell_2 = p_2 - p_1$. Write $\ell_1 - \ell_2 = m\omega_1 + n\omega_2$ with $m, n \in \mathbb{Z}$, and $p_k = a_k\omega_1 + b_k\omega_2$ with $0 \leq a_k, b_k < 1$ ($k = 1, 2$). Then

$$m\omega_1 + n\omega_2 = (a_2 - a_1)\omega_1 + (b_2 - b_1)\omega_2,$$

and by linear independence $m = a_2 - a_1$ and $n = b_2 - b_1$. Since $0 \leq a_1, a_2 < 1$ we have $-1 < a_2 - a_1 < 1$, and the only integer in $(-1, 1)$ is 0; so $m = 0$, and similarly $n = 0$. Hence $p_1 = p_2$ and $\ell_1 = \ell_2$.

Finally, if $p \in \Pi \setminus \partial\Pi$, then $p = a\omega_1 + b\omega_2$ with $0 < a, b < 1$, so $p \in \Pi'$, and any other representation $x = \ell' + p'$ with $p' \in \Pi$ and $p' \notin \partial\Pi$ also has $p' \in \Pi'$; by the above $p' = p$. ■

Correction. In my handwritten notes the existence part concludes “ $(u - m), (v - n) \notin \partial\Pi$ ”. This is not true in general: if $u \in \mathbb{Z}$, then $u - m = 0$ and p lies on the edge spanned by ω_2 . What is true is $0 \leq u - m < 1$ and $0 \leq v - n < 1$, i.e. $p \in \Pi'$. Points of $\partial\Pi$ really do have two (or four, for the corners) representations with $p \in \Pi$, e.g. $\omega_1 = 0 + \omega_1 = \omega_1 + 0$.

Example 2.2.5. Let $L = \mathbb{Z}[i]$ (so $\omega_1 = 1, \omega_2 = i$) and $x = \frac{5}{2} + \frac{5}{2}i$. Then $u = v = \frac{5}{2}$, $m = n = 2$, and

$$x = \underbrace{(2 + 2i)}_{\ell \in L} + \underbrace{\left(\frac{1}{2} + \frac{1}{2}i\right)}_{p \in \Pi'}.$$

Correction. In my handwritten example the decomposition was written as $\ell = 2 + i$ and $p = \frac{1}{2} + \frac{3}{2}i$. But $\frac{1}{2} + \frac{3}{2}i = \frac{1}{2}\omega_1 + \frac{3}{2}\omega_2$ does not lie in Π (the coefficient of ω_2 exceeds 1); the correct decomposition is the one above.

Proposition 2.2.6

Let L be the lattice generated by ω_1, ω_2 and let $\omega'_1, \omega'_2 \in \mathbb{C}$. Then ω'_1, ω'_2 generate the same lattice, i.e.

$$L = \{a\omega_1 + b\omega_2 \mid a, b \in \mathbb{Z}\} = \{a\omega'_1 + b\omega'_2 \mid a, b \in \mathbb{Z}\} = L',$$

if and only if there is an integer 2×2 matrix A with $\det A = \pm 1$ such that $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = A \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$.

Proof. ($\Rightarrow$) Let $L = \{m\omega_1 + n\omega_2 \mid m, n \in \mathbb{Z}\}$ and $L' = \{k\omega'_1 + l\omega'_2 \mid k, l \in \mathbb{Z}\}$, and suppose $L = L'$. Since $\omega'_1, \omega'_2 \in L' = L$, there are $a, b, c, d \in \mathbb{Z}$ with

$$\omega'_1 = a\omega_1 + b\omega_2, \quad \omega'_2 = c\omega_1 + d\omega_2, \quad \text{i.e.} \quad \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}, \quad M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Similarly, since $\omega_1, \omega_2 \in L = L'$, there are $p, q, r, s \in \mathbb{Z}$ with

$$\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = N \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix}, \quad N = \begin{pmatrix} p & q \\ r & s \end{pmatrix}.$$

Combining these two systems,

$$\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \left(N \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} \right) = MN \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix}, \quad \text{so} \quad (MN - I) \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Each row (x_1, x_2) of the integer matrix $MN - I$ gives a relation $x_1\omega'_1 + x_2\omega'_2 = 0$ with real coefficients. Since ω'_1, ω'_2 generate the lattice L' , they are linearly independent over $\mathbb{R}$, so $x_1 = x_2 = 0$. Hence $MN = I$, and

$$\det(M) \det(N) = \det(MN) = \det(I) = 1.$$

As $\det M$ and $\det N$ are integers whose product is 1, the only possibility is $\det M = \det N = \pm 1$.

($\Leftarrow$) Suppose $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$ with $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2 \times 2}(\mathbb{Z})$ and $\varepsilon := \det M = ad - bc = \pm 1$. Then $\omega'_1 = a\omega_1 + b\omega_2 \in L$ and $\omega'_2 = c\omega_1 + d\omega_2 \in L$; since L is an additive group, $L' \subseteq L$. Conversely, since $\varepsilon \neq 0$, M is invertible with

$$M^{-1} = \frac{1}{\varepsilon} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \varepsilon \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \in M_{2 \times 2}(\mathbb{Z})$$

(using $1/\varepsilon = \varepsilon$ for $\varepsilon = \pm 1$). So

$$\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \varepsilon \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} \implies \omega_1 = \varepsilon(d\omega'_1 - b\omega'_2) \in L', \quad \omega_2 = \varepsilon(-c\omega'_1 + a\omega'_2) \in L'.$$

Hence $L \subseteq L'$, and therefore $L = L'$. (Note that ω'_1, ω'_2 are then automatically linearly independent over $\mathbb{R}$, because M is invertible.) ■

Definition 2.2.7 (Elliptic function)

Let $L = \{m\omega_1 + n\omega_2\}$ be a lattice. A meromorphic function f on $\mathbb{C}$ is called an *elliptic function with respect to L* if it is *doubly periodic* with periods ω_1, ω_2 :

$$f(z + \omega_1) = f(z) \quad \text{and} \quad f(z + \omega_2) = f(z) \quad \text{for all } z \in \mathbb{C},$$

or equivalently, $f(z + \ell) = f(z)$ for all $\ell \in L$ and $z \in \mathbb{C}$.

Remark 2.2.8. An elliptic function is determined by its values on Π , and its values on opposite sides of Π agree: $f(a\omega_1 + \omega_2) = f(a\omega_1)$ and $f(\omega_1 + b\omega_2) = f(b\omega_2)$. So an elliptic function can be thought of as a function on the parallelogram Π with opposite sides glued together, i.e. on the *torus* $\mathbb{C}/L$.

Definition 2.2.9

The set of all elliptic functions with respect to L is denoted by $\mathcal{E}_L$. It is a subfield of the field of all meromorphic functions on $\mathbb{C}$ (Proposition 2.2.10).

Proposition 2.2.10

For all $f, g \in \mathcal{E}_L$ we have $f \pm g, fg, f' \in \mathcal{E}_L$, and $f/g \in \mathcal{E}_L$ if $g \not\equiv 0$.

Proof. Sums, products and quotients (by a function which is not identically zero) of meromorphic functions are meromorphic, and so is the derivative. For $\ell \in L$ and $z \in \mathbb{C}$:

$$(f \pm g)(z + \ell) = f(z + \ell) \pm g(z + \ell) = f(z) \pm g(z) = (f \pm g)(z),$$

and in the same way $(fg)(z + \ell) = (fg)(z)$ and $(f/g)(z + \ell) = (f/g)(z)$. Finally, differentiating the identity $f(z + \ell) = f(z)$ with respect to z (chain rule) gives $f'(z + \ell) = f'(z)$. ■

Proposition 2.2.11

A function $f \in \mathcal{E}_L$ which has no poles in Π is constant.

Proof. Since f is meromorphic and has no poles in Π , by periodicity it has no poles at all (every $z \in \mathbb{C}$ is $z = \ell + p$ with $p \in \Pi$ by Proposition 2.2.4, and f has a pole at z iff it has a pole at p). So f is entire. Since Π is compact (closed and bounded) and f is continuous, f is bounded on Π , say $|f| \leq M$ on Π . Again by periodicity, $|f(z)| = |f(p)| \leq M$ for all $z \in \mathbb{C}$. Thus f is a bounded entire function, and by Liouville's theorem (recalled below) f is constant. ■

Correction. In my handwritten notes it was concluded from “ f is holomorphic on the bounded region Π ” directly that f is constant. This is not enough: a function which is holomorphic and bounded on a bounded region need not be constant (e.g. $f(z) = z$). Liouville’s theorem needs f to be bounded and holomorphic on all of $\mathbb{C}$, and this is exactly what double periodicity provides.

Recall (Liouville’s theorem). *If f is entire and bounded, then f is constant.* Indeed, let $|f| \leq M$ on $\mathbb{C}$ and fix $s \in \mathbb{C}$. For every $R > 0$, Cauchy’s integral formula on the circle $C_R : z = s + Re^{i\theta}$, $0 \leq \theta \leq 2\pi$, gives

$$f'(s) = \frac{1}{2\pi i} \oint_{C_R} \frac{f(z)}{(z-s)^2} dz = \frac{1}{2\pi i} \int_0^{2\pi} \frac{f(s + Re^{i\theta})}{(Re^{i\theta})^2} iRe^{i\theta} d\theta.$$

By the integral inequality,

$$|f'(s)| \leq \frac{1}{2\pi} \int_0^{2\pi} \frac{|f(s + Re^{i\theta})|}{R} d\theta \leq \frac{1}{2\pi} \int_0^{2\pi} \frac{M}{R} d\theta = \frac{M}{R}.$$

Letting $R \rightarrow \infty$ gives $|f'(s)| \leq 0$, so $f'(s) = 0$. As s was arbitrary, $f' \equiv 0$ and f is constant. (In my handwritten notes the factor $\frac{1}{2\pi i}$ was written as $\frac{1}{2\pi}$ and the exponent of $(z-s)$ was first written as 1; the formula above is the correct one for f' .)

From now on we orient the basis so that $\text{Im}(\omega_2/\omega_1) > 0$ (otherwise swap ω_1 and ω_2); then the boundary $\partial(\alpha + \Pi)$, traversed $\alpha \rightarrow \alpha + \omega_1 \rightarrow \alpha + \omega_1 + \omega_2 \rightarrow \alpha + \omega_2 \rightarrow \alpha$, is positively oriented (counterclockwise).

Notes. A meromorphic function has only finitely many zeros and poles in any compact set K (unless it is identically zero): its poles are isolated, so if there were infinitely many in K they would accumulate at a point of K by compactness, which would be a non-isolated singularity; the same argument applies to the zeros of $f \not\equiv 0$, by the identity theorem. Consequently, for $f \in \mathcal{E}_L$, $f \not\equiv 0$, we can always choose $\alpha \in \mathbb{C}$ such that f has no zeros and no poles on $\partial(\alpha + \Pi)$: the zeros and poles of f in a slightly larger compact set are finitely many, and we only need to move α off finitely many lines. (In my handwritten notes this point was marked “no idea”.)

Proposition 2.2.12

Let $f \in \mathcal{E}_L$ and $\alpha \in \mathbb{C}$ such that f has no poles on $\partial(\alpha + \Pi)$. Then the sum of the residues of f at its poles in $\alpha + \Pi$ is 0.

Proof. By the previous Notes, f has finitely many poles in $\alpha + \Pi$, all in the interior. At each such pole z_0 , f has a Laurent expansion $f(z) = \sum_{n \geq 1} \frac{b_n}{(z-z_0)^n} + \sum_{n \geq 0} a_n(z-z_0)^n$, and the residue is b_1 . By the residue theorem,

$$\sum_{z_0 \in \alpha + \Pi} \text{Res}_{z=z_0} f = \frac{1}{2\pi i} \oint_{\partial(\alpha + \Pi)} f(z) dz.$$

The boundary consists of four sides:

$$C_1 : \alpha \rightarrow \alpha + \omega_1, \quad C_2 : \alpha + \omega_1 \rightarrow \alpha + \omega_1 + \omega_2, \quad C_3 : \alpha + \omega_1 + \omega_2 \rightarrow \alpha + \omega_2, \quad C_4 : \alpha + \omega_2 \rightarrow \alpha.$$

The side C_3 is the translate of C_1 by ω_2 , traversed in the opposite direction. Substituting

$z = w + \omega_2$ and using $f(w + \omega_2) = f(w)$,

$$\int_{C_3} f(z) dz = - \int_{C_1} f(w + \omega_2) dw = - \int_{C_1} f(w) dw.$$

Similarly C_2 is the translate of $-C_4$ by ω_1 , and $\int_{C_2} f = - \int_{C_4} f$. Hence

$$\oint_{\partial(\alpha + \Pi)} f(z) dz = \int_{C_1} f + \int_{C_2} f + \int_{C_3} f + \int_{C_4} f = \int_{C_1} f - \int_{C_4} f - \int_{C_1} f + \int_{C_4} f = 0. \quad \blacksquare$$

Recall (Argument principle). Let f be meromorphic on an open set U , and let C be a positively oriented simple closed contour in U (with its interior in U) on which f has no zeros and no poles. Then

$$\frac{1}{2\pi i} \oint_C \frac{f'(z)}{f(z)} dz = \#\{\text{zeros of } f \text{ inside } C\} - \#\{\text{poles of } f \text{ inside } C\},$$

where zeros and poles are counted with multiplicity (order).

Proposition 2.2.13

Let $f \in \mathcal{E}_L$, $f \not\equiv 0$, and let $\alpha \in \mathbb{C}$ be such that f has no zeros and no poles on $\partial(\alpha + \Pi)$. Then, counting with multiplicity,

$$\#\{\text{zeros of } f \text{ in } \alpha + \Pi\} = \#\{\text{poles of } f \text{ in } \alpha + \Pi\}.$$

Proof. Since $f \in \mathcal{E}_L$, also $f' \in \mathcal{E}_L$ and $f'/f \in \mathcal{E}_L$ (Proposition 2.2.10). The poles of f'/f are among the zeros and poles of f , so f'/f has no poles on $\partial(\alpha + \Pi)$. By the argument principle and Proposition 2.2.12 applied to f'/f ,

$$\#\{\text{zeros of } f \text{ in } \alpha + \Pi\} - \#\{\text{poles of } f \text{ in } \alpha + \Pi\} = \frac{1}{2\pi i} \oint_{\partial(\alpha + \Pi)} \frac{f'(z)}{f(z)} dz = 0. \quad \blacksquare$$

Notes. Since every point of $\mathbb{C}$ has a unique representative in the half-open translate $\alpha + \Pi'$, and there are no zeros or poles on $\partial(\alpha + \Pi)$, the counts in Proposition 2.2.13 are the numbers of zeros and poles of f modulo L , i.e. on the torus $\mathbb{C}/L$; in particular they do not depend on α . We call the common value the *order* of f . If f does have zeros or poles on $\partial\Pi$, one can either move α , or (as drawn in my handwritten notes) indent the contour around them: a zero or pole on an edge is then counted with weight $\frac{1}{2}$ on each of the two opposite edges, and one at a corner with weight $\frac{1}{4}$ at each of the four corners, which again adds up to 1.

Proposition 2.2.14

There is no $f \in \mathcal{E}_L$ which has exactly one pole modulo L , this pole being simple.

Proof. Suppose, by way of contradiction, that $f \in \mathcal{E}_L$ has exactly one pole z_0 in $\alpha + \Pi$ (with α chosen such that there is no pole on the boundary), and that it is simple. Then its residue $c = \text{Res}_{z=z_0} f$ is nonzero: at a simple pole $f(z) = \frac{c}{z-z_0} + (\text{holomorphic})$ with $c \neq 0$. But by Proposition 2.2.12 the sum of the residues in $\alpha + \Pi$, which is c , equals 0. Contradiction. $\blacksquare$

Thus a nonconstant elliptic function has at least two poles (counted with multiplicity) modulo L . The simplest example, with one double pole, is the Weierstrass $\wp$ -function.

2.3 The Weierstrass $\wp$ -function

Definition 2.3.1 (Weierstrass $\wp$ -function)

For a lattice $L = \{m\omega_1 + n\omega_2\}$ we define

$$\wp(z) = \wp(z; L) = \frac{1}{z^2} + \sum_{\substack{\ell \in L \\ \ell \neq 0}} \left(\frac{1}{(z - \ell)^2} - \frac{1}{\ell^2} \right) \quad (z \in \mathbb{C} \setminus L). \quad (2.2)$$

We first show that the series (2.2) converges. We need two lemmas.

Lemma 2.3.2

There is a constant $C_L > 0$ such that for all real $R \geq 1$,

$$\#\{\ell \in L : R \leq |\ell| < R + 1\} \leq C_L R.$$

Proof. Let $d = \max_{p \in \Pi} |p|$ be the maximal distance from 0 to a point of Π , and let $A = \text{area}(\Pi) > 0$. The translates $\ell + \Pi'$, $\ell \in L$, are pairwise disjoint (Proposition 2.2.4), and each has area A .

Let $\ell \in L$ with $R \leq |\ell| < R + 1$. For every $z = \ell + p \in \ell + \Pi'$ we have $|p| \leq d$, so

$$|z| = |\ell + p| \leq |\ell| + |p| < R + 1 + d \quad \text{and} \quad |z| = |\ell + p| \geq |\ell| - |p| \geq R - d.$$

Hence all these (disjoint) tiles lie in the region $\{\max(0, R - d) \leq |z| < R + 1 + d\}$, and comparing areas,

$$\#\{\ell : R \leq |\ell| < R + 1\} \cdot A \leq \pi((R + 1 + d)^2 - \max(0, R - d)^2).$$

If $R \geq d$, the right-hand side is $\pi(2R + 1)(1 + 2d) \leq 3\pi(1 + 2d)R$ (using $1 \leq R$). If $R < d$, it is $\pi(R + 1 + d)^2 \leq \pi(1 + 2d)^2 \leq \pi(1 + 2d)^2 R$. In both cases the claim holds with

$$C_L = \frac{3\pi(1 + 2d)^2}{A},$$

which depends only on L (through d and A). ■

Correction. In my handwritten notes the annulus was taken to be $R - d \leq |z| < R + 1 + d$, giving the constant $C_L = 3\pi(1 + 2d)/A$. This is fine for $R \geq d$, but when $R < d$ the inner radius $R - d$ is negative and the area of the region is $\pi(R + 1 + d)^2$, not $\pi((R + 1 + d)^2 - (R - d)^2)$. Enlarging the constant as above covers both cases.

Lemma 2.3.3

For every real $k > 2$ the series $\sum_{\substack{\ell \in L \\ \ell \neq 0}} \frac{1}{|\ell|^k}$ converges.

Proof. Split the sum according to the size of $|\ell|$:

$$\sum_{\ell \neq 0} \frac{1}{|\ell|^k} = \sum_{0 < |\ell| < 1} \frac{1}{|\ell|^k} + \sum_{R=1}^{\infty} \left(\sum_{\ell \in A_R} \frac{1}{|\ell|^k} \right), \quad A_R = \{\ell \in L : R \leq |\ell| < R + 1\}.$$

The first sum has only finitely many terms (only finitely many lattice points lie in the bounded disc $|z| < 1$, since the tiles $\ell + \Pi'$ are disjoint of area A); call it C_0 . For $\ell \in A_R$ we have $R \leq |\ell|$,

so $\frac{1}{|\ell|^k} \leq \frac{1}{R^k}$, and by Lemma 2.3.2,

$$\sum_{\ell \neq 0} \frac{1}{|\ell|^k} \leq C_0 + \sum_{R=1}^{\infty} \frac{\#A_R}{R^k} \leq C_0 + \sum_{R=1}^{\infty} \frac{C_L R}{R^k} = C_0 + C_L \sum_{R=1}^{\infty} \frac{1}{R^{k-1}} < \infty,$$

since $k - 1 > 1$. ■

Notes. In my handwritten notes this series was called “absolutely and uniformly convergent”; since it does not depend on a variable, only convergence (of a series of positive terms) is meant.

Proposition 2.3.4

For $0 \neq \ell \in L$ put $f_\ell(z) = \frac{1}{(z-\ell)^2} - \frac{1}{\ell^2}$. The series $\sum_{\ell \neq 0} f_\ell(z)$, and hence the series (2.2), converges absolutely and uniformly on every compact subset K of $\mathbb{C} \setminus L$.

Proof. First,

$$f_\ell(z) = \frac{1}{(z-\ell)^2} - \frac{1}{\ell^2} = \frac{\ell^2 - (z-\ell)^2}{\ell^2(z-\ell)^2} = \frac{\ell^2 - (z^2 - 2z\ell + \ell^2)}{\ell^2(z-\ell)^2} = \frac{2z\ell - z^2}{\ell^2(z-\ell)^2} = \frac{2z - z^2/\ell}{\ell(z-\ell)^2}.$$

Let $K \subseteq \mathbb{C} \setminus L$ be compact. Since K is bounded, there is $R > 0$ with $|z| \leq R$ for all $z \in K$. Then

$$|f_\ell(z) \ell^3| = \left| \frac{2z - z^2/\ell}{\ell(z-\ell)^2} \ell^3 \right| = \left| \frac{2z - z^2/\ell}{\ell^2(1 - z/\ell)^2} \ell^2 \right| = \left| \frac{2z - z^2/\ell}{(1 - z/\ell)^2} \right|.$$

As $|\ell| \rightarrow \infty$ this tends to $|2z|$, uniformly for $z \in K$. To get an explicit bound, let $|\ell| \geq 2R$. Then $|z/\ell| \leq \frac{1}{2}$, so $|1 - z/\ell| \geq 1 - |z/\ell| \geq \frac{1}{2}$ and $|2z - z^2/\ell| \leq 2R + R \cdot \frac{1}{2} = \frac{5}{2}R$. Therefore

$$|f_\ell(z)| \leq \frac{\frac{5}{2}R}{(\frac{1}{2})^2} \cdot \frac{1}{|\ell|^3} = \frac{10R}{|\ell|^3} =: M_\ell \quad (z \in K, |\ell| \geq 2R).$$

By Lemma 2.3.3 with $k = 3$, $\sum_{|\ell| \geq 2R} M_\ell = 10R \sum_{|\ell| \geq 2R} |\ell|^{-3} < \infty$. By the Weierstrass M -test (Lemma 1.3.2), $\sum_{|\ell| \geq 2R} f_\ell(z)$ converges absolutely and uniformly on K . The remaining terms, with $0 < |\ell| < 2R$, are finitely many, and each of them is a continuous (hence bounded) function on K , since $K \cap L = \emptyset$. Adding finitely many such terms does not affect absolute and uniform convergence. ■

Correction. In my handwritten notes it was concluded from $\lim_{|\ell| \rightarrow \infty} |f_\ell(z) \ell^3| = |2z|$ that $|f_\ell(z) \ell^3| = |2z| \leq 2R$. A limit does not give an inequality for all ℓ ; the correct statement is the explicit bound $|f_\ell(z)| \leq 10R |\ell|^{-3}$ for $|\ell| \geq 2R$ above. Also, the convergence is uniform on compact subsets K of $\mathbb{C} \setminus L$, not on all of $\mathbb{C} \setminus L$.

Proposition 2.3.5

$\wp \in \mathcal{E}_L$, and its only poles are double poles, exactly at the lattice points. Moreover $\wp$ is even, and

$$\wp'(z) = -2 \sum_{\ell \in L} \frac{1}{(z-\ell)^3}. \quad (2.3)$$

Proof. $\wp$ is meromorphic with double poles exactly at L . By Proposition 2.3.4, the partial sums of (2.2), which are holomorphic on $\mathbb{C} \setminus L$, converge uniformly on compact subsets of $\mathbb{C} \setminus L$; by Lemma 1.1.3, $\wp$ is holomorphic on $\mathbb{C} \setminus L$. Now fix $\ell_0 \in L$ and a closed disc $\bar{D}$ around ℓ_0 containing no other lattice point. Applying the estimate of Proposition 2.3.4 to the compact set $\bar{D}$ and the lattice points $\ell \neq 0, \ell_0$ (the bound only used $|z| \leq R$ and $z \neq \ell$), we see that

$$\wp(z) - \frac{1}{(z - \ell_0)^2}$$

is holomorphic on D (for $\ell_0 \neq 0$ we also add the constant $-1/\ell_0^2$ and the holomorphic term $1/z^2$). Hence $\wp$ has a double pole at ℓ_0 , with principal part $\frac{1}{(z - \ell_0)^2}$. So $\wp$ is meromorphic on $\mathbb{C}$, and its poles are exactly the points of L , all double.

$\wp$ is even. Note that $\frac{1}{(-z)^2} = \frac{1}{z^2}$ and $\frac{1}{(-z - \ell)^2} = \frac{1}{(z + \ell)^2}$, and that $\ell \mapsto -\ell$ is a bijection of $L \setminus \{0\}$; summing over ℓ or over $-\ell$ is the same (the series converges absolutely, so it may be rearranged). Thus

$$\wp(-z) = \frac{1}{(-z)^2} + \sum_{\ell \neq 0} \left(\frac{1}{(-z - \ell)^2} - \frac{1}{\ell^2} \right) = \frac{1}{z^2} + \sum_{\ell \neq 0} \left(\frac{1}{(z + \ell)^2} - \frac{1}{\ell^2} \right) = \frac{1}{z^2} + \sum_{\ell \neq 0} \left(\frac{1}{(z - \ell)^2} - \frac{1}{\ell^2} \right) = \wp(z).$$

The derivative. A locally uniformly convergent series of holomorphic functions may be differentiated term by term (Weierstrass' theorem, [6, Ch. 2, Thm. 5.3]). Hence

$$\wp'(z) = \frac{-2}{z^3} + \sum_{\ell \neq 0} \frac{-2}{(z - \ell)^3} = -2 \sum_{\ell \in L} \frac{1}{(z - \ell)^3},$$

and this series converges absolutely and uniformly on compact subsets of $\mathbb{C} \setminus L$ by the same argument (or directly by Lemma 2.3.3 with $k = 3$).

$\wp'$ is periodic. For $\omega \in \{\omega_1, \omega_2\}$, substituting $\ell' = \ell - \omega$ (a bijection of L),

$$\wp'(z + \omega) = -2 \sum_{\ell \in L} \frac{1}{(z + \omega - \ell)^3} = -2 \sum_{\ell' \in L} \frac{1}{(z - \ell')^3} = \wp'(z).$$

$\wp$ is periodic. Since $\frac{d}{dz}(\wp(z + \omega_1) - \wp(z)) = \wp'(z + \omega_1) - \wp'(z) = 0$ on the connected open set $\mathbb{C} \setminus L$, there is a constant C with

$$\wp(z + \omega_1) - \wp(z) = C \quad (z \in \mathbb{C} \setminus L).$$

Put $z = -\frac{\omega_1}{2}$, which is not a lattice point (as ω_1, ω_2 are linearly independent, $\frac{1}{2}\omega_1 \notin L$). We get

$$C = \wp\left(\frac{\omega_1}{2}\right) - \wp\left(-\frac{\omega_1}{2}\right) = 0,$$

since $\wp$ is even. Thus $\wp(z + \omega_1) = \wp(z)$, and analogously $\wp(z + \omega_2) = \wp(z)$. Therefore $\wp$ is doubly periodic, i.e. $\wp \in \mathcal{E}_L$. ■

Corollary 2.3.6 (1) $\wp$ has exactly 2 zeros modulo L (counted with multiplicity).

(2) $\wp'$ has exactly 3 zeros modulo L , namely the *half-periods*

$$\frac{\omega_1}{2}, \quad \frac{\omega_2}{2}, \quad \frac{\omega_3}{2} := \frac{\omega_1 + \omega_2}{2},$$

and each of them is a *simple* zero.

Proof. (1) Modulo L , $\wp$ has exactly one pole (at 0), which is double. By Proposition 2.2.13 (and the Notes after it),

$$\#\{\text{zeros of } \wp \bmod L\} = \#\{\text{poles of } \wp \bmod L\} = 1 \cdot 2 = 2.$$

(2) By (2.3), modulo L , $\wp'$ has exactly one pole (at 0), which is triple; so $\wp'$ has exactly 3 zeros modulo L , counted with multiplicity. Since $\wp$ is even, differentiating $\wp(z) = \wp(-z)$ gives $\wp'(z) = -\wp'(-z)$, i.e. $\wp'$ is odd. Let ω be one of $\omega_1, \omega_2, \omega_3$; then $\omega \in L$ but $\frac{\omega}{2} \notin L$. By periodicity and oddness,

$$\wp'\left(\frac{\omega}{2}\right) = \wp'\left(\frac{\omega}{2} - \omega\right) = \wp'\left(-\frac{\omega}{2}\right) = -\wp'\left(\frac{\omega}{2}\right) \implies 2\wp'\left(\frac{\omega}{2}\right) = 0 \implies \wp'\left(\frac{\omega}{2}\right) = 0.$$

The three half-periods are pairwise distinct modulo L (their differences $\frac{\omega_1}{2} - \frac{\omega_2}{2}, \frac{\omega_2}{2}, \frac{\omega_1}{2}$ are not in L). So $\wp'$ has at least three distinct zeros modulo L ; as the total number with multiplicity is 3, these are all the zeros and each is simple. ■

Correction. In my handwritten notes it is stated (and later used) that $\wp'$ has “exactly 3 double zeros”. This is impossible, since $\wp'$ has only 3 zeros counted with multiplicity: the zeros of $\wp'$ at the half-periods are *simple*. What has double zeros at the half-periods is $(\wp')^2$ (and, as we show next, the functions $\wp(z) - \wp(\omega/2)$).

Proposition 2.3.7

Put $e_i = \wp(\frac{\omega_i}{2})$ for $i = 1, 2, 3$. Let $u \in \mathbb{C}$. Then $\wp(z) - u$ has a double zero if and only if $u \in \{e_1, e_2, e_3\}$. In that case, for $u = e_i$, the only zero of $\wp - e_i$ modulo L is the double zero at $\frac{\omega_i}{2}$. Moreover, e_1, e_2, e_3 are pairwise distinct.

Proof. ($\Leftarrow$) Let $u = e_1$ (the other two cases are analogous) and $f(z) = \wp(z) - u$. At $z = \frac{\omega_1}{2}$,

$$f\left(\frac{\omega_1}{2}\right) = \wp\left(\frac{\omega_1}{2}\right) - u = u - u = 0,$$

so $\frac{\omega_1}{2}$ is a zero of f . Differentiating, $f'(z) = \wp'(z)$, so $f'(\frac{\omega_1}{2}) = \wp'(\frac{\omega_1}{2}) = 0$ by Corollary 2.3.6. Therefore $f(\frac{\omega_1}{2}) = f'(\frac{\omega_1}{2}) = 0$, and $z = \frac{\omega_1}{2}$ is a zero of multiplicity ≥ 2 of $f = \wp - e_1$. On the other hand, $f = \wp - u$ (with u a constant) has exactly one double pole modulo L , hence exactly 2 zeros modulo L (Proposition 2.2.13). So $\frac{\omega_1}{2}$ is a zero of multiplicity exactly 2, and it is the only zero modulo L .

($\Rightarrow$) If $\wp - u$ has a double zero at z_0 , then $z_0 \notin L$ (these are poles) and $\wp'(z_0) = 0$. By Corollary 2.3.6, $z_0 \equiv \frac{\omega_i}{2} \pmod{L}$ for some i , and then $u = \wp(z_0) = \wp(\frac{\omega_i}{2}) = e_i$.

Distinctness. Suppose, say, $e_1 = e_2$. Then $\wp - e_1$ has a double zero at $\frac{\omega_1}{2}$ and (since $e_2 = e_1$) also a double zero at $\frac{\omega_2}{2} \not\equiv \frac{\omega_1}{2}$, i.e. at least 4 zeros modulo L ; but it has exactly 2. Contradiction. ■

Remark 2.3.8. Just as $\sin$ and $\cos$ play a basic role in the theory of periodic functions on $\mathbb{R}$, the functions $\wp$ and $\wp'$ play the basic role in the theory of elliptic functions: we now show that every elliptic function is a rational function of $\wp$ and $\wp'$ (no infinite series needed).

The field of elliptic functions

Let $\mathcal{E}_L^+ \subseteq \mathcal{E}_L$ be the subfield of even elliptic functions. For a meromorphic function $f \neq 0$ and $a \in \mathbb{C}$ we write $\text{ord}_a f \in \mathbb{Z}$ for the order of f at a : it is $m > 0$ if f has a zero of order m at a , $-m < 0$ if f has a pole of order m , and 0 otherwise.

Lemma 2.3.9

Let $f \in \mathcal{E}_L^+$, $f \neq 0$, and $a \in \mathbb{C}$.

- (i) $\text{ord}_{-a} f = \text{ord}_a f$, and $\text{ord}_{a+\ell} f = \text{ord}_a f$ for $\ell \in L$.
- (ii) If $2a \in L$ (i.e. $a \in L$ or a is congruent to a half-period), then $\text{ord}_a f$ is even.

Proof. (i) Write the Laurent expansion $f(a+w) = \sum_{n \geq m} c_n w^n$ with $c_m \neq 0$, $m = \text{ord}_a f$. Since f is even, $f(-a-w) = f(a+w) = \sum_n c_n w^n$, i.e. $f(-a+w') = \sum_n c_n (-w')^n = \sum_n (-1)^n c_n w'^n$ with $w' = -w$; the lowest coefficient is $(-1)^m c_m \neq 0$, so $\text{ord}_{-a} f = m$. Periodicity gives $f(a+\ell+w) = f(a+w)$, hence $\text{ord}_{a+\ell} f = \text{ord}_a f$.

(ii) Suppose $2a \in L$ and put $g(w) = f(a+w)$. Then, using evenness and periodicity,

$$g(-w) = f(a-w) = f(-(a-w)) = f(w-a) = f(w-a+2a) = f(a+w) = g(w).$$

So g is an even function, its Laurent expansion $\sum_n c_n w^n$ at $w = 0$ contains only even powers of w , and hence $\text{ord}_a f = \text{ord}_0 g$ is even. ■

Proposition 2.3.10

$\mathcal{E}_L^+ = \mathbb{C}(\wp)$, i.e. every even elliptic function is a rational function of $\wp$.

Proof. Clearly $\mathbb{C}(\wp) \subseteq \mathcal{E}_L^+$, since $\wp$ is even and elliptic. Conversely, let $f \in \mathcal{E}_L^+$, $f \neq 0$. The idea is to construct $g \in \mathbb{C}(\wp)$ with the same zeros and poles as f ; then f/g has no zeros and no poles, hence is constant.

Consider the half-open parallelogram Π' ; every point of $\mathbb{C}$ differs by a lattice point from exactly one point of Π' . Let $a \in \Pi'$, $a \neq 0$. Define its *symmetric point* $a^* \in \Pi'$ as the representative of $-a$, i.e.

$$a^* = \begin{cases} \omega_1 + \omega_2 - a & \text{if } a \text{ lies in the interior of } \Pi, \\ \omega_1 - a \text{ or } \omega_2 - a & \text{if } a \text{ lies on the edge spanned by } \omega_1 \text{ resp. } \omega_2. \end{cases}$$

Then $a^* = a$ if and only if $2a \in L$, i.e. a is one of the half-periods $\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_3}{2}$. By Lemma 2.3.9(i), $\text{ord}_{a^*} f = \text{ord}_a f$, and by (ii) $\text{ord}_a f$ is even when a is a half-period.

Now choose a set $R \subseteq \Pi' \setminus \{0\}$ containing exactly one point of each pair $\{a, a^*\}$ with $a \neq a^*$, together with the three half-periods. For $a \in R$ put

$$m_a = \begin{cases} \text{ord}_a f & \text{if } a \text{ is not a half-period,} \\ \frac{1}{2} \text{ord}_a f & \text{if } a \text{ is a half-period,} \end{cases}$$

so $m_a \in \mathbb{Z}$, and $m_a \neq 0$ for only finitely many a (the zeros and poles of f in Π'). Define

$$g(z) = \prod_{a \in R} (\wp(z) - \wp(a))^{m_a} \in \mathbb{C}(\wp).$$

(The zeros of f give factors with $m_a > 0$ in the numerator and the poles of f give factors with $m_a < 0$ in the denominator, as in my handwritten notes.) We claim that $\text{ord}_z g = \text{ord}_z f$ for every $z \notin L$. For $a \in R$, the function $\wp(z) - \wp(a)$ has a double pole at 0 modulo L , hence exactly two zeros modulo L :

- if a is not a half-period, these zeros are a and $a^* \equiv -a$ (as $\wp$ is even), they are distinct, so both are simple;
- if a is a half-period, it is a double zero at a and there is no other zero (Proposition 2.3.7).

So, for $z_0 \in \Pi' \setminus \{0\}$: if $z_0 \in \{a, a^*\}$ with $a \in R$ not a half-period, then $\text{ord}_{z_0} g = m_a = \text{ord}_a f = \text{ord}_{z_0} f$; if $z_0 = a$ is a half-period, then $\text{ord}_{z_0} g = 2m_a = \text{ord}_{z_0} f$. This proves the claim (by periodicity, for all $z \notin L$).

Hence $h = f/g \in \mathcal{E}_L$ has no zeros and no poles outside L . Modulo L , its only possible zero or pole is at 0. If $\text{ord}_0 h = k > 0$, then h would have k zeros and no poles modulo L , contradicting Proposition 2.2.13; similarly $k < 0$ is impossible. So $\text{ord}_0 h = 0$, and h has no poles at all. By Proposition 2.2.11, $h = c$ is a constant (nonzero, as $f \neq 0$). Therefore $f = cg \in \mathbb{C}(\wp)$. ■

Correction. In my handwritten notes it was claimed that g has poles at the half-periods “because $\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_1+\omega_2}{2} \in L \setminus \{0\}$ and $\wp$ has double poles at $L \setminus \{0\}$ ”. This is wrong: the half-periods are *not* lattice points. The correct reason that the orders at the half-periods match is that $\wp(z) - \wp(\frac{\omega}{2})$ has a *double zero* at $\frac{\omega}{2}$ (Proposition 2.3.7), which is why the exponent m_a is taken to be $\frac{1}{2} \text{ord}_a f$ there. The comparison of orders at $z = 0$ (which my notes did by noting that the orders of f sum to 0 and those of g sum to 0) is the argument with $h = f/g$ above.

Proposition 2.3.11

$\mathcal{E}_L = \mathbb{C}(\wp, \wp')$, i.e. every elliptic function is a rational function of $\wp$ and $\wp'$. More precisely, every $f \in \mathcal{E}_L$ can be written as $f = R_1(\wp) + \wp' R_2(\wp)$ with rational functions $R_1, R_2 \in \mathbb{C}(X)$.

Proof. For $f \in \mathcal{E}_L$ write

$$f(z) = f_{\text{even}}(z) + f_{\text{odd}}(z), \quad f_{\text{even}}(z) = \frac{f(z) + f(-z)}{2}, \quad f_{\text{odd}}(z) = \frac{f(z) - f(-z)}{2}.$$

Both are elliptic, f_{even} is even and f_{odd} is odd. Since $\wp'$ is odd and $\wp' \neq 0$, the quotient $h = f_{\text{odd}}/\wp'$ is even and elliptic (odd divided by odd). By Proposition 2.3.10, $f_{\text{even}} = R_1(\wp)$ and $h = R_2(\wp)$ for rational functions R_1, R_2 . Hence $f = R_1(\wp) + \wp' R_2(\wp)$. ■

2.4 The Weierstrass equation and Eisenstein series

We have obtained that every elliptic function lies in $\mathbb{C}(\wp, \wp')$, and every even elliptic function in $\mathbb{C}(\wp)$. Since $\wp'$ is odd, $(\wp')^2$ is even: $(\wp'(-z))^2 = (-\wp'(z))^2 = (\wp'(z))^2$. So $(\wp')^2$ is a rational function of $\wp$. Which one?

First way: comparing zeros and poles

Proposition 2.4.1

With $e_i = \wp(\frac{\omega_i}{2})$ as above,

$$(\wp'(z))^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3).$$

Proof. We follow the construction of g in the proof of Proposition 2.3.10 for $f = (\wp')^2 \in \mathcal{E}_L^+$.

- (1) Since $\wp$ has a double pole at 0 (with principal part $\frac{1}{z^2}$), $\wp'$ has a triple pole there (principal part $-\frac{2}{z^3}$), and $(\wp')^2$ has a pole of order 6 at 0.
- (2) By Corollary 2.3.6, $\wp'$ has simple zeros exactly at the three half-periods, so $(\wp')^2$ has *double* zeros exactly at $\frac{\omega_1}{2}, \frac{\omega_2}{2}, \frac{\omega_3}{2}$ and no other zeros modulo L .

So in the construction we get three half-periods with $m_a = \frac{1}{2} \cdot 2 = 1$, and no other points with $m_a \neq 0$. Hence

$$(\wp'(z))^2 = C (\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3)$$

for a constant C . To determine C , compare the leading terms at $z = 0$: since $\wp(z) = \frac{1}{z^2} + \dots$ and $\wp'(z) = -\frac{2}{z^3} + \dots$, the left-hand side is $\frac{4}{z^6} + \dots$ while the right-hand side is $\frac{C}{z^6} + \dots$. So $C = 4$. ■

Thus $(\wp')^2$ is a cubic polynomial in $\wp$. To find its coefficients in terms of L we compute Laurent expansions.

Second way: Laurent expansions

Definition 2.4.2 (Eisenstein series of a lattice)

For an integer $k > 2$ we put

$$G_k(L) = \sum_{\substack{\ell \in L \\ \ell \neq 0}} \frac{1}{\ell^k}.$$

Proposition 2.4.3

The series $G_k(L)$ converges absolutely for $k > 2$, and $G_k(L) = 0$ for odd $k > 2$.

Proof. Absolute convergence is Lemma 2.3.3. For odd k : since $\ell \mapsto -\ell$ is a bijection of $L \setminus \{0\}$ and the series converges absolutely (so it can be rearranged),

$$G_k(L) = \sum_{\ell \neq 0} \frac{1}{\ell^k} = \sum_{\ell \neq 0} \frac{1}{(-\ell)^k} = \sum_{\ell \neq 0} \frac{1}{-\ell^k} = -G_k(L),$$

hence $G_k(L) = 0$. ■

Lemma 2.4.4 (Laurent expansion of $\wp$)

Let $c = \min\{|\ell| : 0 \neq \ell \in L\}$. For $0 < |z| < c$,

$$\wp(z) = \frac{1}{z^2} + \sum_{k \geq 1} (2k+1)G_{2k+2}(L) z^{2k} = \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + 7G_8 z^6 + \dots$$

Proof. Fix $0 < r < 1$ and let $|z| \leq rc$. For $\ell \neq 0$ we have $|z/\ell| \leq r < 1$, and differentiating the geometric series $\frac{1}{1-w} = \sum_{j \geq 0} w^j$ gives $\frac{1}{(1-w)^2} = \sum_{j \geq 0} (j+1)w^j$ for $|w| < 1$. Hence

$$\frac{1}{(z-\ell)^2} - \frac{1}{\ell^2} = \frac{1}{\ell^2} \left(\frac{1}{(1-z/\ell)^2} - 1 \right) = \frac{1}{\ell^2} \sum_{j \geq 1} (j+1) \left(\frac{z}{\ell} \right)^j = \sum_{j \geq 1} (j+1) \frac{z^j}{\ell^{j+2}} = \frac{2z}{\ell^3} + \frac{3z^2}{\ell^4} + \frac{4z^3}{\ell^5} + \dots$$

We claim that the double series $\sum_{\ell \neq 0} \sum_{j \geq 1} (j+1) \frac{z^j}{\ell^{j+2}}$ converges absolutely. Indeed, using $|z/\ell| \leq r$,

$$\sum_{j \geq 1} (j+1) \frac{|z|^j}{|\ell|^{j+2}} = \frac{|z|}{|\ell|^3} \sum_{j \geq 1} (j+1) \left| \frac{z}{\ell} \right|^{j-1} \leq \frac{|z|}{|\ell|^3} \sum_{i \geq 0} (i+2) r^i \leq \frac{|z|}{|\ell|^3} \cdot \frac{2}{(1-r)^2},$$

since $\sum_{i \geq 0} (i+2) r^i = \frac{1}{(1-r)^2} + \frac{1}{1-r} \leq \frac{2}{(1-r)^2}$. And $\sum_{\ell \neq 0} \frac{2|z|}{(1-r)^2} \cdot \frac{1}{|\ell|^3} < \infty$ by Lemma 2.3.3. So we may interchange the order of summation:

$$\wp(z) = \frac{1}{z^2} + \sum_{j \geq 1} (j+1) \left(\sum_{\ell \neq 0} \frac{1}{\ell^{j+2}} \right) z^j = \frac{1}{z^2} + \sum_{j \geq 1} (j+1) G_{j+2}(L) z^j.$$

By Proposition 2.4.3 the terms with odd $j + 2$, i.e. odd j , vanish. Writing $j = 2k$ gives the claim. Since $r < 1$ was arbitrary, the expansion holds for $0 < |z| < c$. ■

Correction. In my handwritten notes the Taylor expansion was first written as “ $\wp(z) = 1 + \wp'(z)z + \wp(z)z^2 + \dots$ ”, which is meaningless; the correct expansion is the one above. Further, in the bound for the double series, the inequality $1 + \frac{3}{2}r + \frac{4}{2}r^2 + \dots < 1 + r + r^2 + \dots$ is false, but the final bound $\frac{2|z|}{|\ell|^3(1-r)^2}$ is correct, as shown. Finally, in the expansion “ $\frac{1}{z^2} + 2G_3 + 3G_4z^2 + 4G_5z^3$ ” the first term should read $2G_3z$ (it vanishes anyway).

Theorem 2.4.5 (Weierstrass equation)

Put

$$g_2 = g_2(L) = 60 G_4(L) = 60 \sum_{\ell \neq 0} \frac{1}{\ell^4}, \quad g_3 = g_3(L) = 140 G_6(L) = 140 \sum_{\ell \neq 0} \frac{1}{\ell^6}.$$

Then

$$(\wp'(z))^2 = 4\wp(z)^3 - g_2\wp(z) - g_3,$$

i.e. $(\wp')^2 = f(\wp)$ with $f(x) = 4x^3 - g_2x - g_3$. Moreover $f(x) = 4(x - e_1)(x - e_2)(x - e_3)$, so f has three distinct roots.

Proof. We want to find a cubic polynomial $f(x) = ax^3 + bx^2 + cx + d$ such that the Laurent expansion of $f(\wp(z))$ at 0 agrees with that of $(\wp'(z))^2$ in all negative powers of z and in the constant term. Then $(\wp')^2 - f(\wp)$ is an elliptic function whose Laurent expansion at 0 has no negative powers, hence it has no pole at 0 and therefore no poles at all (its only possible poles are at L); by Proposition 2.2.11 it is a constant, and since its constant term at 0 is 0, it tends to 0 as $z \rightarrow 0$, so it is identically 0.

By Lemma 2.4.4, writing G_k for $G_k(L)$,

$$\begin{aligned} \wp(z) &= \frac{1}{z^2} + 3G_4z^2 + 5G_6z^4 + \dots, \\ \wp'(z) &= -\frac{2}{z^3} + 6G_4z + 20G_6z^3 + \dots. \end{aligned}$$

Squaring and cubing (keeping all terms up to z^0):

$$\begin{aligned} (\wp'(z))^2 &= \frac{4}{z^6} - 2 \cdot \frac{2}{z^3} \cdot 6G_4z - 2 \cdot \frac{2}{z^3} \cdot 20G_6z^3 + \dots = \frac{4}{z^6} - \frac{24G_4}{z^2} - 80G_6 + \dots, \\ \wp(z)^2 &= \frac{1}{z^4} + 2 \cdot \frac{1}{z^2} \cdot 3G_4z^2 + \dots = \frac{1}{z^4} + 6G_4 + \dots, \\ \wp(z)^3 &= \frac{1}{z^6} + 3 \cdot \frac{1}{z^4} (3G_4z^2 + 5G_6z^4) + \dots = \frac{1}{z^6} + \frac{9G_4}{z^2} + 15G_6 + \dots, \end{aligned}$$

where $\dots$ denotes terms of order z^2 and higher (all expansions involve only even powers). Now

$$a\wp^3 + b\wp^2 + c\wp + d = \frac{a}{z^6} + \frac{b}{z^4} + \frac{9aG_4 + c}{z^2} + (15aG_6 + 6bG_4 + d) + \dots.$$

Comparing with $(\wp')^2 = \frac{4}{z^6} + \frac{0}{z^4} - \frac{24G_4}{z^2} - 80G_6 + \dots$:

$$\begin{aligned} z^{-6}: \quad & a = 4, \\ z^{-4}: \quad & b = 0, \\ z^{-2}: \quad & 9aG_4 + c = -24G_4 \implies c = -24G_4 - 4 \cdot 9G_4 = -60G_4, \\ z^0: \quad & 15aG_6 + 6bG_4 + d = -80G_6 \implies d = -80G_6 - 4 \cdot 15G_6 = -140G_6. \end{aligned}$$

Hence $(\wp')^2 = 4\wp^3 - 60G_4\wp - 140G_6 = 4\wp^3 - g_2\wp - g_3$.

For the last claim: by Proposition 2.4.1, $4\prod_i(\wp(z) - e_i) = (\wp'(z))^2 = f(\wp(z))$ for all $z \notin L$. Since $\wp$ is nonconstant, it takes infinitely many values, and two polynomials agreeing at infinitely many points are equal. So $f(x) = 4(x - e_1)(x - e_2)(x - e_3)$, and the roots e_1, e_2, e_3 are distinct by Proposition 2.3.7. $\blacksquare$

Proposition 2.4.6

For every even $k \geq 4$ we have $G_k(L) \in \mathbb{Q}[G_4(L), G_6(L)]$, i.e. G_k is a polynomial in G_4 and G_6 with rational coefficients.

Proof. First we derive a differential equation for $\wp''$. Differentiating $(\wp')^2 = 4\wp^3 - g_2\wp - g_3$ gives $2\wp'\wp'' = 12\wp^2\wp' - g_2\wp'$, and dividing by $2\wp' \neq 0$,

$$\wp'' = 6\wp^2 - \frac{g_2}{2} = 6\wp^2 - 30G_4. \quad (2.4)$$

We use strong induction on even k . For $k = 4$ and $k = 6$ the claim is trivial. Let $k \geq 8$ be even and assume that $G_m \in \mathbb{Q}[G_4, G_6]$ for all even m with $4 \leq m \leq k - 2$.

By Lemma 2.4.4, $\wp(z) = \frac{1}{z^2} + \sum_{\substack{m \geq 4 \\ m \text{ even}}} (m-1)G_m z^{m-2}$. Differentiating twice,

$$\wp''(z) = \frac{6}{z^4} + \sum_{\substack{m \geq 4 \\ m \text{ even}}} (m-1)(m-2)(m-3)G_m z^{m-4},$$

so the coefficient of z^{k-4} in $\wp''$ is $(k-1)(k-2)(k-3)G_k$.

On the other hand,

$$6\wp(z)^2 - 30G_4 = 6\left(\frac{1}{z^2} + \sum_{\substack{m \geq 4 \\ m \text{ even}}} (m-1)G_m z^{m-2}\right)\left(\frac{1}{z^2} + \sum_{\substack{m \geq 4 \\ m \text{ even}}} (m-1)G_m z^{m-2}\right) - 30G_4.$$

For $k \geq 8$ the constant $-30G_4$ only contributes to z^0 , not to z^{k-4} . The coefficient of z^{k-4} in the product comes from:

- (1) $\frac{1}{z^2} \cdot (k-1)G_k z^{k-2} = (k-1)G_k z^{k-4}$; this occurs twice (from the first factor times the second, and vice versa), giving $2(k-1)G_k$;
- (2) products of two terms $(i-1)G_i z^{i-2} \cdot (j-1)G_j z^{j-2}$ with $i+j-4 = k-4$, i.e. $i+j = k$, $i, j \geq 4$ even, giving $\sum_{i+j=k} (i-1)(j-1)G_i G_j$.

So the coefficient of z^{k-4} in $6\wp^2 - 30G_4$ is $6(2(k-1)G_k + \sum_{i+j=k, i,j \geq 4} (i-1)(j-1)G_i G_j)$. By (2.4) the two coefficients are equal:

$$(k-1)(k-2)(k-3)G_k = 12(k-1)G_k + 6 \sum_{\substack{i+j=k \\ i,j \geq 4}} (i-1)(j-1)G_i G_j.$$

Subtracting $12(k-1)G_k$ from both sides,

$$\begin{aligned} (k-1)[(k-2)(k-3) - 12]G_k &= (k-1)(k^2 - 5k - 6)G_k = (k-1)(k-6)(k+1)G_k \\ &= 6 \sum_{\substack{i+j=k \\ i,j \geq 4}} (i-1)(j-1)G_i G_j, \end{aligned}$$

and since $k \geq 8$, the factor $(k-1)(k-6)(k+1)$ is nonzero, so

$$G_k = \frac{6}{(k-1)(k-6)(k+1)} \sum_{\substack{i+j=k \\ i,j \geq 4}} (i-1)(j-1)G_iG_j.$$

Since $i+j=k$ and $i,j \geq 4$, we have $i,j \leq k-4 < k-2$; by the induction hypothesis $G_i, G_j \in \mathbb{Q}[G_4, G_6]$. Hence $G_k \in \mathbb{Q}[G_4, G_6]$. ■

Correction. In my handwritten notes the differential equation was written as “ $\wp'' = 6\wp^2 - 30G_4 + \dots$ ”; it is in fact an exact identity, see (2.4). At the end, “ $G_k \in \mathbb{Q}(G_2, G_4)$ ” should read $G_k \in \mathbb{Q}[G_4, G_6]$.

Example 2.4.7. For $k=8$: $G_8 = \frac{6}{7 \cdot 2 \cdot 9} \cdot 3 \cdot 3 G_4^2 = \frac{3}{7} G_4^2$. (The only term is $i=j=4$.)

2.5 The torus $\mathbb{C}/L$ as an elliptic curve

The differential equation $(\wp'(z))^2 = f(\wp(z))$, $f(x) = 4x^3 - g_2x - g_3$, has a geometric interpretation. Consider the map

$$\mathbb{C}/L \longrightarrow \mathbb{P}^2(\mathbb{C}), \quad z \longmapsto \begin{cases} [\wp(z) : \wp'(z) : 1] & z \notin L, \\ [0 : 1 : 0] & z \in L. \end{cases}$$

It is well defined on $\mathbb{C}/L$ since $\wp, \wp'$ are L -periodic. The value at $z \in L$ is forced by continuity: as $z \rightarrow 0$, $\wp(z), \wp'(z) \rightarrow \infty$, but in $\mathbb{P}^2(\mathbb{C})$ we may rescale by the nonzero factor $1/\wp'(z)$:

$$[\wp(z) : \wp'(z) : 1] = \left[\frac{\wp(z)}{\wp'(z)} : 1 : \frac{1}{\wp'(z)} \right] \longrightarrow [0 : 1 : 0],$$

because $\frac{\wp(z)}{\wp'(z)} = \frac{z^{-2} + \dots}{-2z^{-3} + \dots} = -\frac{z}{2} + \dots \rightarrow 0$ and $\frac{1}{\wp'(z)} \rightarrow 0$.

Proposition 2.5.1

The image of the map $\mathbb{C}/L \rightarrow \mathbb{P}^2(\mathbb{C})$ above lies on the elliptic curve

$$E : Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3. \quad (2.5)$$

Proof. We write the Weierstrass equation $y^2 = 4x^3 - g_2x - g_3$ and move it into $\mathbb{P}^2$ by homogenisation: we replace (x, y) by $[X : Y : Z]$ via $x = \frac{X}{Z}$, $y = \frac{Y}{Z}$. We obtain

$$\left(\frac{Y}{Z}\right)^2 = 4\left(\frac{X}{Z}\right)^3 - g_2\left(\frac{X}{Z}\right) - g_3,$$

and multiplying both sides by Z^3 gives (2.5), the homogeneous equation of the elliptic curve in $\mathbb{P}^2(\mathbb{C})$. Now:

- For $z \notin L$ we have $X = \wp(z)$, $Y = \wp'(z)$, $Z = 1$, so (2.5) reads $(\wp'(z))^2 \cdot 1 = 4\wp(z)^3 - g_2\wp(z) \cdot 1^2 - g_3 \cdot 1^3$, which is Theorem 2.4.5.
- For $z \in L$ we have $X = 0$, $Y = 1$, $Z = 0$, and (2.5) reads $0 = 0$. ■

Theorem 2.5.2

The map

$$F : \mathbb{C}/L \longrightarrow E, \quad F(z) = [\wp(z) : \wp'(z) : 1] \ (z \notin L), \quad F(0) = [0 : 1 : 0] = O,$$

is an analytic bijection from the torus $\mathbb{C}/L$ onto the elliptic curve $E : y^2 = 4x^3 - g_2x - g_3$ in $\mathbb{P}^2(\mathbb{C})$ (including its point at infinity O).

Proof. Let E be the set of points of $Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$ in $\mathbb{P}^2(\mathbb{C})$; as explained after Definition 2.1.9, these are the points $[x_0 : y_0 : 1]$ with $y_0^2 = f(x_0)$ together with $O = [0 : 1 : 0]$. By Proposition 2.5.1, F maps $\mathbb{C}/L$ into E .

Surjectivity. Let $P \in E$ be arbitrary. If $P = O$, then $F(0) = P$. Otherwise $P = [x_0 : y_0 : 1]$ is a finite point; we need $z \in \mathbb{C}/L$, $z \neq 0$, with $\wp(z) = x_0$ and $\wp'(z) = y_0$.

- (1) The function $\wp - x_0$ is elliptic with exactly one pole modulo L , of order 2. By Proposition 2.2.13, it has exactly 2 zeros modulo L (counted with multiplicity); in particular it has at least one. Let z_1 be such a zero; then $z_1 \notin L$ and $\wp(z_1) = x_0$.
- (2) Since Theorem 2.4.5 holds at z_1 , $(\wp'(z_1))^2 = f(\wp(z_1)) = f(x_0)$.
- (3) Since P is on the curve, $y_0^2 = f(x_0)$.
- (4) Thus $(\wp'(z_1))^2 = y_0^2$, so $\wp'(z_1) = \pm y_0$.
- (5) If $\wp'(z_1) = y_0$, choose $z = z_1$. If $\wp'(z_1) = -y_0$, use that $\wp$ is even and $\wp'$ is odd: $\wp(-z_1) = \wp(z_1) = x_0$ and $\wp'(-z_1) = -\wp'(z_1) = -(-y_0) = y_0$; choose $z = -z_1$.
- (6) In both cases $z \notin L$, $\wp(z) = x_0$ and $\wp'(z) = y_0$, so $F(z) = P$.

Injectivity. Let $z_1, z_2 \in \mathbb{C}/L$ with $F(z_1) = F(z_2)$.

- (1) If $F(z_1) = F(z_2) = O = [0 : 1 : 0]$: for $z \notin L$ the point $F(z)$ has third coordinate 1 $\neq 0$, so it is not O . Hence $z_1, z_2 \in L$, i.e. $z_1 = z_2 = 0$ in $\mathbb{C}/L$.
- (2) If $F(z_1) = F(z_2) = [x_0 : y_0 : 1]$, then $z_1, z_2 \notin L$, $\wp(z_1) = \wp(z_2) = x_0$ and $\wp'(z_1) = \wp'(z_2) = y_0$. The function $\wp - x_0$ has exactly two zeros modulo L (with multiplicity), and z_1 and $-z_1$ are zeros (as $\wp$ is even). If $z_1 \not\equiv -z_1$, these are the two zeros; if $z_1 \equiv -z_1$, then z_1 is a half-period and it is a double zero by Proposition 2.3.7. In either case $z_2 \equiv z_1$ or $z_2 \equiv -z_1 \pmod{L}$.
 - If $z_2 \equiv z_1$, we are done.
 - If $z_2 \equiv -z_1$, then $y_0 = \wp'(z_2) = \wp'(-z_1) = -\wp'(z_1) = -y_0$, so $y_0 = 0$ and $\wp'(z_1) = 0$. By Corollary 2.3.6, z_1 is a half-period, so $2z_1 \in L$, i.e. $-z_1 \equiv z_1$. Hence $z_2 \equiv -z_1 \equiv z_1$.

Thus $z_1 = z_2$ in $\mathbb{C}/L$, and F is injective.

Analyticity. Near a point $z_0 \notin L$, the map is given by $z \mapsto [\wp(z) : \wp'(z) : 1]$, whose coordinates are holomorphic functions. Near a lattice point, it is given by $z \mapsto [\frac{\wp(z)}{\wp'(z)} : 1 : \frac{1}{\wp'(z)}]$, whose coordinates are holomorphic near $z \in L$ (they have removable singularities there with value 0, since $\wp'$ has a pole). ■

Correction. In the injectivity part of my handwritten notes, the case $z_2 = -z_1$, $\wp'(z_1) = 0$ was “dismissed since $z_1, z_2 \in \mathbb{C}/L$ ”. The right conclusion is not that this case cannot occur

(it does occur: for $z_1 = \frac{\omega_1}{2}$), but that in this case z_1 is a half-period, so that $-z_1 \equiv z_1$ and therefore $z_2 = z_1$ in $\mathbb{C}/L$. Also, the statement “the equation $\wp(z) = x_0$ must have exactly 2 solutions” (which I had marked with a question mark) should be understood as: $\wp - x_0$ has exactly 2 zeros modulo L *counted with multiplicity*, by Proposition 2.2.13, since its only pole modulo L is the double pole at 0.

Remark 2.5.3. The bijection F transports the group structure of $\mathbb{C}/L$ to E ; the resulting group law on E is the usual chord-and-tangent addition law with neutral element O [2, §I.6]. Conversely, one can show that every elliptic curve $y^2 = 4x^3 - ax - b$ over $\mathbb{C}$ with $a^3 - 27b^2 \neq 0$ arises in this way from some lattice L (the *uniformisation theorem*); this uses the modular function j and will be discussed after the theory of modular forms of level 1.

Chapter 3

The Modular Group and Congruence Subgroups

In this chapter we study the full modular group $\mathrm{SL}_2(\mathbb{Z})$, its action on the upper half-plane $\mathfrak{H} = \{z \in \mathbb{C} : \mathrm{Im} z > 0\}$, and its congruence subgroups $\Gamma(N)$, $\Gamma_1(N)$, $\Gamma_0(N)$: their indices, fundamental domains, generators and cusps. Modular forms of congruence subgroups will be functions on $\mathfrak{H}$ with a transformation law under these groups. We follow [2, §III.1]; the last section contains my solutions to the exercises of that section.

3.1 $\mathrm{SL}_2(\mathbb{Z})$ and its action on the upper half-plane

Definition 3.1.1 (General linear group)

For any commutative ring R , the *general linear group* is

$$\mathrm{GL}_2(R) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2 \times 2}(R) \mid ad - bc \in R^\times \right\},$$

where $R^\times$ is the multiplicative group of invertible elements of R .

Proposition 3.1.2

$\mathrm{GL}_2(R)$ is a group under matrix multiplication.

Proof. Matrix multiplication is associative, and $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(R)$ since $\det I = 1 \in R^\times$. Since R is commutative, $\det(AB) = \det(A)\det(B)$, and a product of units is a unit; so $\mathrm{GL}_2(R)$ is closed under multiplication. Finally, for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $\Delta = ad - bc \in R^\times$, put

$$A^{-1} = \Delta^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

A direct computation gives $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = \Delta I$, and similarly in the other order, so $AA^{-1} = A^{-1}A = I$. Also $\det(A^{-1}) = \Delta^{-2} \cdot \Delta = \Delta^{-1} \in R^\times$, so $A^{-1} \in \mathrm{GL}_2(R)$. ■

Definition 3.1.3 (Special linear group)

The *special linear group* $\mathrm{SL}_2(R)$ is the subgroup of $\mathrm{GL}_2(R)$ consisting of the matrices of determinant 1 (it is the kernel of the homomorphism $\det : \mathrm{GL}_2(R) \rightarrow R^\times$).

We write

$$\Gamma = \mathrm{SL}_2(\mathbb{Z})$$

and call it the *full modular group*. We put $\bar{\Gamma} = \Gamma/\{\pm I\}$, and more generally, for a subgroup $G \subseteq \Gamma$,

$$\bar{G} = \begin{cases} G/\{\pm I\} & \text{if } -I \in G, \\ G & \text{if } -I \notin G, \end{cases}$$

which we regard as a subgroup of $\bar{\Gamma}$ (the image of G in $\bar{\Gamma}$). We also use the two special matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad Sz = -\frac{1}{z}, \quad Tz = z + 1.$$

Proposition 3.1.4

The group $\mathrm{GL}_2^+(\mathbb{R}) = \{\gamma \in \mathrm{GL}_2(\mathbb{R}) : \det \gamma > 0\}$, and in particular Γ , acts on $\mathfrak{H}$ by Möbius transformations

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

Moreover $\mathrm{Im}(\gamma z) = \frac{\det \gamma \cdot \mathrm{Im} z}{|cz + d|^2}$, and $-I$ acts trivially, so $\bar{\Gamma}$ acts on $\mathfrak{H}$.

Proof. Let $z \in \mathfrak{H}$ and $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R})$. First, $cz + d \neq 0$: if $c \neq 0$ then $\mathrm{Im}(cz + d) = c \mathrm{Im} z \neq 0$, and if $c = 0$ then $d \neq 0$ (as $ad = \det \gamma > 0$). Next,

$$\gamma z = \frac{(az + b)(c\bar{z} + d)}{|cz + d|^2} = \frac{ac|z|^2 + adz + bc\bar{z} + bd}{|cz + d|^2},$$

and the imaginary part of the numerator is $ad \mathrm{Im} z + bc \mathrm{Im} \bar{z} = (ad - bc) \mathrm{Im} z$. Hence $\mathrm{Im}(\gamma z) = \det \gamma \cdot \mathrm{Im} z / |cz + d|^2 > 0$, i.e. $\gamma z \in \mathfrak{H}$.

Clearly $Iz = z$. For $\gamma_1, \gamma_2 \in \mathrm{GL}_2^+(\mathbb{R})$ one checks $\gamma_1(\gamma_2 z) = (\gamma_1 \gamma_2)z$: writing $\gamma_2 z = \frac{a_2 z + b_2}{c_2 z + d_2}$,

$$\gamma_1(\gamma_2 z) = \frac{a_1 \frac{a_2 z + b_2}{c_2 z + d_2} + b_1}{c_1 \frac{a_2 z + b_2}{c_2 z + d_2} + d_1} = \frac{(a_1 a_2 + b_1 c_2)z + (a_1 b_2 + b_1 d_2)}{(c_1 a_2 + d_1 c_2)z + (c_1 b_2 + d_1 d_2)} = (\gamma_1 \gamma_2)z.$$

So we have a group action. Finally $(-I)z = \frac{-z}{-1} = z$; more generally γ and $-\gamma$ act in the same way, so the action of Γ factors through $\bar{\Gamma}$. ■

Cusps

The same formula defines an action of Γ on $\mathbb{Q} \cup \{\infty\} = \mathbb{P}^1(\mathbb{Q})$, with the usual conventions $\gamma \infty = \frac{a}{c}$ (and $= \infty$ if $c = 0$), and $\gamma(-\frac{d}{c}) = \infty$. We put

$$\mathfrak{H}^* = \mathfrak{H} \cup \{\infty\} \cup \mathbb{Q}.$$

The points of $\mathbb{Q} \cup \{\infty\}$ are called *cusps*: ∞ is the “point at infinity” $i\infty$, and the rational numbers lie on the boundary $\mathbb{R}$ of $\mathfrak{H}$.

Proposition 3.1.5

Γ acts transitively on $\mathbb{Q} \cup \{\infty\}$. More precisely, for every $\frac{a}{c} \in \mathbb{Q}$ in lowest terms there is $\gamma \in \Gamma$ with $\gamma \infty = \frac{a}{c}$.

Proof. Let $\frac{a}{c}$ with $\gcd(a, c) = 1$. By Bézout’s theorem there are $b, d \in \mathbb{Z}$ with $ad - bc = 1$. Then $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $\gamma \infty = \frac{a}{c}$. Hence every cusp is in the orbit of ∞ , and the action is transitive. ■

So Γ has only one cusp up to equivalence (“ Γ has one cusp, at ∞ ”). A subgroup $\Gamma' \subseteq \Gamma$ still permutes $\mathbb{Q} \cup \{\infty\}$, but in general not transitively; its *cusps* are the Γ' -equivalence classes of $\mathbb{Q} \cup \{\infty\}$ (Section 3.4).

3.2 Congruence subgroups

Definition 3.2.1 (Congruence subgroups)

Let N be a positive integer. We define

$$\begin{aligned}\Gamma(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid a \equiv d \equiv 1, b \equiv c \equiv 0 \pmod{N} \right\} = \left\{ \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}, \\ \Gamma_1(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid a \equiv d \equiv 1, c \equiv 0 \pmod{N} \right\} = \left\{ \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}, \\ \Gamma_0(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\} = \left\{ \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\}.\end{aligned}$$

Clearly $\Gamma(N) \subseteq \Gamma_1(N) \subseteq \Gamma_0(N) \subseteq \Gamma$, and $\Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \Gamma$. That these are subgroups follows from Proposition 3.2.4 below: they are the preimages, under the reduction homomorphism $\mathrm{SL}_2(\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, of the subgroups $\{I\}$, $\left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\}$ and $\left\{ \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \right\}$.

Example 3.2.2. For $N = 2$:

$$\begin{aligned}\Gamma(2) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid a, d \text{ odd}, b, c \text{ even} \right\} \\ &= \left\{ \begin{pmatrix} 2k+1 & 2l \\ 2m & 2n+1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid k, l, m, n \in \mathbb{Z} \right\} \\ &= \left\{ 2 \begin{pmatrix} k & l \\ m & n \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid k, l, m, n \in \mathbb{Z} \right\},\end{aligned}$$

and $\bar{\Gamma}(2) = \Gamma(2)/\{\pm I\}$, since $-I \in \Gamma(2)$.

Example 3.2.3. (1) $\begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \in \Gamma_0(2)$, since $ad - bc = 3 - 2 = 1$ and $c = 2 \equiv 0 \pmod{2}$.

(2) $\begin{pmatrix} 4 & 1 \\ 3 & 1 \end{pmatrix} \in \Gamma_1(3)$, since $ad - bc = 4 - 3 = 1$, $a = 4 \equiv 1 \pmod{3}$, $d = 1 \equiv 1 \pmod{3}$, and $c = 3 \equiv 0 \pmod{3}$. In general, the entries a, d of an element of $\Gamma_1(3)$ are of the form $\dots, -2, 1, 4, 7, \dots$.

Proposition 3.2.4

The reduction map

$$\phi : \mathrm{SL}_2(\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}), \quad \phi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a \bmod N & b \bmod N \\ c \bmod N & d \bmod N \end{pmatrix},$$

is a group homomorphism with $\ker \phi = \Gamma(N)$. Consequently $\Gamma(N)$ is a normal subgroup of Γ , and hence also of $\Gamma_1(N)$ and $\Gamma_0(N)$. Moreover $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$.

Proof. ϕ is well defined: if $ad - bc = 1$, then $(a \bmod N)(d \bmod N) - (b \bmod N)(c \bmod N) = 1 \bmod N$. Let

$$A = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}, \quad B = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Since reduction modulo N is a ring homomorphism $\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$, i.e. it respects sums and products,

$$\begin{aligned} \phi(AB) &= \phi \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix} = \begin{pmatrix} (a_1a_2 + b_1c_2) \bmod N & (a_1b_2 + b_1d_2) \bmod N \\ (c_1a_2 + d_1c_2) \bmod N & (c_1b_2 + d_1d_2) \bmod N \end{pmatrix} \\ &= \begin{pmatrix} a_1 \bmod N & b_1 \bmod N \\ c_1 \bmod N & d_1 \bmod N \end{pmatrix} \begin{pmatrix} a_2 \bmod N & b_2 \bmod N \\ c_2 \bmod N & d_2 \bmod N \end{pmatrix} = \phi(A)\phi(B). \end{aligned}$$

So ϕ is a group homomorphism. Its kernel is

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

By the lemma below, $\Gamma(N) \trianglelefteq \Gamma$; a normal subgroup of Γ is normal in every intermediate subgroup.

For the last claim consider $\psi : \Gamma_0(N) \rightarrow (\mathbb{Z}/N\mathbb{Z})^\times$, $\psi \begin{pmatrix} a & b \\ c & d \end{pmatrix} = d \bmod N$. This is well defined: $ad - bc = 1$ and $c \equiv 0$ give $ad \equiv 1 \pmod{N}$, so $d \bmod N$ is a unit. It is a homomorphism: for $A, B \in \Gamma_0(N)$ as above, the lower right entry of AB is $c_1b_2 + d_1d_2 \equiv d_1d_2 \pmod{N}$ since $c_1 \equiv 0$. Its kernel is $\{c \equiv 0, d \equiv 1\}$; for such matrices also $a \equiv ad \equiv 1$, so $\ker \psi = \Gamma_1(N)$, which is therefore normal in $\Gamma_0(N)$. ■

Lemma 3.2.5 (Recall)

Let $f : G \rightarrow G_1$ be a homomorphism of groups. Then $\ker f$ is a normal subgroup of G .

Proof. Since $f(e_G) = e_{G_1}$, $e_G \in \ker f$, so $\ker f \neq \emptyset$. Let $a, b \in \ker f$. Then

$$f(ab^{-1}) = f(a)f(b^{-1}) = f(a)f(b)^{-1} = e_{G_1}(e_{G_1})^{-1} = e_{G_1},$$

so $ab^{-1} \in \ker f$ and $\ker f \leq G$. Now let $a \in G$ and $h \in \ker f$. Then

$$f(aha^{-1}) = f(a)f(h)f(a^{-1}) = f(a)e_{G_1}f(a)^{-1} = e_{G_1}.$$

Thus $aha^{-1} \in \ker f$, and hence $\ker f \trianglelefteq G$. ■

Proposition 3.2.6

The map $\phi : \mathrm{SL}_2(\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective.

Proof. Let $\bar{A} = \begin{pmatrix} \bar{a} & \bar{b} \\ \bar{c} & \bar{d} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ be arbitrary. This means $\bar{a}\bar{d} - \bar{b}\bar{c} \equiv 1 \pmod{N}$. Choose lifts $a, b, c, d \in \mathbb{Z}$ with $a \equiv \bar{a}, b \equiv \bar{b}, c \equiv \bar{c}, d \equiv \bar{d} \pmod{N}$; we may and do choose $a > 0$ (replace a by $a + kN$ for large k). Then

$$ad - bc \equiv 1 \pmod{N}, \quad \text{i.e.} \quad ad - bc + Nk = 1 \quad \text{for some } k \in \mathbb{Z}.$$

In general $ad - bc \neq 1$, so we have to modify the lifts.

Step 1: a lift $b' \equiv b \pmod{N}$ with $\gcd(a, b') = 1$. By the fundamental theorem of arithmetic we can write $a = a_1a_2$ with coprime positive integers a_1, a_2 , where

- (1) a_1 is the product of the prime powers in the factorisation of a whose primes divide N ;

- (2) a_2 is the product of the prime powers in the factorisation of a whose primes do not divide N .

By construction $\gcd(a_2, N) = 1$ and $\gcd(a_1, a_2) = 1$.

We claim $\gcd(b, a_1) = 1$. Suppose, by way of contradiction, that a prime p divides both b and a_1 . Since $p \mid a_1$, we have $p \mid a$ and $p \mid N$ (by the definition of a_1). From $ad - bc + Nk = 1$ and $p \mid a$, $p \mid b$, $p \mid N$ we get $p \mid 1$, a contradiction.

Since $\gcd(N, a_2) = 1$, the Chinese remainder theorem gives $b' \in \mathbb{Z}$ with

$$b' \equiv b \pmod{N} \quad \text{and} \quad b' \equiv 1 \pmod{a_2}.$$

Then $b' = b + Nl$ for some $l \in \mathbb{Z}$. If p is a prime dividing a_1 , then $p \mid N$ and $p \nmid b$, so $p \nmid b + Nl = b'$; hence $\gcd(b', a_1) = 1$. Since $b' \equiv 1 \pmod{a_2}$, also $\gcd(b', a_2) = 1$. As $a = a_1 a_2$, we conclude $\gcd(b', a) = 1$.

Step 2: correcting c and d . By Bézout's theorem there are $x, y \in \mathbb{Z}$ with $ax - b'y = 1$. Put

$$\Delta = ad - b'c, \quad c' = c + y(1 - \Delta), \quad d' = d + x(1 - \Delta), \quad M = \begin{pmatrix} a & b' \\ c' & d' \end{pmatrix}.$$

Then

$$\begin{aligned} \det M &= ad' - b'c' = a(d + x(1 - \Delta)) - b'(c + y(1 - \Delta)) \\ &= ad - b'c + (ax - b'y)(1 - \Delta) = \Delta + 1 \cdot (1 - \Delta) = 1, \end{aligned}$$

so $M \in \mathrm{SL}_2(\mathbb{Z})$. Moreover, since $b' \equiv b \pmod{N}$,

$$\Delta = ad - b'c \equiv ad - bc \equiv 1 \pmod{N},$$

so $1 - \Delta \equiv 0 \pmod{N}$, and therefore $c' \equiv c \equiv \bar{c}$ and $d' \equiv d \equiv \bar{d} \pmod{N}$. Also $a \equiv \bar{a}$ and $b' \equiv \bar{b}$. Hence $\phi(M) = \bar{A}$, and ϕ is surjective. $\blacksquare$

Correction. In my handwritten notes this proposition was announced as “ ϕ is injective”; it is *surjective* (it is certainly not injective, its kernel being $\Gamma(N)$). We also need the lift a to be nonzero for the factorisation $a = a_1 a_2$ to make sense; this is why we chose $a > 0$.

Proposition 3.2.7

$[\Gamma : \Gamma(N)] < \infty$; more precisely $\Gamma/\Gamma(N) \cong \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, so $[\Gamma : \Gamma(N)] = |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|$.

Proof. The index $[\Gamma : \Gamma(N)]$ is the number of distinct left cosets of $\Gamma(N)$ in Γ , i.e. the order $|\Gamma/\Gamma(N)|$ of the quotient group. By Propositions 3.2.4 and 3.2.6, ϕ is a surjective homomorphism with $\ker \phi = \Gamma(N)$ and $\mathrm{Im} \phi = \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. By the first isomorphism theorem,

$$\Gamma/\ker \phi \cong \mathrm{Im} \phi \implies \Gamma/\Gamma(N) \cong \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

So $|\Gamma/\Gamma(N)| = |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})| \leq N^4 < \infty$. $\blacksquare$

Counting $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$

Lemma 3.2.8 (Chinese remainder theorem)

Let $N = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ be the prime factorisation of N . Then the map

$$\Phi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_k^{e_k}\mathbb{Z}, \quad \Phi(a \bmod N) = (a \bmod p_1^{e_1}, \dots, a \bmod p_k^{e_k}),$$

is a ring isomorphism. Consequently it induces isomorphisms

$$(\mathbb{Z}/N\mathbb{Z})^\times \cong \prod_{i=1}^k (\mathbb{Z}/p_i^{e_i}\mathbb{Z})^\times,$$

$$\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \cong \prod_{i=1}^k \mathrm{SL}_2(\mathbb{Z}/p_i^{e_i}\mathbb{Z}), \quad \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \cong \prod_{i=1}^k \mathrm{GL}_2(\mathbb{Z}/p_i^{e_i}\mathbb{Z}).$$

Proof. Φ is well defined since $p_i^{e_i} \mid N$. It respects sums:

$$\Phi(\bar{a} + \bar{b}) = ((a+b) \bmod p_1^{e_1}, \dots) = (a \bmod p_1^{e_1}, \dots) + (b \bmod p_1^{e_1}, \dots) = \Phi(\bar{a}) + \Phi(\bar{b}),$$

and products in the same way: $\Phi(\bar{a}\bar{b}) = \Phi(\bar{a})\Phi(\bar{b})$; and $\Phi(\bar{1}) = (1, \dots, 1)$.

Injective. Let $\bar{a} \in \ker \Phi$, i.e. $a \equiv 0 \pmod{p_i^{e_i}}$, i.e. $p_i^{e_i} \mid a$, for all $1 \leq i \leq k$. Since $\gcd(p_i^{e_i}, p_j^{e_j}) = 1$ for $i \neq j$, the product $p_1^{e_1} \cdots p_k^{e_k} = N$ divides a , i.e. $\bar{a} = 0$. So $\ker \Phi = \{0\}$ and Φ is injective.

Surjective. Both sides have $N = p_1^{e_1} \cdots p_k^{e_k}$ elements, and an injective map between finite sets of the same size is surjective.

A ring isomorphism maps units to units, which gives the first induced isomorphism. Applying Φ entrywise to 2×2 matrices gives a bijection $M_{2 \times 2}(\mathbb{Z}/N\mathbb{Z}) \rightarrow \prod_i M_{2 \times 2}(\mathbb{Z}/p_i^{e_i}\mathbb{Z})$ compatible with matrix multiplication and with determinants: $\bar{a}\bar{d} - \bar{b}\bar{c} = \bar{1}$ in $\mathbb{Z}/N\mathbb{Z}$ if and only if $ad - bc \equiv 1 \pmod{p_i^{e_i}}$ for all i . Restricting to determinant 1 (resp. unit determinant) gives the isomorphisms for SL_2 and GL_2 . ■

Proposition 3.2.9

Let $\varphi(N) = |(\mathbb{Z}/N\mathbb{Z})^\times|$ be Euler's totient function. Then

$$\varphi(N) = N \prod_{p \mid N} \left(1 - \frac{1}{p}\right) \quad \text{and} \quad |\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})| = \varphi(N) |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|.$$

Proof. The group of units $(\mathbb{Z}/N\mathbb{Z})^\times$ is the set of residues which have an inverse modulo N , i.e. $\{\bar{a} \in \mathbb{Z}/N\mathbb{Z} : \gcd(a, N) = 1\}$. For example, for $N = 6$, $(\mathbb{Z}/6\mathbb{Z})^\times = \{\bar{1}, \bar{5}\}$ since $\gcd(1, 6) = \gcd(5, 6) = 1$.

By Lemma 3.2.8, $|(\mathbb{Z}/N\mathbb{Z})^\times| = \prod_{i=1}^k |(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^\times|$. The ring $\mathbb{Z}/p^e\mathbb{Z}$ has the p^e elements $0, 1, \dots, p^e - 1$, and such an element is a unit iff it is coprime to p^e , i.e. not a multiple of p . The multiples of p among them are $0, p, 2p, \dots, (p^{e-1} - 1)p$, which are p^{e-1} elements. So

$$|(\mathbb{Z}/p^e\mathbb{Z})^\times| = p^e - p^{e-1} = p^e \left(1 - \frac{1}{p}\right),$$

and therefore

$$\varphi(N) = \prod_{i=1}^k p_i^{e_i} \left(1 - \frac{1}{p_i}\right) = \left(\prod_{i=1}^k p_i^{e_i}\right) \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right) = N \prod_{p \mid N} \left(1 - \frac{1}{p}\right).$$

For the second claim consider the determinant map

$$\det : \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \rightarrow (\mathbb{Z}/N\mathbb{Z})^\times, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc \bmod N.$$

It takes values in $(\mathbb{Z}/N\mathbb{Z})^\times$ by definition of GL_2 , and $\det(AB) = \det(A)\det(B)$, so it is a group homomorphism. Its kernel is $\{A \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) : \det A = 1\} = \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. It is surjective: for $u \in (\mathbb{Z}/N\mathbb{Z})^\times$, the matrix $A = \begin{pmatrix} u & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ has $\det A = u$. By the first isomorphism theorem,

$$\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) / \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \cong (\mathbb{Z}/N\mathbb{Z})^\times,$$

and hence $|\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})| = |(\mathbb{Z}/N\mathbb{Z})^\times| \cdot |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})| = \varphi(N) |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|$. ■

Proposition 3.2.10 (1) $[\Gamma : \Gamma(N)] = N^3 \prod_{p|N} \left(1 - \frac{1}{p^2}\right);$

(2) $[\Gamma_1(N) : \Gamma(N)] = N;$

(3) $[\Gamma_0(N) : \Gamma_1(N)] = \varphi(N) = N \prod_{p|N} \left(1 - \frac{1}{p}\right);$

(4) $[\Gamma : \Gamma_0(N)] = N \prod_{p|N} \left(1 + \frac{1}{p}\right).$

Proof. (1) By Proposition 3.2.7, $[\Gamma : \Gamma(N)] = |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|$, and by Lemma 3.2.8, $|\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})| = \prod_{p^e || N} |\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})|$. We claim

$$|\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})| = p^{3e} \left(1 - \frac{1}{p^2}\right).$$

Case $e = 1$. $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a finite field. The map $\mathrm{GL}_2(\mathbb{F}_p) \rightarrow \mathcal{B}$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (v_1, v_2) = ((\begin{smallmatrix} a \\ c \end{smallmatrix}), (\begin{smallmatrix} b \\ d \end{smallmatrix}))$ is a bijection onto the set $\mathcal{B}$ of ordered bases of $\mathbb{F}_p^2$ (a matrix is invertible iff its columns are linearly independent). We count $\mathcal{B}$: there are $|\mathbb{F}_p| \cdot |\mathbb{F}_p| = p^2$ vectors in $\mathbb{F}_p^2$.

- Choose v_1 : we cannot take 0, so there are $p^2 - 1$ choices.
- Choose v_2 : it must be linearly independent from v_1 , i.e. $v_2 \in \mathbb{F}_p^2 \setminus \mathrm{span}(v_1)$, and $|\mathrm{span}(v_1)| = p$; so there are $p^2 - p$ choices.

Therefore $|\mathrm{GL}_2(\mathbb{F}_p)| = (p^2 - 1)(p^2 - p)$, and by Proposition 3.2.9,

$$|\mathrm{SL}_2(\mathbb{F}_p)| = \frac{|\mathrm{GL}_2(\mathbb{F}_p)|}{\varphi(p)} = \frac{(p^2 - 1)(p^2 - p)}{p - 1} = (p^2 - 1)p = p^3 - p = p^3 \left(1 - \frac{1}{p^2}\right).$$

Case $e > 1$. Consider the reduction map

$$\lambda : \mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}), \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bmod p^e \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bmod p.$$

As in Proposition 3.2.4, λ is a homomorphism. It is surjective: by Proposition 3.2.6 every element of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ lifts to $\mathrm{SL}_2(\mathbb{Z})$, and reducing that lift modulo p^e gives a preimage. Its kernel is

$$\ker \lambda = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bmod p^e \in \mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod p \right\}.$$

We count it. Given residues $b, c, d \bmod p^e$ with $b \equiv c \equiv 0$ and $d \equiv 1 \pmod{p}$, there is a *unique* $a \bmod p^e$ with $ad - bc \equiv 1 \pmod{p^e}$, namely

$$a \equiv (1 + bc) d^{-1} \pmod{p^e};$$

here $d^{-1} \bmod p^e$ exists because $d \equiv 1 \pmod{p}$ implies $\gcd(d, p^e) = 1$. This a automatically satisfies $a \equiv (1 + 0) \cdot 1 = 1 \pmod{p}$. So $|\ker \lambda|$ is the number of choices of (b, c, d) :

- $b \equiv 0 \pmod{p}$: $b \in \{0, p, 2p, \dots, (p^{e-1} - 1)p\}$, which gives p^{e-1} choices;
- $c \equiv 0 \pmod{p}$: likewise p^{e-1} choices;
- $d \equiv 1 \pmod{p}$: $d \in \{1, 1 + p, 1 + 2p, \dots, 1 + (p^{e-1} - 1)p\}$, also p^{e-1} choices.

Thus $|\ker \lambda| = p^{e-1} \cdot p^{e-1} \cdot p^{e-1} = p^{3(e-1)}$. By the first isomorphism theorem, $\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})/\ker \lambda \cong \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, so

$$|\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})| = p^{3(e-1)} \cdot p^3 \left(1 - \frac{1}{p^2}\right) = p^{3e} \left(1 - \frac{1}{p^2}\right).$$

Therefore

$$[\Gamma : \Gamma(N)] = \prod_{p^e \parallel N} p^{3e} \left(1 - \frac{1}{p^2}\right) = \left(\prod_{p^e \parallel N} p^e\right)^3 \prod_{p \mid N} \left(1 - \frac{1}{p^2}\right) = N^3 \prod_{p \mid N} \left(1 - \frac{1}{p^2}\right).$$

(2) Consider $\beta : \Gamma_1(N) \rightarrow \mathbb{Z}/N\mathbb{Z}$ (additive group), $\beta\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = b \pmod{N}$. For $A, B \in \Gamma_1(N)$ we have $A \equiv \begin{pmatrix} 1 & b_1 \\ 0 & 1 \end{pmatrix}$ and $B \equiv \begin{pmatrix} 1 & b_2 \\ 0 & 1 \end{pmatrix} \pmod{N}$, so

$$AB \equiv \begin{pmatrix} 1 & b_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & b_1 + b_2 \\ 0 & 1 \end{pmatrix} \pmod{N},$$

i.e. $\beta(AB) = \beta(A) + \beta(B)$. The map is surjective, since $T^b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in \Gamma_1(N)$ and $\beta(T^b) = b \pmod{N}$. Its kernel is $\{a \equiv d \equiv 1, b \equiv c \equiv 0\} = \Gamma(N)$. By the first isomorphism theorem $\Gamma_1(N)/\Gamma(N) \cong \mathbb{Z}/N\mathbb{Z}$, so $[\Gamma_1(N) : \Gamma(N)] = N$.

(3) The homomorphism $\psi : \Gamma_0(N) \rightarrow (\mathbb{Z}/N\mathbb{Z})^\times$, $\gamma \mapsto d \pmod{N}$, from the proof of Proposition 3.2.4 has kernel $\Gamma_1(N)$. It is surjective: for $u \in (\mathbb{Z}/N\mathbb{Z})^\times$, the matrix $\begin{pmatrix} u^{-1} & 0 \\ 0 & u \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ lifts, by Proposition 3.2.6, to some $\gamma \in \Gamma$, which lies in $\Gamma_0(N)$ (its lower left entry is $\equiv 0$) and has $\psi(\gamma) = u$. So $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ and $[\Gamma_0(N) : \Gamma_1(N)] = \varphi(N)$.

(4) Using $[A : C] = [A : B][B : C]$ for $C \leq B \leq A$ and (1)–(3),

$$[\Gamma : \Gamma_0(N)] = \frac{[\Gamma : \Gamma(N)]}{[\Gamma_0(N) : \Gamma_1(N)][\Gamma_1(N) : \Gamma(N)]} = \frac{N^3 \prod_{p \mid N} (1 - \frac{1}{p^2})}{N \prod_{p \mid N} (1 - \frac{1}{p}) \cdot N} = N \prod_{p \mid N} \left(1 + \frac{1}{p}\right),$$

since $1 - \frac{1}{p^2} = (1 - \frac{1}{p})(1 + \frac{1}{p})$. ■

Notes. In my handwritten notes only (1) was proved, and (2), (3) were stated to follow “in the same way”, using $[A : B][B : C] = [A : C]$; we have written out the homomorphisms β and ψ , and added (4), which will be used repeatedly (e.g. in Exercise 3.4.7). In the proof of injectivity of Φ in Lemma 3.2.8, my notes said “by CRT”; what is used is only that a product of pairwise coprime divisors of a divides a .

Proposition 3.2.11

$\Gamma(N)$, $\Gamma_1(N)$ and $\Gamma_0(N)$ act on $\mathfrak{H}$ via Möbius transformations.

Proof. This is clear, since $\Gamma(N) \subseteq \Gamma_1(N) \subseteq \Gamma_0(N) \subseteq \Gamma$ and Γ acts on $\mathfrak{H}$ via Möbius transformations (Proposition 3.1.4); the restriction of a group action to a subgroup is an action. ■

Definition 3.2.12 (Congruence subgroup)

A subgroup $\Gamma' \subseteq \Gamma$ is called a *congruence subgroup of level N* if it contains $\Gamma(N)$. (Similarly, a subgroup of $\bar{\Gamma}$ is a congruence subgroup of level N if it contains $\bar{\Gamma}(N)$.) Every congruence subgroup has finite index in Γ , by Proposition 3.2.7.

3.3 Fundamental domains

Definition 3.3.1 (Fundamental domain)

Let $G \subseteq \bar{\Gamma}$ be a subgroup. A closed subset $F_G \subseteq \mathfrak{H}$ is a *fundamental domain* for G if

- (1) for every $z \in \mathfrak{H}$ there is $\gamma \in G$ with $\gamma z \in F_G$;
- (2) no two distinct points of the interior $\text{int}(F_G)$ are G -equivalent, i.e. for $z, z' \in \text{int}(F_G)$ with $z \neq z'$ and all $\gamma \in G$ we have $z' \neq \gamma z$.

We recall the following standard facts about the full modular group, see [2, §III.1], [4, Ch. VII, §1].

Theorem 3.3.2

Let $F = \{z \in \mathfrak{H} : |\text{Re } z| \leq \frac{1}{2}, |z| \geq 1\}$.

- (1) F is a fundamental domain for $\bar{\Gamma}$.
- (2) $\bar{\Gamma}$ is generated by (the images of) S and T .
- (3) If $z \in \text{int}(F)$, then the stabiliser of z in $\bar{\Gamma}$ is trivial.

Proof sketch. For $z \in \mathfrak{H}$, by Proposition 3.1.4, $\text{Im}(\gamma z) = \text{Im } z / |cz + d|^2$, and since $\{cz + d\}$ is a lattice, $|cz + d|$ attains a minimum over the bottom rows (c, d) of elements of Γ ; so there is γ in the subgroup $\langle S, T \rangle$ with $\text{Im}(\gamma z)$ maximal. Translating by a power of T we get $|\text{Re}(\gamma z)| \leq \frac{1}{2}$, and then $|\gamma z| \geq 1$, since otherwise $\text{Im}(S\gamma z) = \text{Im}(\gamma z) / |\gamma z|^2$ would be larger. This proves the first condition in (1), with $\gamma \in \langle S, T \rangle$. For the second condition and (3) one checks that if $z, \gamma z \in F$ with $\gamma \neq \pm I$, then $|cz + d| \leq 1$ forces $|c| \leq 1$, and a short case analysis shows that z is on the boundary of F . Finally (2) follows by applying (1) to a point $z_0 \in \text{int}(F)$: for $\gamma \in \Gamma$ choose $\delta \in \langle S, T \rangle$ with $\delta \gamma z_0 \in F$; by (ii) and (3), $\delta \gamma = \pm I$. ■

Figure 3.1 shows F and some of its translates γF ; the region labelled γF is the image of F under γ .

Proposition 3.3.3

Let $\Gamma' \subseteq \bar{\Gamma}$ be a subgroup of finite index n , and let $\bar{\Gamma} = \bigsqcup_{i=1}^n \alpha_i \Gamma'$ be the decomposition into distinct left cosets, with $\alpha_i \in \bar{\Gamma}$. Put

$$F' = \bigcup_{i=1}^n \alpha_i^{-1} F.$$

Then:

- (1) for every $z \in \mathfrak{H}$ there is $\gamma' \in \Gamma'$ with $\gamma' z \in F'$;
- (2) no two distinct points of $\bigcup_{i=1}^n \alpha_i^{-1} \text{int}(F)$ are Γ' -equivalent.

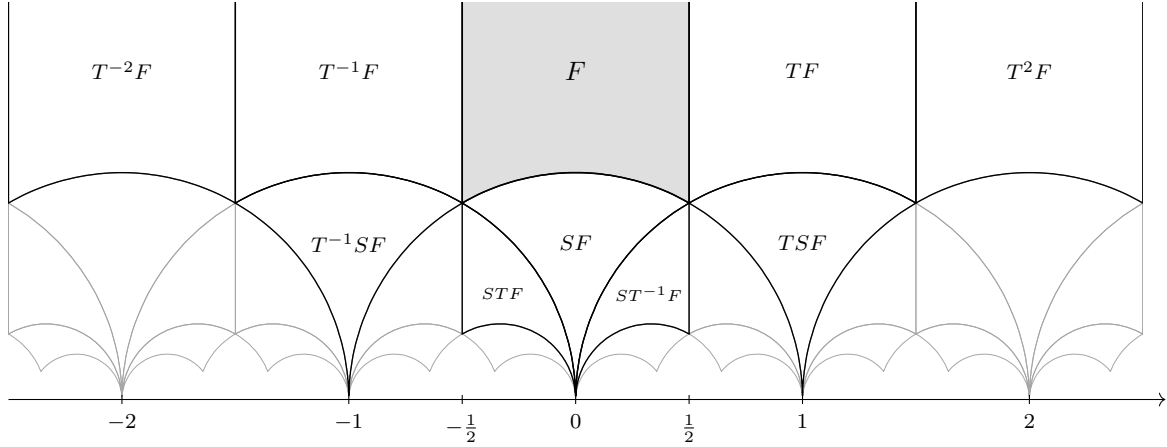


Figure 3.1: The fundamental domain F of $\bar{\Gamma}$ (shaded) and some of its translates γF .

In particular F' is a fundamental domain for Γ' : by Problem 3.4.4 below, (2) even holds for any two distinct points of $\text{int}(F')$.

Proof. (1) Let $z \in \mathfrak{H}$ be arbitrary. Since F is a fundamental domain for $\bar{\Gamma}$, there is $\gamma \in \bar{\Gamma}$ with $\gamma z \in F$. Since $\gamma \in \bar{\Gamma} = \bigcup_i \alpha_i \Gamma'$, there is i with

$$\gamma = \alpha_i \gamma' \quad (\gamma' \in \Gamma') \quad \implies \quad \gamma' = \alpha_i^{-1} \gamma.$$

Thus $\gamma' z = \alpha_i^{-1}(\gamma z) \in \alpha_i^{-1} F \subseteq F'$.

(2) Suppose $z \in \alpha_i^{-1} \text{int}(F)$, $z' \in \alpha_j^{-1} \text{int}(F)$, and $z' = \gamma' z$ with $\gamma' \in \Gamma'$. Put

$$w_1 = \alpha_i z \in \text{int}(F), \quad w_2 = \alpha_j z' \in \text{int}(F).$$

Substituting $z = \alpha_i^{-1} w_1$ and $z' = \alpha_j^{-1} w_2$ into $z' = \gamma' z$ gives

$$\alpha_j^{-1} w_2 = \gamma' \alpha_i^{-1} w_1 \quad \implies \quad w_2 = (\alpha_j \gamma' \alpha_i^{-1}) w_1.$$

Since $\alpha_j \gamma' \alpha_i^{-1} \in \bar{\Gamma}$ and $w_1, w_2 \in \text{int}(F)$, the fact that F is a fundamental domain for $\bar{\Gamma}$ forces $w_1 = w_2$. Then $\alpha_j \gamma' \alpha_i^{-1}$ fixes $w_1 \in \text{int}(F)$, so by Theorem 3.3.2(3), $\alpha_j \gamma' \alpha_i^{-1} = 1$ in $\bar{\Gamma}$, i.e. $\alpha_j \gamma' = \alpha_i$. Hence $\alpha_i \in \alpha_j \Gamma'$, so the cosets $\alpha_i \Gamma'$ and $\alpha_j \Gamma'$ coincide and $i = j$. Then $\gamma' = \alpha_j^{-1} \alpha_i = 1$, and $z' = \gamma' z = z$. ■

Notes. In my handwritten notes, (2) was stated for $z, z' \in \text{int}(F')$, and it was argued that then $w_1 = \alpha_i z$ and $w_2 = \alpha_j z'$ lie in $\text{int}(F)$ and that $i \neq j$. Neither is automatic: a point of $\text{int}(F')$ may lie on the common boundary of two adjacent translates $\alpha_i^{-1} F$ and $\alpha_j^{-1} F$ (so $\alpha_i z \in \partial F$), and $i = j$ is possible. The version above avoids this by working with $\bigcup_i \alpha_i^{-1} \text{int}(F)$, which differs from $\text{int}(F')$ only by finitely many arcs of boundary; and it uses Theorem 3.3.2(3) to handle the case $w_1 = w_2$. This is the precise meaning of “ F' is a fundamental domain” used in practice (and it is the key step in Problem 3.4.4). Note also that we must work in $\bar{\Gamma}$: in Γ , the element $-I$ fixes every point.

Example 3.3.4 (Fundamental domain for $\Gamma(2)$). By Proposition 3.2.10(1), $[\Gamma : \Gamma(2)] = 2^3(1 - \frac{1}{4}) = 6$, and since $-I \in \Gamma(2)$, also $[\bar{\Gamma} : \bar{\Gamma}(2)] = 6$. Reducing modulo 2, the six matrices

$$I, \quad T^{-1} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$T^{-1}S \equiv \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad ST^{-1} \equiv \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad T^{-1}ST^{-1} \equiv \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

are the six distinct elements of $\mathrm{SL}_2(\mathbb{F}_2)$, so they represent the six cosets of $\Gamma(2)$. As $\Gamma(2)$ is normal, left and right cosets agree, and the inverses of a complete set of representatives again form a complete set of representatives. Hence, by Proposition 3.3.3 (applied with α_i^{-1} running through this list),

$$F(2) = T^{-1}F \cup F \cup T^{-1}SF \cup SF \cup T^{-1}ST^{-1}F \cup ST^{-1}F$$

is a fundamental domain for $\bar{\Gamma}(2)$; see Figure 3.2 (left), where the regions are numbered in this order. This is the domain drawn in my handwritten notes. The index $[\Gamma : \Gamma(N)]$ is exactly the number of copies of F needed.

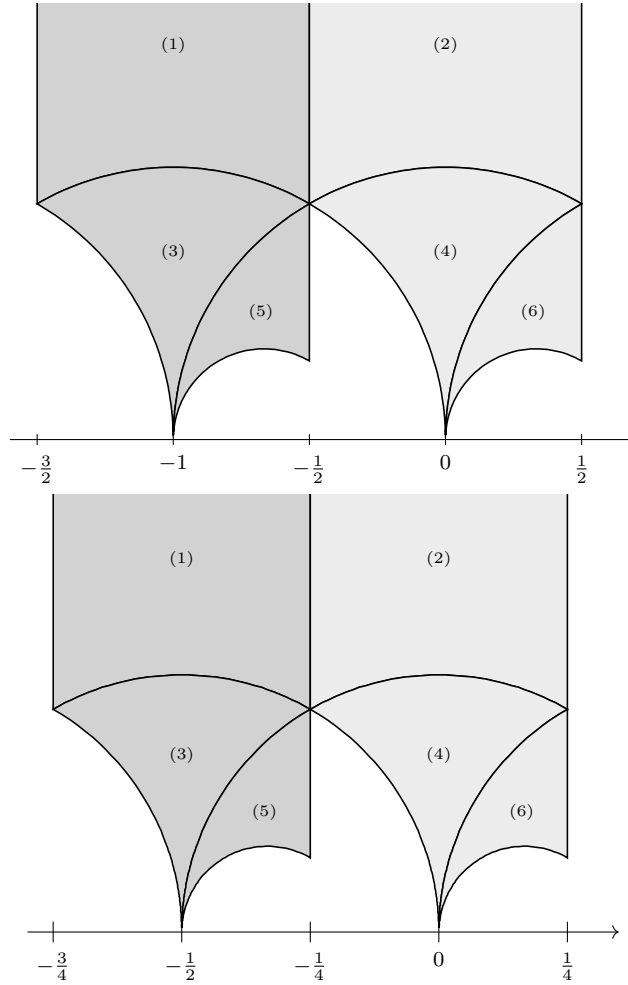


Figure 3.2: Left: the fundamental domain $F(2)$ for $\bar{\Gamma}(2)$ of Example 3.3.4; the regions (1)–(6) are $T^{-1}F$, F , $T^{-1}SF$, SF , $T^{-1}ST^{-1}F$, $ST^{-1}F$. Right: the fundamental domain $F_0(4) = M F(2) = \frac{1}{2}F(2)$ for $\bar{\Gamma}_0(4)$ (Problem 3.4.3).

3.4 Exercises from Koblitz, §III.1

Notes. At the end of the section my handwritten notes list the problems 8–12, 14, 16(b) and 17 (with 13 crossed out at first, and later done anyway). Below are my solutions, completed where the handwritten ones stopped. The numbering is that of [2, §III.1].

Report Problem 3.4.1 (Problem 8)

Find a simple isomorphism from $\Gamma(N)$ to a subgroup of $\Gamma_0(N^2)$ having index $\varphi(N)$ in $\Gamma_0(N^2)$. In particular, $\Gamma(2)$ and $\Gamma_0(4)$ are isomorphic.

Solution. Let $M = \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{Q})$, so $M^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/N \end{pmatrix}$, and define

$$\phi : \Gamma(N) \rightarrow \mathrm{GL}_2(\mathbb{Q}), \quad \phi(\gamma) = M\gamma M^{-1}.$$

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(N)$,

$$\phi(\gamma) = \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/N \end{pmatrix} = \begin{pmatrix} a & b \\ Nc & Nd \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/N \end{pmatrix} = \begin{pmatrix} a & b/N \\ Nc & d \end{pmatrix}.$$

Since $\gamma \in \Gamma(N)$, we have $a \equiv d \equiv 1$ and $b \equiv c \equiv 0 \pmod{N}$, so $b = kN$ and $c = jN$ for some $k, j \in \mathbb{Z}$. This implies $b/N = k \in \mathbb{Z}$ and $Nc = jN^2$, and therefore

$$\phi(\gamma) = \begin{pmatrix} a & k \\ jN^2 & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N^2}, \quad \det \phi(\gamma) = \det \gamma = 1.$$

So $\phi(\gamma) \in \Gamma_0(N^2)$, with moreover $A \equiv D \equiv 1 \pmod{N}$ for $\phi(\gamma) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Define

$$H = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_0(N^2) \mid D \equiv 1 \pmod{N} \right\}.$$

(For such matrices also $A \equiv 1 \pmod{N}$: since $C \equiv 0 \pmod{N}$, $1 = AD - BC \equiv AD \equiv A \pmod{N}$.)

H is a subgroup of $\Gamma_0(N^2)$. Clearly $I \in H$. For $M_1, M_2 \in H$, with $M_k = \begin{pmatrix} A_k & B_k \\ C_k & D_k \end{pmatrix}$, we have $A_1, A_2, D_1, D_2 \equiv 1 \pmod{N}$ and $C_1, C_2 \equiv 0 \pmod{N^2}$, and $\det M_2 = 1$, so

$$M_1 M_2^{-1} = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \begin{pmatrix} D_2 & -B_2 \\ -C_2 & A_2 \end{pmatrix} = \begin{pmatrix} A_1 D_2 - B_1 C_2 & -A_1 B_2 + B_1 A_2 \\ C_1 D_2 - D_1 C_2 & -C_1 B_2 + D_1 A_2 \end{pmatrix}.$$

Here $C_1 D_2 - D_1 C_2 \equiv 0 \pmod{N^2}$ and $-C_1 B_2 + D_1 A_2 \equiv 0 + 1 \cdot 1 = 1 \pmod{N}$. So $M_1 M_2^{-1} \in H$. Thus $\phi : \Gamma(N) \rightarrow H$, $\phi(A) = MAM^{-1}$.

ϕ is a homomorphism:

$$\begin{aligned} \phi(AB) &= M(AB)M^{-1} = M(AIB)M^{-1} = M(AM^{-1}MB)M^{-1} \\ &= (MAM^{-1})(MBM^{-1}) = \phi(A)\phi(B). \end{aligned}$$

ϕ is injective: if $\phi(A) = \phi(B)$, then $MAM^{-1} = MBM^{-1}$; multiplying by M^{-1} on the left and by M on the right gives $A = B$.

ϕ is surjective onto H : let $Y = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in H$, so $D \equiv 1 \pmod{N}$, $C \equiv 0 \pmod{N^2}$ and (as shown) $A \equiv 1 \pmod{N}$. Choose

$$X = \begin{pmatrix} A & NB \\ C/N & D \end{pmatrix}.$$

Then X has integer entries, $\det X = AD - BC = 1$, $A \equiv D \equiv 1$, $NB \equiv 0$, and $C/N \equiv 0 \pmod{N}$ (as $N^2 \mid C$); so $X \in \Gamma(N)$, and

$$\phi(X) = \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} A & NB \\ C/N & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{N} \end{pmatrix} = \begin{pmatrix} A & NB \\ C & ND \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{N} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = Y.$$

Therefore $\Gamma(N) \cong H \leq \Gamma_0(N^2)$.

The index. Define $\psi : \Gamma_0(N^2) \rightarrow (\mathbb{Z}/N\mathbb{Z})^\times$, $\psi\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = D \pmod{N}$.

- *Well defined*: $AD - BC = 1$ and $C \equiv 0 \pmod{N}$ give $AD \equiv 1 \pmod{N}$, so $D \pmod{N}$ is a unit.
- *Homomorphism*: for $X = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, Y = \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \in \Gamma_0(N^2)$, the lower right entry of XY is $C_1B_2 + D_1D_2 \equiv D_1D_2 \pmod{N}$, since $C_1 \equiv 0$. So $\psi(XY) = \psi(X)\psi(Y)$.
- *Surjective*: let $u \in (\mathbb{Z}/N\mathbb{Z})^\times$, represented by an integer u with $\gcd(u, N) = 1$, hence $\gcd(u, N^2) = 1$. By Bézout there are $x, y \in \mathbb{Z}$ with $ux - N^2y = 1$. Then the matrix $X = \begin{pmatrix} x & y \\ N^2 & u \end{pmatrix}$ has $\det X = ux - N^2y = 1$ and $N^2 \equiv 0 \pmod{N^2}$, so $X \in \Gamma_0(N^2)$ and $\psi(X) = u \pmod{N}$.
- *Kernel*: $\ker \psi = \{D \equiv 1 \pmod{N}\} = H$.

By the first isomorphism theorem $\Gamma_0(N^2)/H \cong (\mathbb{Z}/N\mathbb{Z})^\times$, so $[\Gamma_0(N^2) : H] = \varphi(N)$.

In particular, for $N = 2$ we have $\varphi(2) = 1$, so $[\Gamma_0(4) : H] = 1$, i.e. $H = \Gamma_0(4)$, and $\Gamma(2) \cong \Gamma_0(4)$ via $\gamma \mapsto M\gamma M^{-1}$, $M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. ■

Correction. In my handwritten notes the matrix $M = \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix}$ was said to lie in $\mathrm{SL}_2(\mathbb{R})$; its determinant is N , so $M \in \mathrm{GL}_2^+(\mathbb{Q})$. This does not matter for the argument, since conjugation by any invertible matrix is an injective homomorphism. Also, “ $c \equiv d \equiv 0 \pmod{N}$ ” should read $b \equiv c \equiv 0 \pmod{N}$.

Report Problem 3.4.2 (Problem 9)

Suppose that Γ_1 and Γ_2 are two subgroups of finite index in Γ , and $\Gamma_1 = \alpha\Gamma_2\alpha^{-1}$ for some $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. If F_2 is a fundamental domain for Γ_2 , prove that αF_2 is a fundamental domain for Γ_1 .

Solution. Recall that F is a fundamental domain for a group G acting on $\mathfrak{H}$ if (1) for all $z \in \mathfrak{H}$ there is $\gamma \in G$ with $\gamma z \in F$, and (2) for all $z, z' \in \mathrm{int}(F)$ with $z \neq z'$ and all $\gamma \in G$, $z' \neq \gamma z$. Since F_2 is a fundamental domain for Γ_2 :

- (i) for all $z \in \mathfrak{H}$ there is $\gamma \in \Gamma_2$ with $\gamma z \in F_2$;
- (ii) for all $z, z' \in \text{int}(F_2)$ with $z \neq z'$ and all $\gamma \in \Gamma_2$, $z' \neq \gamma z$.

We prove that αF_2 satisfies (1) and (2) for Γ_1 . Note that since $\det \alpha > 0$, α and α^{-1} map $\mathfrak{H}$ to itself (Proposition 3.1.4: $\text{Im}(\alpha^{-1}z) = \det(\alpha^{-1}) \text{Im } z / |cz + d|^2 > 0$), and $z \mapsto \alpha z$ is a homeomorphism of $\mathfrak{H}$, so $\text{int}(\alpha F_2) = \alpha \text{int}(F_2)$.

(1). Let $z \in \mathfrak{H}$ be arbitrary. Then $\alpha^{-1}z \in \mathfrak{H}$, and since F_2 is a fundamental domain for Γ_2 , there is $\gamma_2 \in \Gamma_2$ such that $\gamma_2(\alpha^{-1}z) \in F_2$. Thus

$$\alpha(\gamma_2(\alpha^{-1}z)) \in \alpha F_2 \implies (\alpha\gamma_2\alpha^{-1})z \in \alpha F_2.$$

We have $\alpha\Gamma_2\alpha^{-1} = \Gamma_1$ and $\gamma_2 \in \Gamma_2$; if we denote $\gamma_1 = \alpha\gamma_2\alpha^{-1} \in \Gamma_1$, we finally have $\gamma_1 z \in \alpha F_2$.

(2). Suppose, by way of contradiction, that there is $\gamma_1 \in \Gamma_1$ with $z' = \gamma_1 z$ for some $z, z' \in \text{int}(\alpha F_2)$, $z \neq z'$. Since $z, z' \in \text{int}(\alpha F_2) = \alpha \text{int}(F_2)$, there are $w, w' \in \text{int}(F_2)$ with

$$z = \alpha w, \quad z' = \alpha w' \implies w = \alpha^{-1}z, \quad w' = \alpha^{-1}z'. \quad (\text{i})$$

Note $w \neq w'$, since $w = w'$ would give $z = \alpha w = \alpha w' = z'$. Since $\gamma_1 \in \Gamma_1 = \alpha\Gamma_2\alpha^{-1}$,

$$\gamma_1 = \alpha\gamma_2\alpha^{-1} \quad \text{for some } \gamma_2 \in \Gamma_2. \quad (\text{ii})$$

Substituting (i) and (ii) into $z' = \gamma_1 z$:

$$\alpha w' = (\alpha\gamma_2\alpha^{-1})(\alpha w) \implies w' = \alpha^{-1}(\alpha\gamma_2\alpha^{-1})(\alpha w) = (\alpha^{-1}\alpha)\gamma_2(\alpha^{-1}\alpha)w = \gamma_2 w.$$

Hence $w \neq w'$ in $\text{int}(F_2)$ are Γ_2 -equivalent, contradicting (ii). ■

Correction. Koblitz states the problem for $\alpha \in \text{GL}_2(\mathbb{Q})$. For the action on $\mathfrak{H}$ we need $\det \alpha > 0$ (otherwise α maps $\mathfrak{H}$ to the lower half-plane), as was implicitly used in my solution; so $\alpha \in \text{GL}_2^+(\mathbb{Q})$. (Alternatively, if $\det \alpha < 0$ one can replace α by $\alpha \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, but then the fundamental domain changes accordingly.) In my handwritten solution, “there exists $\gamma_2 \in F_2$ ” should read $\gamma_2 \in \Gamma_2$.

Report Problem 3.4.3 (Problem 10)

Using the previous problem, draw a fundamental domain for $\Gamma_0(4)$.

Solution. From Problem 3.4.1 we have $\Gamma_0(4) = M\Gamma(2)M^{-1}$ with $M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, and hence the same for the images in $\bar{\Gamma}$. From Problem 3.4.2, if $\Gamma_1 = \alpha\Gamma_2\alpha^{-1}$ then $F_1 = \alpha F_2$ is a fundamental domain for Γ_1 . Thus a fundamental domain for $\Gamma_0(4)$ is

$$F_0(4) = M F(2),$$

with $F(2)$ from Example 3.3.4. For any $z \in \mathfrak{H}$ we have $Mz = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} z = \frac{1 \cdot z + 0}{0 \cdot z + 2} = \frac{z}{2}$, so $F_0(4)$ is obtained by shrinking every point of $F(2)$ by the factor $\frac{1}{2}$; see Figure 3.2 (right). The cusps $\infty, -1, 0$ of $F(2)$ become the cusps $\infty, -\frac{1}{2}, 0$ of $F_0(4)$. ■

Report Problem 3.4.4 (Problem 11)

Suppose that $\bar{\Gamma} = \bigcup_{j=1}^n \alpha_j \bar{\Gamma}'$, where $\bar{\Gamma}'$ is a subgroup of index n in $\bar{\Gamma}$. Let $F' = \bigcup_{j=1}^n \alpha_j^{-1} F$. Show that no two distinct points in the interior of F' are $\bar{\Gamma}'$ -equivalent.

Solution. Suppose, by way of contradiction, that there are $z_1 \neq z_2$ in $\text{int}(F')$ which are $\bar{\Gamma}'$ -equivalent; then there is $\bar{\gamma}' \in \bar{\Gamma}'$, $\bar{\gamma}' \neq 1$, with $z_2 = \bar{\gamma}' z_1$. Since $z_1, z_2 \in \text{int}(F')$, there exist $\varepsilon_1, \varepsilon_2 > 0$ such that

$$B_{\varepsilon_1}(z_1) \subseteq F' \quad \text{and} \quad B_{\varepsilon_2}(z_2) \subseteq F'.$$

Let $f(z) = \bar{\gamma}' z$. Since fractional linear transformations are homeomorphisms of $\mathfrak{H}$, $f^{-1}(B_{\varepsilon_2}(z_2))$ is open and contains z_1 . Define

$$U = B_{\varepsilon_1}(z_1) \cap f^{-1}(B_{\varepsilon_2}(z_2)),$$

a nonempty open set with $U \subseteq F'$ and $f(U) \subseteq F'$.

Let $B = \bigcup_{i=1}^n \alpha_i^{-1}(\partial F)$ be the union of the boundaries of the translates. It is a finite union of arcs of circles and segments of lines, hence closed with empty interior, and so is $f^{-1}(B)$. Hence the open set $U' = U \setminus (B \cup f^{-1}(B))$ is nonempty. Let $w \in U'$. Since $w \in F'$, we have $w \in \alpha_i^{-1} F$ for some i , and since $w \notin \alpha_i^{-1}(\partial F)$, in fact $w \in \alpha_i^{-1} \text{int}(F)$. In the same way $f(w) \in F' \setminus B$ gives $f(w) \in \alpha_j^{-1} \text{int}(F)$ for some j . By Proposition 3.3.3(2), $f(w) = w$.

So $\bar{\gamma}'$ fixes every point of the nonempty open set U' . But a fractional linear transformation $z \mapsto \frac{az+b}{cz+d}$ fixing infinitely many points satisfies $cz^2 + (d-a)z - b = 0$ for all of them, so $c = b = 0$ and $a = d$, i.e. $\bar{\gamma}' = 1$ in $\bar{\Gamma}$. This contradicts $\bar{\gamma}' \neq 1$. ■

Notes. My handwritten solution started exactly in this way (with the balls $B_{\varepsilon_1}(z_1)$, $B_{\varepsilon_2}(z_2)$ and the open set U) but stopped there; the remaining step is to move to a point of U which is not on any of the boundary arcs, and to apply Proposition 3.3.3(2) there.

Report Problem 3.4.5 (Problem 12)

Describe all congruence subgroups Γ' of level 2, i.e. all groups contained between Γ and $\Gamma(2)$: $\Gamma(2) \subseteq \Gamma' \subseteq \Gamma$. For each such Γ' , find a simply connected fundamental domain by taking a suitable part of the fundamental domain $F(2)$ for $\Gamma(2)$.

Solution. By the correspondence theorem, the groups between $\Gamma(2)$ and Γ correspond to the subgroups of $\Gamma/\Gamma(2) \cong \text{SL}_2(\mathbb{Z}/2\mathbb{Z})$ (Proposition 3.2.7), via $\Gamma' \mapsto \Gamma'/\Gamma(2)$.

$\text{SL}_2(\mathbb{F}_2) \cong S_3$. $\text{SL}_2(\mathbb{F}_2) = \text{GL}_2(\mathbb{F}_2)$ has $(4-1)(4-2) = 6$ elements, and it permutes the three nonzero vectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

of $\mathbb{F}_2^2$. This gives an injective homomorphism $\text{SL}_2(\mathbb{F}_2) \rightarrow S_3$ (a matrix fixing v_1 and v_2 is I), hence an isomorphism since both groups have order 6. Now

$$S_3 = \{e, (12), (13), (23), (123), (132)\}$$

has the two trivial subgroups $\{e\}$ and S_3 , and four nontrivial ones: three of order 2, namely $\langle (12) \rangle$, $\langle (13) \rangle$, $\langle (23) \rangle$, and one of order 3, namely $\langle (123) \rangle = \{e, (123), (132)\}$ (note $(132) = (123)^2$ generates the same subgroup). So, besides $\Gamma(2)$ and Γ , there are exactly 4 groups between $\Gamma(2)$ and Γ . We identify them by checking the action on v_1, v_2, v_3 .

(1) $\langle T \rangle \bmod 2$: since $T^2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \equiv I \pmod{2}$, we get $\langle T \rangle \bmod 2 = \{I, T\}$, of order 2. And

$$Tv_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = v_1, \quad Tv_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = v_3, \quad Tv_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v_2 \pmod{2},$$

so T swaps v_2, v_3 , and $\langle T \rangle \bmod 2$ corresponds to $\langle (23) \rangle$.

(2) $\langle S \rangle \bmod 2$: $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $S^2 \equiv I \pmod{2}$, so $\langle S \rangle \bmod 2 = \{I, S\}$. And $Sv_1 \equiv v_2, Sv_2 \equiv v_1, Sv_3 \equiv v_3$, so S swaps v_1, v_2 : this corresponds to $\langle (12) \rangle$.

(3) $\langle T^{-1}ST \rangle \bmod 2$: $T^{-1}ST \equiv \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $(T^{-1}ST)^2 \equiv \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \equiv I \pmod{2}$. And $T^{-1}STv_1 \equiv \begin{pmatrix} 1 \\ 1 \end{pmatrix} = v_3, T^{-1}STv_2 \equiv v_2, T^{-1}STv_3 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} = v_1$, so it swaps v_1, v_3 : this corresponds to $\langle (13) \rangle$.

(4) $\langle ST \rangle \bmod 2$: $ST \equiv \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $(ST)^2 \equiv \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $(ST)^3 \equiv \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \equiv I \pmod{2}$, so $\langle ST \rangle \bmod 2 = \{I, ST, (ST)^2\}$. And $STv_1 \equiv v_2, STv_2 \equiv v_3, STv_3 \equiv v_1$, i.e. $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_1$: this corresponds to $\langle (123) \rangle$.

Therefore the groups strictly between $\Gamma(2)$ and Γ are the preimages of these four subgroups:

$$\begin{aligned} \langle T \rangle \bmod 2 : \quad & \Gamma_0(2) = \{c \equiv 0 \pmod{2}\}, \\ \langle S \rangle \bmod 2 : \quad & \Gamma_\theta = \mathfrak{G}(2) = \{\gamma \equiv I \text{ or } \equiv S \pmod{2}\} = \{a \equiv d, b \equiv c \pmod{2}\}, \\ \langle T^{-1}ST \rangle \bmod 2 : \quad & \Gamma^0(2) = \{b \equiv 0 \pmod{2}\}, \\ \langle ST \rangle \bmod 2 : \quad & \text{the unique subgroup of index 2 containing } \Gamma(2). \end{aligned}$$

Fundamental domains. By Proposition 3.3.3, for left coset representatives α_i of Γ' we may take $\bigcup_i \alpha_i^{-1}F$; the following choices give simply connected domains made of regions of $F(2)$ (Figure 3.2, left) up to translation by $T^{\pm 1}$:

- (1) $\Gamma_0(2)$ (index 3): cosets $\{I, T\}, S\{I, T\} = \{S, ST\}, T^{-1}S\{I, T\}$; representatives $I, S, T^{-1}S$, and $F_0(2) = F \cup SF \cup S^{-1}TF = F \cup SF \cup STF$ (in $\bar{\Gamma}$). Alternatively, the representatives I, S, TS (note $TS \equiv \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ is in neither $\{I, T\}$ nor $\{S, ST\}$ modulo 2) give $F \cup SF \cup ST^{-1}F$, which is the right half (regions (2), (4), (6)) of $F(2)$, as in my notes.
- (2) $\mathfrak{G}(2)$ (index 3): cosets $\{I, S\}, T\{I, S\}, ST\{I, S\}$; representatives I, T, ST , and $F_{\mathfrak{G}} = F \cup T^{-1}F \cup T^{-1}SF$.
- (3) $\Gamma^0(2)$ (index 3): representatives I, T, S , and $F^0(2) = F \cup T^{-1}F \cup SF$.
- (4) $\langle ST \rangle$ -group (index 2): representatives I, T (since $T \not\equiv I, ST, (ST)^2$), and the fundamental domain $F \cup T^{-1}F$.

■

Notes. For (1) one can check the representatives directly: $T^{-1}S \equiv \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ is neither in

$\{I, T\}$ nor in $\{S, ST\} \equiv \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right\}$ modulo 2. The domain $F \cup SF \cup STF$ is also the one obtained in Problem 3.4.7 for $p = 2$, $e = 1$. Here Γ_θ is the *theta group*, which will reappear with the Jacobi theta function.

Report Problem 3.4.6 (Problem 13) (a) Prove that $\pm S$ and T^2 generate one of the subgroups of level 2 in the previous problem. That group was denoted $\mathfrak{G}(2)$ by Hecke.

(b) Prove that TST and T^2 generate $\bar{\Gamma}^0(2) \stackrel{\text{def}}{=} \left\{ \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \pmod{2} \right\}$; and that T and ST^2S generate $\bar{\Gamma}_0(2)$.

(c) Prove that the elements T^2 and $ST^{-2}S$ generate $\bar{\Gamma}(2)$.

(d) Prove that T and $ST^{-4}S$ generate $\bar{\Gamma}_0(4)$.

Solution. (a) Let $G = \langle S, T^2 \rangle \subseteq \Gamma$; note $-I = S^2 \in G$. Since $S \equiv S$ and $T^2 \equiv I \pmod{2}$, we have $G \subseteq \mathfrak{G}(2) = \{\gamma \equiv I \text{ or } S \pmod{2}\}$. We show $\mathfrak{G}(2) \subseteq G$ by induction on $|c|$, where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathfrak{G}(2)$.

Case $c = 0$. Then $ad = 1$, so $a = d = \pm 1$ and $\gamma = \pm \begin{pmatrix} 1 & \pm b \\ 0 & 1 \end{pmatrix}$. Since $c = 0$ is even, $\gamma \equiv I \pmod{2}$ (not $\equiv S$), so b is even and $\gamma = \pm T^{\pm b} \in G$.

Case $c \neq 0$. For $n \in \mathbb{Z}$,

$$T^{2n}\gamma = \begin{pmatrix} 1 & 2n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + 2nc & b + 2nd \\ c & d \end{pmatrix}.$$

Choose n such that $a' := a + 2nc$ satisfies $-|c| < a' \leq |c|$ (possible, since the residues modulo $2|c|$ can be represented in $(-|c|, |c|]$). We claim $|a'| < |c|$. Otherwise $a' = \pm c$, and since $\gcd(a', c) = 1$ (the determinant of $T^{2n}\gamma$ is 1), this forces $|c| = 1$ and a', c both odd. But $a' \equiv a \pmod{2}$, and $\gamma \equiv I \text{ or } S \pmod{2}$ means that exactly one of a, c is odd. This contradiction proves $|a'| < |c|$. Now

$$ST^{2n}\gamma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a' & b' \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ a' & b' \end{pmatrix}$$

lies in $\mathfrak{G}(2)$ (as $G \subseteq \mathfrak{G}(2)$) and has lower left entry a' with $|a'| < |c|$. By the induction hypothesis $ST^{2n}\gamma \in G$, hence $\gamma = T^{-2n}S^{-1}(ST^{2n}\gamma) \in G$.

So $\mathfrak{G}(2) = \langle S, T^2 \rangle$ (with $\pm$, i.e. $-I = S^2$ is included).

(b) *Conjugation lemma.* If a subgroup H of a group G is generated by x, y , then for all $g \in G$, gHg^{-1} is generated by gxg^{-1}, gyg^{-1} . Indeed, every $h \in H$ is a product $h = s_1s_2 \cdots s_k$ with $s_i \in \{x, y, x^{-1}, y^{-1}\}$, and

$$ghg^{-1} = g(s_1s_2 \cdots s_k)g^{-1} = g s_1(g^{-1}g)s_2(g^{-1}g) \cdots s_k g^{-1} = (gs_1g^{-1})(gs_2g^{-1}) \cdots (gs_kg^{-1}),$$

with $gs_ig^{-1} \in \{gxg^{-1}, gyg^{-1}, (gxg^{-1})^{-1}, (gyg^{-1})^{-1}\}$; conversely these elements lie in gHg^{-1} .

Claims. (1) $\Gamma^0(2) = T \mathfrak{G}(2) T^{-1}$; (2) $\Gamma_0(2) = (ST) \mathfrak{G}(2) (ST)^{-1}$.

Since $\Gamma(2)$ is normal in Γ , for $g \in \Gamma$ the group $g \mathfrak{G}(2) g^{-1}$ is the preimage of $g\{I, S\}g^{-1} \pmod{2}$. We compute:

$$TST^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \pmod{2},$$

which is the nontrivial element of $\Gamma^0(2)/\Gamma(2)$; this proves (1). Next, $ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, $(ST)^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$, and

$$\begin{aligned} (ST)S(ST)^{-1} &= \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \pmod{2}, \end{aligned}$$

the nontrivial element of $\Gamma_0(2)/\Gamma(2)$; this proves (2).

Using the conjugation lemma and $\mathfrak{G}(2) = \langle S, T^2 \rangle$:

$$\Gamma^0(2) = \langle TST^{-1}, TT^2T^{-1} \rangle = \langle TST^{-1}, T^2 \rangle = \langle TST, T^2 \rangle,$$

because $TST^{-1} \cdot T^2 = TST$ (so each pair generates the other). And

$$\Gamma_0(2) = \langle (ST)S(ST)^{-1}, (ST)T^2(ST)^{-1} \rangle = \left\langle \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix}, ST^2S^{-1} \right\rangle,$$

where $(ST)T^2(ST)^{-1} = ST^3T^{-1}S^{-1} = ST^2S^{-1}$, and

$$ST^2S^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

Now

$$\begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -1+2 & -1 \\ 2-2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = T^{-1},$$

so T lies in the group, and conversely $(ST)S(ST)^{-1} = T^{-1}(ST^2S^{-1})^{-1}$. Since $ST^2S = -ST^2S^{-1}$ (as $S^{-1} = -S$), in $\bar{\Gamma}$ we obtain $\bar{\Gamma}_0(2) = \langle T, ST^2S \rangle$.

(c) Put $V = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. We compute

$$ST^{-2}S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -2 & -1 \end{pmatrix} = -V,$$

so in $\bar{\Gamma}$ we must show $\bar{\Gamma}(2) = \langle T^2, V \rangle$. Let $G = \langle T^2, V, -I \rangle \subseteq \Gamma(2)$. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2)$ (a, d odd, b, c even) we use induction on $\max(|a|, |c|)$, using

$$T^{2n}\gamma = \begin{pmatrix} a+2nc & b+2nd \\ c & d \end{pmatrix}, \quad V^m\gamma = \begin{pmatrix} 1 & 0 \\ 2m & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c+2ma & d+2mb \end{pmatrix}.$$

- If $c = 0$, then $a = d = \pm 1$ and b is even, so $\gamma = \pm T^{\pm b} \in G$.
- If $c \neq 0$ and $|a| > |c|$: choose n with $-|c| < a+2nc \leq |c|$. Since $a+2nc$ is odd and c is even, $a+2nc \neq \pm c$, so $|a+2nc| < |c| < |a|$, and max decreases.
- If $|a| < |c|$ (equality is impossible, as a is odd and c even): choose m with $-|a| < c+2ma \leq |a|$. Since $c+2ma$ is even and a odd, $|c+2ma| < |a| < |c|$, and max decreases.

In each step we multiply γ on the left by an element of G and stay in $\Gamma(2)$; after finitely many steps we reach $c = 0$. Hence $\gamma \in G$, i.e. $\Gamma(2) = \langle T^2, V, -I \rangle$ and $\bar{\Gamma}(2) = \langle T^2, ST^{-2}S \rangle$.

(d) We already proved $\Gamma_0(4) = M\Gamma(2)M^{-1}$ with $M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ (Problem 3.4.1), and $\Gamma(2) = \langle T^2, V, -I \rangle$ by (c). By the conjugation lemma (with three generators), $\Gamma_0(4) = \langle MT^2M^{-1}, MV M^{-1}, -I \rangle$, and

$$MT^2M^{-1} = \begin{pmatrix} 1 & 2/2 \\ 0 & 1 \end{pmatrix} = T, \quad MV M^{-1} = \begin{pmatrix} 1 & 0 \\ 2 \cdot 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix} = -ST^{-4}S$$

(using the formula $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b/2 \\ 2c & d \end{pmatrix}$ from Problem 3.4.1, and the same computation as in (c) with -4 in place of -2). Hence $\bar{\Gamma}_0(4) = \langle T, ST^{-4}S \rangle$. ■

Notes. My handwritten solution of (a) wrote $g \in \mathfrak{G}(2)$ as a word $\pm S^{a_1} T^{b_1} ST^{b_2} S \dots ST^{b_l}$ (using Theorem 3.3.2(2)) and shortened the word using the relations $S^2 = (ST)^3 = 1$ in $\bar{\Gamma}$, e.g. $TST = ST^{-1}S$, and $T^2S \cdot (ST^{-1}S) = TS \cdot T^{-1} \dots$; the bookkeeping of that reduction was only sketched (“proven by induction”). The descent on $|c|$ above is a complete version of the same idea. For (b), my notes verified $\langle STST^{-1}S, ST^2S \rangle = \langle T, ST^2S \rangle$ in $\bar{\Gamma}$ via the computation $(ST^2S)^{-1}(STST^{-1}S)^{-1} = ST^{-2}SSTSTS = ST^{-2}TSTS = \dots = T$ using $(ST)^3 = 1$; the matrix computation above is equivalent. For (c), my notes suggested to multiply $g \in \Gamma(2)$ repeatedly by T^2 and $ST^{-2}S$ until it “shrinks”; the descent above makes this precise.

Report Problem 3.4.7 (Problem 14) (a) Prove that the following is a complete set of coset representatives $\{\alpha_j\}$ for $\Gamma_0(p^e)$ in Γ , i.e. $\Gamma = \bigsqcup_j \alpha_j \Gamma_0(p^e)$ is a disjoint union:

$$1; \quad T^{-k}S, \quad k = 0, 1, \dots, p^e - 1; \quad ST^{-kp}S, \quad k = 1, 2, \dots, p^{e-1} - 1.$$

(b) Use part (a) to draw a fundamental domain for $\Gamma_0(4)$.

(c) Using part (a), describe a fundamental domain for $\Gamma_0(p)$. Draw the fundamental domain for $\Gamma_0(3)$.

Solution. (a) *A criterion.* Let $N = p^e$. For $\gamma_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ and $\gamma_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$ in Γ we have $\gamma_1 \Gamma_0(N) = \gamma_2 \Gamma_0(N)$ iff $\gamma_2^{-1} \gamma_1 \in \Gamma_0(N)$. Since $\gamma_2^{-1} = \begin{pmatrix} d_2 & -b_2 \\ -c_2 & a_2 \end{pmatrix}$, the lower left entry of $\gamma_2^{-1} \gamma_1$ is $-c_2 a_1 + a_2 c_1$. Hence

$$\gamma_1 \Gamma_0(N) = \gamma_2 \Gamma_0(N) \iff a_2 c_1 - a_1 c_2 \equiv 0 \pmod{N}.$$

So the coset of γ only depends on its first column (a, c) .

First columns of the representatives.

$$1 : (a, c) = (1, 0); \quad T^{-k}S = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -k & -1 \\ 1 & 0 \end{pmatrix} : (a, c) = (-k, 1);$$

$$\begin{aligned} ST^{-kp}S &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -kp \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 \\ 1 & -kp \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -kp & -1 \end{pmatrix} : (a, c) = (-1, -kp). \end{aligned}$$

Disjointness. We check $a_2c_1 - a_1c_2 \not\equiv 0 \pmod{p^e}$ for each pair of distinct representatives γ_1, γ_2 :

- 1 and $T^{-k}S$: $(-k) \cdot 0 - 1 \cdot 1 = -1 \not\equiv 0$.
- 1 and $ST^{-kp}S$: $(-1) \cdot 0 - 1 \cdot (-kp) = kp$, and $0 < kp < p^e$ since $1 \leq k \leq p^{e-1} - 1$; so $kp \not\equiv 0$.
- $T^{-k}S$ and $T^{-k'}S$, $k \neq k'$: $(-k') \cdot 1 - (-k) \cdot 1 = k - k'$, and $0 < |k - k'| < p^e$.
- $T^{-k}S$ and $ST^{-k'}S$: $(-1) \cdot 1 - (-k)(-k'p) = -1 - kk'p \equiv -1 \pmod{p}$, so it is not divisible by p , let alone by p^e .
- $ST^{-kp}S$ and $ST^{-k'}S$, $k \neq k'$: $(-1)(-kp) - (-1)(-k'p) = (k - k')p$, and $0 < |k - k'|p < p^e$.

So the $1 + p^e + (p^{e-1} - 1) = p^e + p^{e-1}$ cosets $\alpha_j \Gamma_0(p^e)$ are pairwise distinct.

Completeness. By Proposition 3.2.10(4), $[\Gamma : \Gamma_0(p^e)] = p^e(1 + \frac{1}{p}) = p^e + p^{e-1}$. So we have found as many distinct cosets as there are cosets in total; hence $\Gamma = \bigsqcup_j \alpha_j \Gamma_0(p^e)$.

(b) For $p = 2$, $e = 2$ the representatives are 1; $T^{-k}S$ for $k = 0, 1, 2, 3$, i.e. $S, T^{-1}S, T^{-2}S, T^{-3}S$; and $ST^{-2}S$ ($k = 1$). Their inverses in $\bar{\Gamma}$ are 1, $S^{-1}T^k = ST^k$ and $S^{-1}T^2S^{-1} = ST^2S$. By Proposition 3.3.3,

$$F_0(4) = F \cup SF \cup STF \cup ST^2F \cup ST^3F \cup ST^2SF$$

is a fundamental domain for $\bar{\Gamma}_0(4)$ (6 copies of F , as $[\Gamma : \Gamma_0(4)] = 6$); see Figure 3.3 (right). Its cusps are $\infty, 0$ (the cusp of ST^kF) and $ST^2S\infty = ST^2(0) = S(2) = -\frac{1}{2}$.

(c) For $e = 1$ the third family is empty, and the representatives are 1 and $T^{-k}S$, $0 \leq k \leq p - 1$. Hence

$$F_0(p) = F \cup \bigcup_{k=0}^{p-1} ST^kF = F \cup SF \cup STF \cup \dots \cup ST^{p-1}F$$

($p + 1$ copies of F). Geometrically, ST^kF is the image of the translate T^kF under $z \mapsto -\frac{1}{z}$; all these regions have the cusp 0, and together they fill the region below F between the arc $|z| = 1$ and the cusp 0, on the left of $\operatorname{Re} z = \frac{1}{2}$. For $p = 3$:

$$F_0(3) = F \cup SF \cup STF \cup ST^2F,$$

see Figure 3.3 (left). ■

Correction. In my handwritten solution of (b), the last representative was taken to be T^3S instead of $ST^{-2}S$, and the resulting domain $F \cup SF \cup STF \cup ST^2F \cup ST^3F$ consists of only 5 copies of F , while $[\bar{\Gamma} : \bar{\Gamma}_0(4)] = 6$. The missing region is ST^2SF , which contains the cusp $-\frac{1}{2}$ ($\Gamma_0(4)$ has three cusps: $\infty, 0, \frac{1}{2} \sim -\frac{1}{2}$). Note that this $F_0(4)$ and the domain $MF(2)$ of Problem 3.4.3 are different, but both are fundamental domains for $\bar{\Gamma}_0(4)$.

Cusps of subgroups

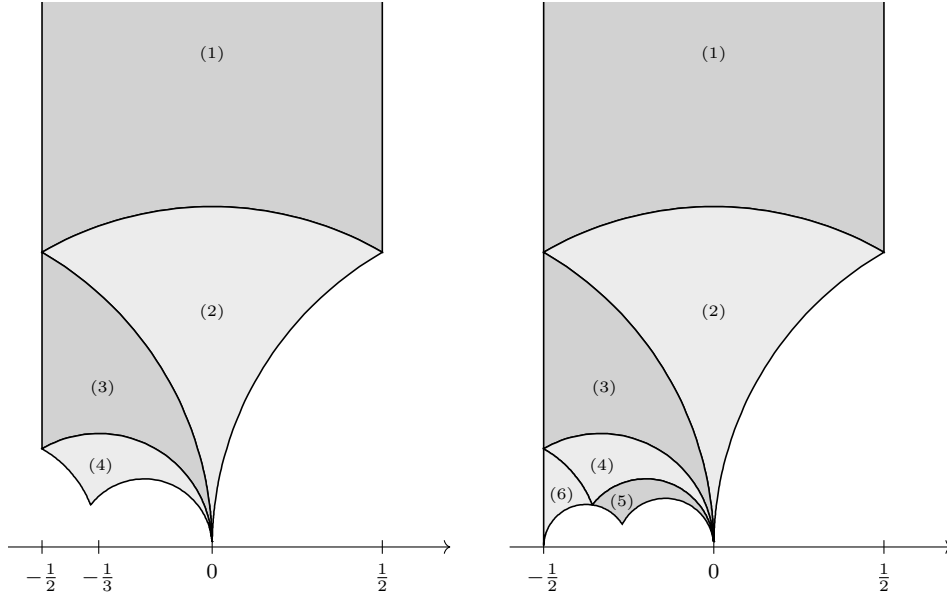


Figure 3.3: Left: $F_0(3) = F \cup SF \cup STF \cup ST^2F$, regions (1)–(4) in this order. Right: $F_0(4) = F \cup SF \cup STF \cup ST^2F \cup ST^3F \cup ST^2SF$, regions (1)–(6) in this order (Problem 3.4.7).

Report Problem 3.4.8 (Problem 16) (a) The assertion at the end of the section that $\Gamma(2)$ has 3 cusps (i.e. that there are three $\Gamma(2)$ -equivalence classes of cusps) does not immediately follow from the appearance of the fundamental domain, because boundary points of the fundamental domain *may* be $\Gamma(2)$ -equivalent. But prove directly that $\infty, 0, -1$ are $\Gamma(2)$ -inequivalent to one another.

(b) How many cusps does each of the congruence subgroups in Problem 3.4.5 have?

Solution. Recall that z, z' are Γ' -inequivalent if $z' \neq \gamma z$ for all $\gamma \in \Gamma'$. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$: $\gamma\infty = \frac{a}{c}$ (or ∞ if $c = 0$) and $\gamma 0 = \frac{b}{d}$ (or ∞ if $d = 0$).

(a) Suppose some two of them are $\Gamma(2)$ -equivalent; recall a, d are odd and b, c even.

- ∞ and 0 : there is $\gamma_1 \in \Gamma(2)$ with $\gamma_1\infty = \frac{a_1}{c_1} = 0$, so $a_1 = 0$, contradicting $a_1 \equiv 1 \pmod{2}$.
- ∞ and -1 : there is $\gamma_2 \in \Gamma(2)$ with $\gamma_2\infty = \frac{a_2}{c_2} = -1$, so $a_2 = -c_2$, contradicting a_2 odd and c_2 even.
- 0 and -1 : there is $\gamma_3 \in \Gamma(2)$ with $\gamma_3(0) = \frac{b_3}{d_3} = -1$, so $b_3 = -d_3$, contradicting b_3 even and d_3 odd.

Therefore $\infty, 0, -1$ are pairwise $\Gamma(2)$ -inequivalent. Together with Problem 3.4.9 (every cusp is equivalent to some $\alpha_j^{-1}\infty$, and for the representatives of Example 3.3.4 these are $\infty, \infty, 0, -1, 0, -1$), $\Gamma(2)$ has exactly 3 cusps.

(b) We use Problem 3.4.9: the cusps of Γ' are among the $\alpha_j^{-1}\infty$.

$\Gamma_0(2)$. Since $[\Gamma : \Gamma_0(2)] = 3$, we take the left coset representatives of Problem 3.4.7 ($p = 2, e = 1$):

$$\alpha_1 = I, \quad \alpha_2 = S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \alpha_3 = T^{-1}S = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then $\alpha_1^{-1}\infty = \infty$, $\alpha_2^{-1}\infty = S^{-1}\infty = 0$, and $\alpha_3^{-1}\infty = S^{-1}T\infty = S^{-1}\infty = 0$. So every cusp is equivalent to ∞ or 0 .

- ∞ and 0 are inequivalent: if $\gamma\infty = \frac{a}{c} = 0$, then $a = 0$, so $ad - bc = -bc = 1$ and $c = \pm 1$, contradicting $c \equiv 0 \pmod{2}$.
- For instance, 0 and -1 are equivalent: $\gamma = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = T^{-1} \in \Gamma_0(2)$ and $\gamma(0) = \frac{1 \cdot 0 - 1}{0 \cdot 0 + 1} = -1$.

So $\Gamma_0(2)$ has 2 cusps: ∞ and $0 \sim -1$.

$\mathfrak{S}(2)$ and $\Gamma^0(2)$. They are conjugate to $\Gamma_0(2)$ in Γ (Problem 3.4.6(b)), and conjugation $\Gamma' \mapsto g\Gamma'g^{-1}$ maps the Γ' -classes of cusps bijectively to the $g\Gamma'g^{-1}$ -classes via $s \mapsto gs$. So both also have 2 cusps. Directly: for $\mathfrak{S}(2)$, the cusps ∞ and -1 are inequivalent, since $\gamma\infty = \frac{a}{c} = -1$ would need $a \equiv c \pmod{2}$, while for $\gamma \equiv I$ or S exactly one of a, c is odd; for $\Gamma^0(2)$, the cusps ∞ and 0 are inequivalent, since $\gamma\infty = 0$ forces $a = 0$, $bc = -1$, contradicting b even.

The $\langle ST \rangle$ -group. Since its index is 2, choose $\alpha_1 = I$, $\alpha_2 = S$; then $\alpha_1^{-1}\infty = \infty$ and $\alpha_2^{-1}\infty = 0$. By Problem 3.4.9, ∞ and 0 are equivalent iff $S^{-1}T^n \in \Gamma'$ for some n . Since $S^{-1} = -S$,

$$S^{-1}T^n = -ST^n = -\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = -\begin{pmatrix} 0 & -1 \\ 1 & n \end{pmatrix} \equiv \begin{pmatrix} 0 & 1 \\ 1 & n \end{pmatrix} \pmod{2}.$$

For $n = 1$ this is $\equiv ST \pmod{2}$, which lies in $\langle ST \rangle \pmod{2}$. So $\infty \sim 0$, and this group has only 1 cusp.

Summary: Γ : 1 cusp; $\langle ST \rangle$ -group: 1; $\Gamma_0(2)$, $\mathfrak{S}(2)$, $\Gamma^0(2)$: 2 each; $\Gamma(2)$: 3. ■

Correction. In my handwritten solution of (a), the case “0 and -1 ” concluded from $\gamma_3(0) = -1$ that $b_3 = 0$, contradicting “ $b \equiv 1 \pmod{2}$ ”. The correct deduction is $b_3 = -d_3$, which contradicts b_3 even and d_3 odd. In (b), my notes used I, S, ST as coset representatives for $\Gamma_0(2)$ (chosen so that their bottom rows cover the possibilities modulo 2); but S and ST lie in the same left coset, as $S^{-1}(ST) = T \in \Gamma_0(2)$. The bottom row determines the *right* coset $\Gamma_0(2)\gamma$, while Problem 3.4.9 uses left cosets $\alpha_j\Gamma'$; the representatives above fix this, and the conclusion (2 cusps) is unchanged. Also “ $bc = 1$ ” should be $-bc = 1$, and in Problem 3.4.9 below the product ST^n was computed as $\begin{pmatrix} 0 & -1 \\ -1 & -n \end{pmatrix}$; the correct product is $\begin{pmatrix} 0 & -1 \\ 1 & n \end{pmatrix}$ (the conclusion modulo 2 is unchanged).

Report Problem 3.4.9 (Problem 17)

Let $\{\alpha_j\}$ be a complete set of coset representatives for $\bar{\Gamma}'$ in $\bar{\Gamma}$, where Γ' is a subgroup of finite index in Γ . Show that the cusps of Γ' are among the set $\{\alpha_j^{-1}\infty\}$, but that $\alpha_j^{-1}\infty$ and $\alpha_k^{-1}\infty$ are Γ' -equivalent if and only if there exists $n \in \mathbb{Z}$ such that $\alpha_k^{-1}T^n\alpha_j \in \bar{\Gamma}'$.

Solution. The cusps of Γ' are among $\{\alpha_j^{-1}\infty\}$. The cusps of the full modular group are $\mathbb{Q} \cup \{\infty\}$, and they are all equivalent to ∞ (Proposition 3.1.5). So for every $s \in \mathbb{Q} \cup \{\infty\}$ there is $\gamma \in \bar{\Gamma}$ such that $s = \gamma(\infty)$. Since

$$\bar{\Gamma} = \bigcup_j \alpha_j \bar{\Gamma}' = \bigcup_j \bar{\Gamma}' \alpha_j^{-1}$$

(taking inverses of all elements turns left cosets into right cosets: $(\alpha_j\Gamma')^{-1} = \Gamma'\alpha_j^{-1}$), we have $\gamma = \gamma'\alpha_j^{-1}$ for some j and some $\gamma' \in \bar{\Gamma}'$. Thus

$$s = \gamma(\infty) = \gamma'\alpha_j^{-1}(\infty) = \gamma'(\alpha_j^{-1}\infty),$$

so s and $\alpha_j^{-1}\infty$ are Γ' -equivalent. Hence every cusp is equivalent to one of the $\alpha_j^{-1}\infty$.

The criterion. ($\Rightarrow$) Suppose $\alpha_j^{-1}\infty$ and $\alpha_k^{-1}\infty$ are Γ' -equivalent. Then there is $\gamma' \in \bar{\Gamma}'$ with $\gamma'(\alpha_j^{-1}\infty) = \alpha_k^{-1}\infty$, hence

$$\alpha_k \gamma' \alpha_j^{-1}(\infty) = \infty.$$

So $M = \alpha_k \gamma' \alpha_j^{-1}$ lies in the stabiliser of ∞ in $\bar{\Gamma}$. If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $M\infty = \frac{a}{c} = \infty$, then $c = 0$, so $ad = 1$, $a = d = \pm 1$, and $M = \pm \begin{pmatrix} 1 & \pm b \\ 0 & 1 \end{pmatrix}$, i.e. $M = T^n$ in $\bar{\Gamma}$ for some $n \in \mathbb{Z}$. Therefore

$$T^n = \alpha_k \gamma' \alpha_j^{-1} \implies \alpha_k^{-1} T^n \alpha_j = \gamma' \in \bar{\Gamma}'.$$

($\Leftarrow$) Let $\gamma' = \alpha_k^{-1} T^n \alpha_j \in \bar{\Gamma}'$. Then

$$\gamma'(\alpha_j^{-1}\infty) = \alpha_k^{-1} T^n \alpha_j \alpha_j^{-1}(\infty) = \alpha_k^{-1} T^n(\infty) = \alpha_k^{-1}(\infty),$$

since $T^n\infty = \infty$. So $\alpha_j^{-1}\infty$ and $\alpha_k^{-1}\infty$ are Γ' -equivalent. ■

Chapter 4

Modular Forms of Level 1 and the Valence Formula

In this chapter we study modular forms for the full modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. We compute the q -expansions of the Eisenstein series, prove the valence formula, and deduce from it the structure of the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$: they are spanned by monomials $E_4^i E_6^j$, and $S_k(\Gamma) = \Delta M_{k-12}(\Gamma)$. Then we study the discriminant Δ and the j -invariant, the (non-modular) Eisenstein series G_2 via Hecke's trick, and the Dedekind η -function, which gives the product expansion $\Delta = (2\pi)^{12} q \prod_{n \geq 1} (1 - q^n)^{24}$. We follow [2, §§III.2–III.3] and [4, Ch. VII].

Notes. In my handwritten Note 4 the pages are not in logical order; I numbered the parts by hand (“1st”, “2nd”, ...). This chapter follows that numbering: first the normalised Eisenstein series, then the valence formula, its corollaries, the dimension formula, and the generation of $M_k(\Gamma)$ by E_4 and E_6 ; then the remaining topics in the order of the notes.

Throughout, $q = e^{2\pi iz}$ for $z \in \mathfrak{H}$; note that $|q| = e^{-2\pi \mathrm{Im} z} < 1$, and $q \rightarrow 0$ as $\mathrm{Im} z \rightarrow \infty$. We write $\rho = e^{\pi i/3} = \frac{1}{2} + \frac{\sqrt{3}}{2}i$, so that $\rho^2 = e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i = \rho - 1$, and F denotes the fundamental domain of Theorem 3.3.2, whose corners are ρ^2 and ρ .

4.1 Modular functions and modular forms

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $z \in \mathfrak{H}$ we write $j(\gamma, z) = cz + d$; it is nonzero by Proposition 3.1.4.

Lemma 4.1.1 (Cocycle relation)

For $\gamma, \gamma' \in \Gamma$ and $z \in \mathfrak{H}$: $j(\gamma\gamma', z) = j(\gamma, \gamma'z)j(\gamma', z)$.

Proof. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\gamma' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$, so the bottom row of $\gamma\gamma'$ is $(ca' + dc', cb' + dd')$. Then

$$\begin{aligned} j(\gamma, \gamma'z)j(\gamma', z) &= \left(c \frac{a'z + b'}{c'z + d'} + d\right)(c'z + d') = c(a'z + b') + d(c'z + d') \\ &= (ca' + dc')z + (cb' + dd') = j(\gamma\gamma', z). \end{aligned}$$

■

Definition 4.1.2 (Weakly modular function)

Let $k \in \mathbb{Z}$. A meromorphic function f on $\mathfrak{H}$ is *weakly modular of weight k* (for Γ) if

$$f(\gamma z) = (cz + d)^k f(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, z \in \mathfrak{H}.$$

Lemma 4.1.3

A meromorphic function f on $\mathfrak{H}$ is weakly modular of weight k if and only if

$$f(z+1) = f(z) \quad \text{and} \quad f\left(-\frac{1}{z}\right) = z^k f(z).$$

Proof. ($\Rightarrow$) Take $\gamma = T$ (so $c = 0, d = 1$) and $\gamma = S$ (so $c = 1, d = 0$).

($\Leftarrow$) Let G be the set of $\gamma \in \Gamma$ with $f(\gamma z) = j(\gamma, z)^k f(z)$ for all z . By assumption $S, T \in G$. If $\gamma, \gamma' \in G$, then by Lemma 4.1.1

$$f(\gamma\gamma'z) = j(\gamma, \gamma'z)^k f(\gamma'z) = j(\gamma, \gamma'z)^k j(\gamma', z)^k f(z) = j(\gamma\gamma', z)^k f(z),$$

so $\gamma\gamma' \in G$. If $\gamma \in G$, then applying the definition at $\gamma^{-1}z$,

$$f(z) = f(\gamma\gamma^{-1}z) = j(\gamma, \gamma^{-1}z)^k f(\gamma^{-1}z), \quad 1 = j(I, z) = j(\gamma, \gamma^{-1}z)j(\gamma^{-1}, z)$$

by Lemma 4.1.1; hence $f(\gamma^{-1}z) = j(\gamma^{-1}, z)^k f(z)$, and $\gamma^{-1} \in G$. So G is a subgroup containing S and T . Since $\bar{\Gamma}$ is generated by S, T (Theorem 3.3.2(2)) and $-I = S^2$, we have $\Gamma = \langle S, T \rangle$, hence $G = \Gamma$. ■

Let f be weakly modular of weight k . Taking $\gamma = -I$ gives $f(z) = (-1)^k f(z)$, so $f = 0$ if k is odd; hence only even weights are interesting. Taking $\gamma = T$ gives $f(z+1) = f(z)$. Therefore f is a function of $q = e^{2\pi iz}$: there is a meromorphic function $\tilde{f}$ on the punctured disc $\{0 < |q| < 1\}$ with $f(z) = \tilde{f}(q)$.

Definition 4.1.4 (Modular function, modular form, cusp form)

Let f be weakly modular of weight k .

- (1) f is a *modular function of weight k* if $\tilde{f}$ is meromorphic at $q = 0$, i.e. $f(z) = \tilde{f}(q) = \sum_{n \geq n_0} a_n q^n$ for some $n_0 \in \mathbb{Z}$ (for $|q|$ small). This is the q -*expansion* (Fourier expansion) of f ; we say f is *meromorphic at ∞* .
- (2) f is a *modular form of weight k* if it is holomorphic on $\mathfrak{H}$ and at ∞ , i.e. $f(z) = \sum_{n \geq 0} a_n q^n$. We then write $f(\infty) = a_0$. The space of such f is $M_k(\Gamma)$.
- (3) f is a *cusp form of weight k* if it is a modular form with $a_0 = 0$. The space of cusp forms is $S_k(\Gamma) \subseteq M_k(\Gamma)$.

Definition 4.1.5 (Order)

Let $f \neq 0$ be a modular function of weight k . For $p \in \mathfrak{H}$, $v_p(f) \in \mathbb{Z}$ is the order of the zero of f at p minus the order of the pole of f at p ; equivalently,

$$v_p(f) = n \iff (z - p)^{-n} f(z) \text{ is holomorphic and non-vanishing at } p.$$

At ∞ we define $v_\infty(f) = v_0(\tilde{f})$, the order of $\tilde{f}$ at $q = 0$.

Proposition 4.1.6

If $f(z) = \sum_{n \geq n_0} a_n q^n$ with $a_{n_0} \neq 0$, then $v_\infty(f) = n_0$.

Proof. Let us denote $\tilde{f}(q) = f(z) = \sum_{n \geq n_0} a_n q^n$. As $z \rightarrow \infty$, i.e. $\text{Im}(z) \rightarrow \infty$, we have $|q| = |e^{2\pi iz}| = |e^{-2\pi y}| = e^{-2\pi y} \rightarrow 0$ ($z = x + iy$). So the study of f at ∞ is the same as the study of $\tilde{f}$ at 0, and we want to check $v_0(\tilde{f}) = n_0$. Note that

$$\frac{\tilde{f}(q)}{(q-0)^{n_0}} = \frac{1}{q^{n_0}} \sum_{n=n_0}^{\infty} a_n q^n = \sum_{n=n_0}^{\infty} a_n q^{n-n_0} = a_{n_0} + \sum_{n=n_0+1}^{\infty} a_n q^{n-n_0} = a_{n_0} + \sum_{m \geq 1} a_{m+n_0} q^m$$

(rewriting $m = n - n_0$). The power series $\sum_{m \geq 1} a_{m+n_0} q^m$ is holomorphic near 0 and vanishes at $q = 0$. Since $a_{n_0} \neq 0$, $\tilde{f}(q)/(q-0)^{n_0}$ is holomorphic and non-vanishing at $q = 0$. Thus $v_\infty(f) = v_0(\tilde{f}) = n_0$. ■

Correction. In my handwritten notes the series $\sum_{m \geq 0} a_{m+n_0} q^m$ was called a “polynomial”; it is a power series, convergent for $|q| < 1$ (in particular holomorphic near 0).

Lemma 4.1.7

Let $f \not\equiv 0$ be a modular function of weight k . Then $v_{\gamma p}(f) = v_p(f)$ for all $\gamma \in \Gamma$ and $p \in \mathfrak{H}$. Moreover, there is $M > 0$ such that f has no zeros and no poles in $\{\text{Im } z > M\}$.

Proof. From $f(\gamma z) = (cz + d)^k f(z)$, with $(cz + d)^k$ holomorphic and non-vanishing near p , the function $f \circ \gamma$ has the same order at p as f . Since γ is biholomorphic near p with $\gamma'(p) = (cp + d)^{-2} \neq 0$, the order of $f \circ \gamma$ at p equals the order of f at γp .

For the second claim: $\tilde{f}(q) = q^{n_0}(a_{n_0} + a_{n_0+1}q + \cdots)$ with $a_{n_0} \neq 0$; the bracket is holomorphic and nonzero for $|q| < r$, some $r > 0$. So $\tilde{f}$ has no zeros or poles in $0 < |q| < r$, i.e. f has none in $\text{Im } z > M := -\frac{\log r}{2\pi}$. ■

Hence the expression $v_p(f)$ only depends on the Γ -orbit of p , and the sum over “ $p \in \Gamma \backslash \mathfrak{H}$ ” in the valence formula below makes sense; by Lemma 4.1.7 and the compactness of $F \cap \{\text{Im } z \leq M\}$, only finitely many terms are nonzero.

4.2 Eisenstein series and their q -expansions

Definition 4.2.1 (Eisenstein series)

For an even integer $k \geq 4$ and $z \in \mathfrak{H}$, let $L_z = \mathbb{Z}z + \mathbb{Z}$ and

$$G_k(z) = G_k(L_z) = \sum_{\substack{(m,n) \in \mathbb{Z}^2 \\ (m,n) \neq (0,0)}} \frac{1}{(mz + n)^k}.$$

Lemma 4.2.2

Let $A, \delta > 0$ and $K = \{z \in \mathfrak{H} : |\text{Re } z| \leq A, \text{Im } z \geq \delta\}$. There is $c > 0$ such that $|mz + n|^2 \geq c(m^2 + n^2)$ for all $z \in K$ and all real m, n .

Proof. For $z = x + iy \in K$: $|mz + n|^2 = (mx + n)^2 + m^2y^2 \geq (mx + n)^2 + m^2\delta^2 =: h(x, m, n)$. On the compact set $\{|x| \leq A\} \times \{m^2 + n^2 = 1\}$ the continuous function h is positive (if $m = 0$ then $n = \pm 1$ and $h = 1$; if $m \neq 0$ then $h \geq m^2\delta^2 > 0$), so it has a minimum $c > 0$. For general $(m, n) \neq (0, 0)$ apply this to $(m, n)/\sqrt{m^2 + n^2}$ and multiply by $m^2 + n^2$ (both sides are homogeneous of degree 2). ■

Proposition 4.2.3

For even $k \geq 4$, the series $G_k(z)$ converges absolutely and uniformly on every set K as in Lemma 4.2.2; G_k is holomorphic on $\mathfrak{H}$, and it is weakly modular of weight k .

Proof. By Lemma 4.2.2, $|mz + n|^{-k} \leq c^{-k/2}(m^2 + n^2)^{-k/2} = c^{-k/2}|mi + n|^{-k}$ on K , and $\sum_{(m,n) \neq 0} |mi + n|^{-k} < \infty$ by Lemma 2.3.3 (for the lattice $\mathbb{Z}[i]$, $k > 2$). By the M -test the series converges absolutely and uniformly on K ; since every compact subset of $\mathfrak{H}$ lies in such a K , G_k is holomorphic by Lemma 1.1.3.

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

$$m(\gamma z) + n = \frac{m(az + b) + n(cz + d)}{cz + d} = \frac{(ma + nc)z + (mb + nd)}{cz + d} = \frac{m'z + n'}{cz + d},$$

with $(m', n') = (m, n) \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Since γ is invertible over $\mathbb{Z}$, $(m, n) \mapsto (m', n')$ is a bijection of $\mathbb{Z}^2 \setminus \{(0, 0)\}$, and as the series converges absolutely we may reorder it:

$$G_k(\gamma z) = \sum_{(m, n) \neq 0} \frac{(cz + d)^k}{(m'z + n')^k} = (cz + d)^k \sum_{(m', n') \neq 0} \frac{1}{(m'z + n')^k} = (cz + d)^k G_k(z). \quad \blacksquare$$

To compute the q -expansion of G_k we need two lemmas.

Lemma 4.2.4 (Partial fractions of the cotangent)

For $z \in \mathbb{C} \setminus \mathbb{Z}$,

$$\pi \cot \pi z = \frac{1}{z} + \sum_{n \geq 1} \left(\frac{1}{z + n} + \frac{1}{z - n} \right) = \frac{1}{z} + \sum_{n \geq 1} \frac{2z}{z^2 - n^2},$$

the series converging absolutely and uniformly on compact subsets of $\mathbb{C} \setminus \mathbb{Z}$.

Proof. Let $g(z)$ be the right-hand side. On a compact $K \subseteq \mathbb{C} \setminus \mathbb{Z}$ with $|z| \leq R$, for $n \geq 2R$ we have $|z^2 - n^2| \geq n^2 - R^2 \geq \frac{3}{4}n^2$, so $\left| \frac{2z}{z^2 - n^2} \right| \leq \frac{8R}{3n^2}$; by the M -test g converges absolutely and uniformly on K , and g is meromorphic on $\mathbb{C}$ with simple poles of residue 1 at the integers (each term $\frac{1}{z \pm n}$ has residue 1 at $\mp n$).

g is odd and 1-periodic. Oddness is clear from the second form. For periodicity, write the symmetric partial sums $g_N(z) = \sum_{n=-N}^N \frac{1}{z+n}$, so $g = \lim_N g_N$. Then $g_N(z+1) - g_N(z) = \frac{1}{z+N+1} - \frac{1}{z-N} \rightarrow 0$ as $N \rightarrow \infty$, hence $g(z+1) = g(z)$.

The function $\pi \cot \pi z = \pi \frac{\cos \pi z}{\sin \pi z}$ is also odd, 1-periodic, with simple poles of residue $\frac{\pi \cos \pi n}{\pi \cos \pi n} = 1$ at each $n \in \mathbb{Z}$. Hence $h(z) = \pi \cot \pi z - g(z)$ is entire (the poles cancel), odd and 1-periodic.

h is bounded. By periodicity it suffices to consider $0 \leq x \leq 1$, $z = x + iy$. On the compact set $\{0 \leq x \leq 1, |y| \leq 1\}$, h is continuous, hence bounded. For $|y| \geq 1$: $\pi \cot \pi z = \pi i \frac{e^{2\pi iz} + 1}{e^{2\pi iz} - 1}$ is bounded (it tends to $\mp \pi i$ as $y \rightarrow \pm \infty$). For g , use $|z \mp n| \geq \max(|y|, n-1)$ and $|z \pm n| \geq \max(|y|, n)$ for $n \geq 1$:

$$\sum_{n \geq 1} \frac{2|z|}{|z-n||z+n|} \leq \sum_{1 \leq n \leq |y|+1} \frac{2(|y|+1)}{y^2} + \sum_{n > |y|+1} \frac{2(|y|+1)}{(n-1)n} \leq \frac{2(|y|+2)(|y|+1)}{y^2} + \frac{2(|y|+1)}{|y|},$$

which is bounded for $|y| \geq 1$. So h is a bounded entire function, hence constant by Liouville's theorem; being odd, $h(0) = -h(0)$, so $h \equiv 0$. $\blacksquare$

Notes. In my handwritten notes this formula was used without proof (“note that $\sum_{n \in \mathbb{Z}} \frac{1}{z+n} = \pi \cot \pi z$ ”). The series $\sum_{n \in \mathbb{Z}} \frac{1}{z+n}$ does not converge absolutely; it is to be understood as the symmetric limit $\lim_N \sum_{n=-N}^N$, which is what the two forms in Lemma 4.2.4 make precise.

Lemma 4.2.5 (Lipschitz formula)

For $z \in \mathfrak{H}$ and $q = e^{2\pi iz}$:

- (i) $\sum_{n \in \mathbb{Z}} \frac{1}{z+n} = \pi \cot \pi z = \pi i - 2\pi i \sum_{n \geq 0} q^n$;
- (ii) for every integer $k \geq 2$, $\sum_{n \in \mathbb{Z}} \frac{1}{(z+n)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{n \geq 1} n^{k-1} q^n$, and for even k the factor is $\frac{(2\pi i)^k}{(k-1)!}$.

Proof. (i) Using $\cos w = \frac{e^{iw} + e^{-iw}}{2}$ and $\sin w = \frac{e^{iw} - e^{-iw}}{2i}$,

$$\pi \cot \pi z = \pi \frac{\frac{e^{i\pi z} + e^{-i\pi z}}{2}}{\frac{e^{i\pi z} - e^{-i\pi z}}{2i}} = i\pi \frac{e^{i\pi z} + e^{-i\pi z}}{e^{i\pi z} - e^{-i\pi z}} \cdot \frac{e^{i\pi z}}{e^{i\pi z}} = i\pi \frac{e^{2\pi iz} + 1}{e^{2\pi iz} - 1} = i\pi \frac{q+1}{q-1} = i\pi \left(1 + \frac{2}{q-1}\right),$$

and since $|q| < 1$, $\frac{2}{q-1} = -\frac{2}{1-q} = -2 \sum_{n \geq 0} q^n$. So $\pi \cot \pi z = i\pi - 2\pi i \sum_{n \geq 0} q^n$.

(ii) Differentiate (i) $k-1$ times with respect to z . On the left, term by term (allowed by the locally uniform convergence in Lemma 4.2.4), $\frac{d^{k-1}}{dz^{k-1}} \frac{1}{z+n} = \frac{(-1)^{k-1} (k-1)!}{(z+n)^k}$. On the right, $\frac{d}{dz} q^n = 2\pi i n q^n$, so the constant $i\pi$ and the term $n=0$ disappear and

$$(-1)^{k-1} (k-1)! \sum_{n \in \mathbb{Z}} \frac{1}{(z+n)^k} = -2\pi i \sum_{n \geq 1} (2\pi i n)^{k-1} q^n = -(2\pi i)^k \sum_{n \geq 1} n^{k-1} q^n.$$

Dividing by $(-1)^{k-1} (k-1)!$ gives $\sum_n \frac{1}{(z+n)^k} = \frac{(-1)^k (2\pi i)^k}{(k-1)!} \sum_{n \geq 1} n^{k-1} q^n$. (For even k , $(-1)^{k-1} = -1$, which is the computation $\frac{-(k-1)!}{(z+n)^k}$ in my notes.) ■

Definition 4.2.6 (Bernoulli numbers)

The Bernoulli numbers B_n are defined by

$$\frac{z}{e^z - 1} = \sum_{n \geq 0} B_n \frac{z^n}{n!} \quad (|z| < 2\pi).$$

So $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}$, $B_8 = -\frac{1}{30}$, $B_{10} = \frac{5}{66}$, $B_{12} = -\frac{691}{2730}$, and $B_n = 0$ for odd $n \geq 3$: indeed $\frac{z}{e^z - 1} + \frac{z}{2} = \frac{z}{2} \cdot \frac{e^z + 1}{e^z - 1} = \frac{z}{2} \coth \frac{z}{2}$ is an even function, and $\frac{z}{2}$ cancels exactly the term $B_1 \frac{z^1}{1!} = -\frac{z}{2}$.

Lemma 4.2.7

For every even $k \geq 2$, $\zeta(k) = -\frac{B_k}{2k} \cdot \frac{(2\pi i)^k}{(k-1)!} = -\frac{B_k (2\pi i)^k}{2 \cdot k!}$. In particular $\zeta(2) = \frac{\pi^2}{6}$, $\zeta(4) = \frac{\pi^4}{90}$, $\zeta(6) = \frac{\pi^6}{945}$.

Proof. Multiplying Lemma 4.2.4 by z :

$$\pi z \cot \pi z = 1 + \sum_{n \geq 1} \left(\frac{z}{z+n} + \frac{z}{z-n} \right) = 1 + \sum_{n \geq 1} \frac{2z^2}{z^2 - n^2} = 1 - 2 \sum_{n \geq 1} \frac{z^2/n^2}{1 - z^2/n^2}.$$

For $|z| < 1$ we have $|z^2/n^2| < 1$, so expanding the geometric series,

$$\pi z \cot \pi z = 1 - 2 \sum_{n \geq 1} \sum_{m \geq 1} \frac{z^{2m}}{n^{2m}} = 1 - 2 \sum_{m \geq 1} \left(\sum_{n \geq 1} \frac{1}{n^{2m}} \right) z^{2m} = 1 - 2 \sum_{m \geq 1} \zeta(2m) z^{2m},$$

where the interchange is allowed since $\sum_{n,m} |z|^{2m} / n^{2m} \leq \sum_n \frac{|z|^2/n^2}{1-|z|^2} < \infty$.

On the other hand, from the definition of B_n and the vanishing of the odd ones, $\frac{z}{e^z-1} + \frac{z}{2} = \sum_{n \geq 0} B_{2n} \frac{z^{2n}}{(2n)!}$ for $|z| < 2\pi$. Replace z by $2\pi iz$ (allowed for $|z| < 1$):

$$\frac{2\pi iz}{e^{2\pi iz} - 1} + \pi iz = \frac{2 \cdot 2\pi iz + 2\pi iz(e^{2\pi iz} - 1)}{2(e^{2\pi iz} - 1)} = \pi iz \frac{e^{2\pi iz} + 1}{e^{2\pi iz} - 1} = \pi z \cot \pi z,$$

using the formula for $\cot$ from Lemma 4.2.5(i). Hence $\pi z \cot \pi z = \sum_{n \geq 0} \frac{(2\pi i)^{2n}}{(2n)!} B_{2n} z^{2n}$. Comparing coefficients of z^{2m} ($m \geq 1$) of the two power series:

$$-2\zeta(2m) = \frac{(2\pi i)^{2m}}{(2m)!} B_{2m}.$$

With $k = 2m$ this is $\zeta(k) = -\frac{B_k(2\pi i)^k}{2 \cdot k!}$. For example $\zeta(4) = -\frac{B_4}{8} \cdot \frac{(2\pi i)^4}{3!} = \frac{1}{240} \cdot \frac{16\pi^4}{6} = \frac{\pi^4}{90}$ and $\zeta(6) = -\frac{B_6}{12} \cdot \frac{(2\pi i)^6}{5!} = -\frac{1}{504} \cdot \frac{-64\pi^6}{120} = \frac{\pi^6}{945}$. ■

Theorem 4.2.8 (q -expansion of G_k)

For even $k \geq 4$,

$$G_k(z) = 2\zeta(k) E_k(z), \quad E_k(z) = 1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n, \quad \sigma_{k-1}(n) = \sum_{d|n} d^{k-1}.$$

E_k is called the *normalised Eisenstein series*; it is a modular form of weight k with $E_k(\infty) = 1$.

Proof. Split the sum defining G_k according to $m = 0$ and $m \neq 0$:

$$\begin{aligned} G_k(z) &= \sum_{n \neq 0} \frac{1}{n^k} + \sum_{m \neq 0} \left(\sum_{n \in \mathbb{Z}} \frac{1}{(mz + n)^k} \right) \\ &= \left(\sum_{n=-\infty}^{-1} \frac{1}{n^k} + \sum_{n=1}^{\infty} \frac{1}{n^k} \right) + \left(\sum_{m=-\infty}^{-1} \sum_{n \in \mathbb{Z}} \frac{1}{(mz + n)^k} + \sum_{m=1}^{\infty} \sum_{n \in \mathbb{Z}} \frac{1}{(mz + n)^k} \right) \\ &= 2\zeta(k) + 2 \sum_{m \geq 1} \sum_{n \in \mathbb{Z}} \frac{1}{(mz + n)^k}, \end{aligned}$$

using $(-n)^k = n^k$ and $(-mz - n)^k = (mz + n)^k$ for even k . By Lemma 4.2.5(ii) applied to $mz \in \mathfrak{H}$ (with $e^{2\pi imz} = q^m$),

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

Hence

$$\begin{aligned} G_k(z) &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \sum_{m \geq 1} \sum_{r \geq 1} r^{k-1} q^{mr} \\ &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \left((q + 2^{k-1}q^2 + 3^{k-1}q^3 + \dots) + (q^2 + 2^{k-1}q^4 + \dots) + (q^3 + \dots) + \dots \right) \\ &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \left(q + (1 + 2^{k-1})q^2 + (1 + 3^{k-1})q^3 + \dots \right) \\ &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \sum_{N \geq 1} \left(\sum_{r|N} r^{k-1} \right) q^N, \end{aligned}$$

where we collected the terms with $mr = N$ (allowed by absolute convergence, as $\sum_{m,r} r^{k-1} |q|^{mr} < \infty$). By Lemma 4.2.7, $2\zeta(k) = -\frac{2B_k}{2k} \cdot \frac{(2\pi i)^k}{(k-1)!}$, so

$$G_k(z) = \frac{2(2\pi i)^k}{(k-1)!} \left(-\frac{B_k}{2k} + \sum_{N \geq 1} \sigma_{k-1}(N) q^N \right).$$

Multiplying both sides by $\frac{1}{2\zeta(k)} = -\frac{k(k-1)!}{B_k(2\pi i)^k}$ gives $\frac{G_k(z)}{2\zeta(k)} = 1 - \frac{2k}{B_k} \sum_N \sigma_{k-1}(N) q^N = E_k(z)$. Finally E_k is weakly modular of weight k (Proposition 4.2.3), holomorphic on $\mathfrak{H}$, and its q -expansion has no negative powers, so $E_k \in M_k(\Gamma)$. ■

Correction. In my handwritten notes the claim “ $E_k \in M_k(\Gamma)$ ” was marked “Proof: Done” without details; the details are Proposition 4.2.3 and Theorem 4.2.8. The normalised series was first written as $E_{2k} = 1 - \frac{4k}{B_{2k}} \sum \sigma_{2k-1}(n) q^n$ (indexing by $2k$), which is the same formula as above.

Example 4.2.9. With $\frac{2k}{B_k} = -240, 504, -480, 264$ for $k = 4, 6, 8, 10$:

$$\begin{aligned} E_4 &= 1 + 240 \sum \sigma_3(n) q^n = 1 + 240(q + (1^3 + 2^3)q^2 + (1^3 + 3^3)q^3 + \cdots) \\ &= 1 + 240q + 2160q^2 + 6720q^3 + \cdots, \\ E_6 &= 1 - 504 \sum \sigma_5(n) q^n = 1 - 504q - 16632q^2 - 122976q^3 - \cdots, \\ E_8 &= 1 + 480 \sum \sigma_7(n) q^n = 1 + 480q + 61920q^2 + 1050240q^3 + \cdots, \\ E_{10} &= 1 - 264 \sum \sigma_9(n) q^n = 1 - 264q - 135432q^2 - 5196576q^3 - \cdots. \end{aligned}$$

The discriminant and the functions g_2, g_3

Recall from Chapter 2 that for a lattice L one defines $g_2(L) = 60G_4(L)$ and $g_3(L) = 140G_6(L)$, and the elliptic curve $y^2 = 4x^3 - g_2(L)x - g_3(L)$ has discriminant (up to a factor 16) $g_2^3 - 27g_3^2$, which is nonzero since the roots e_1, e_2, e_3 are distinct (Theorem 2.4.5; Koblitz, Problem 2 of §I.6). For $z \in \mathfrak{H}$ put

$$g_2(z) = g_2(L_z) = 60G_4(z), \quad g_3(z) = g_3(L_z) = 140G_6(z), \quad \Delta(z) = g_2(z)^3 - 27g_3(z)^2.$$

So $g_2 \in M_4(\Gamma)$, $g_3 \in M_6(\Gamma)$, and $\Delta(z) \neq 0$ for all $z \in \mathfrak{H}$.

Proposition 4.2.10

$g_2 = \frac{4}{3}\pi^4 E_4$, $g_3 = \frac{8}{27}\pi^6 E_6$, and

$$\Delta = \frac{(2\pi)^{12}}{1728} (E_4^3 - E_6^2) = (2\pi)^{12} (q - 24q^2 + \cdots).$$

In particular $\Delta \in S_{12}(\Gamma)$, $\Delta \neq 0$, and $v_\infty(\Delta) = 1$.

Proof. By Theorem 4.2.8 and Lemma 4.2.7,

$$g_2 = 60 \cdot 2\zeta(4) E_4 = 120 \cdot \frac{\pi^4}{90} E_4 = \frac{4}{3}\pi^4 E_4, \quad g_3 = 140 \cdot 2\zeta(6) E_6 = 280 \cdot \frac{\pi^6}{945} E_6 = \frac{8}{27}\pi^6 E_6.$$

Hence

$$\Delta = \left(\frac{4}{3}\pi^4 E_4 \right)^3 - 27 \left(\frac{8}{27}\pi^6 E_6 \right)^2 = \frac{64}{27}\pi^{12} E_4^3 - 27 \cdot \frac{64}{27^2}\pi^{12} E_6^2 = \frac{64}{27}\pi^{12} (E_4^3 - E_6^2) = \frac{(2\pi)^{12}}{1728} (E_4^3 - E_6^2),$$

since $\frac{2^{12}}{1728} = \frac{4096}{1728} = \frac{64}{27}$. It is weakly modular of weight 12 (as E_4^3 and E_6^2 are), holomorphic on $\mathfrak{H}$ and at ∞ . From Example 4.2.9,

$$\begin{aligned} E_4^3 &= 1 + 3 \cdot 240q + (3 \cdot 2160 + 3 \cdot 240^2)q^2 + \cdots = 1 + 720q + 179280q^2 + \cdots, \\ E_6^2 &= 1 - 2 \cdot 504q + (-2 \cdot 16632 + 504^2)q^2 + \cdots = 1 - 1008q + 220752q^2 + \cdots, \end{aligned}$$

so $E_4^3 - E_6^2 = 1728q - 41472q^2 + \cdots = 1728(q - 24q^2 + \cdots)$. The constant term vanishes, so $\Delta \in S_{12}(\Gamma)$, and the coefficient of q is nonzero, so $v_\infty(\Delta) = 1$ (Proposition 4.1.6). ■

Notes. In the corollaries of the valence formula my handwritten notes use the normalisation $\frac{1}{1728}(E_4^3 - E_6^2) = q - 24q^2 + 252q^3 - 1472q^4 + \cdots$ (written there with “ $\frac{1}{1748}$ ”, a slip for $\frac{1}{1728}$), whereas Koblitz’s $\Delta = g_2^3 - 27g_3^2$ carries the extra factor $(2\pi)^{12}$. Since the two differ by a nonzero constant, all statements like $S_{12}(\Gamma) = \mathbb{C}\Delta$ are unaffected; we use Koblitz’s Δ throughout. The further coefficients 252, -1472 will follow from the product formula (Theorem 4.8.4).

4.3 The valence formula

Recall the fundamental domain of $\mathrm{SL}_2(\mathbb{Z})$ acting on $\mathfrak{H}$:

$$F = \left\{ z \in \mathfrak{H} \mid -\frac{1}{2} \leq \mathrm{Re}(z) \leq \frac{1}{2}, |z| \geq 1 \right\},$$

with corners $\rho^2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $\rho = \frac{1}{2} + \frac{\sqrt{3}}{2}i$, and with the point i in the middle of its lower boundary arc. A modular function f is meromorphic on $\mathfrak{H}$, so it has only finitely many zeros and poles in the compact set $F \cap \{\mathrm{Im} z \leq M\}$, and none above (Lemma 4.1.7). It is then natural to count the zeros and poles of f (with multiplicity) in F , by applying the argument principle to f along the boundary of F , avoiding the possible zeros and poles on the boundary (including i, ρ, ρ^2).

Lemma 4.3.1

Let g be meromorphic at p , and let C_r be an arc of the circle of radius r centred at p , of angle α , traversed counterclockwise. Then

$$\lim_{r \rightarrow 0} \int_{C_r} \frac{g'(z)}{g(z)} dz = \alpha i v_p(g).$$

Proof. Since g is meromorphic at p , $g(z) = (z-p)^k h(z)$ where h is holomorphic and non-vanishing at p and $k = v_p(g)$. Then $g'(z) = k(z-p)^{k-1}h(z) + (z-p)^k h'(z)$, so

$$\frac{g'(z)}{g(z)} = \frac{k(z-p)^{k-1}h(z) + (z-p)^k h'(z)}{(z-p)^k h(z)} = \frac{k}{z-p} + \frac{h'(z)}{h(z)}.$$

Parametrise C_r by $z(\theta) = p + re^{i\theta}$, $\theta_0 \leq \theta \leq \theta_0 + \alpha$, $dz = ire^{i\theta} d\theta$:

$$\int_{C_r} \frac{k}{z-p} dz = \int_{\theta_0}^{\theta_0+\alpha} \frac{k}{re^{i\theta}} ire^{i\theta} d\theta = \int_{\theta_0}^{\theta_0+\alpha} ik d\theta = ik[\theta]_{\theta_0}^{\theta_0+\alpha} = ik\alpha.$$

Since h is holomorphic and non-vanishing at p , h'/h is holomorphic in a small closed disc $\overline{D_{r_0}} = \{|z-p| \leq r_0\}$, hence continuous and bounded there: $|h'/h| \leq M$. For $r \leq r_0$ the arc C_r lies in $\overline{D_{r_0}}$ and has length $r\alpha$, so by the ML-inequality $\left| \int_{C_r} \frac{h'}{h} dz \right| \leq M r \alpha \rightarrow 0$ as $r \rightarrow 0$. This proves the lemma. ■

Theorem 4.3.2 (Valence formula)

Let $f \not\equiv 0$ be a modular function of weight k for $\mathrm{SL}_2(\mathbb{Z})$. Then

$$v_\infty(f) + \frac{1}{2}v_i(f) + \frac{1}{3}v_\rho(f) + \sum_{\substack{p \in \Gamma \backslash \mathfrak{H} \\ p \neq i, \rho}} v_p(f) = \frac{k}{12},$$

where the sum runs over a set of representatives of the Γ -orbits in $\mathfrak{H}$ other than those of i and ρ (e.g. over the points of F , counting each pair of boundary points identified by T or S only once).

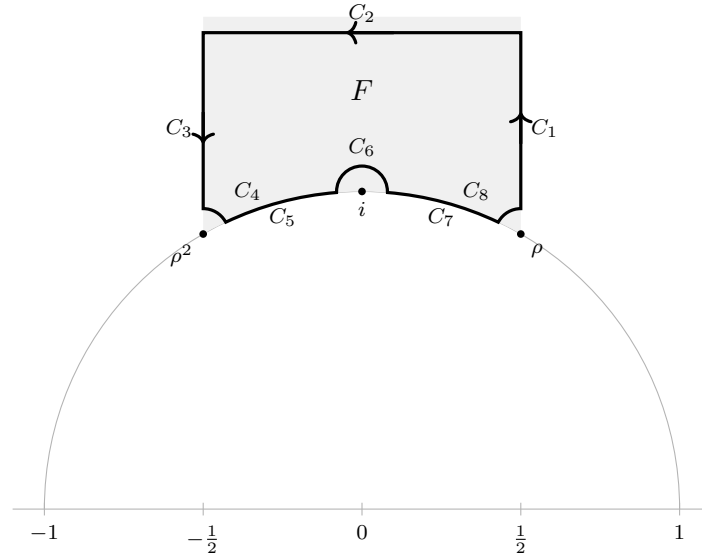


Figure 4.1: The contour $\mathcal{C} = C_1 \cup \cdots \cup C_8$ used in the proof of the valence formula: C_2 at height $\mathrm{Im} z = M$, small arcs C_4, C_6, C_8 around ρ^2, i, ρ , and the arcs C_5, C_7 of the unit circle.

Proof. We count the zeros and poles of f in F using the contour $\mathcal{C} = \bigcup_{i=1}^8 C_i$ of Figure 4.1, traversed counterclockwise:

- (1) the top side C_2 is the segment $\mathrm{Im} z = M$ from $\frac{1}{2} + iM$ to $-\frac{1}{2} + iM$, where M is so large that f has no zeros or poles with $\mathrm{Im} z \geq M$ (Lemma 4.1.7);
- (2) C_3 and C_1 are the vertical sides $\mathrm{Re} z = \mp \frac{1}{2}$ (downwards, resp. upwards);
- (3) C_4, C_6, C_8 are small arcs of radius r around ρ^2, i, ρ , cutting these points out of the region, and C_5, C_7 are the remaining parts of the unit circle, from ρ^2 towards i and from i towards ρ ;
- (4) if f has further zeros or poles on the boundary, we go around them along small semicircles, symmetrically: if we bypass z_0 on C_3 to the inside, we bypass $z_0 + 1$ on C_1 to the outside, and if we bypass z_0 on C_5 to the inside, we bypass $-1/z_0$ on C_7 to the outside. Then each pair of Γ -equivalent boundary points is counted exactly once.

By the argument principle (recalled in Chapter 2),

$$\frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f'(z)}{f(z)} dz = \#\{\text{zeros}\} - \#\{\text{poles}\} \text{ inside } \mathcal{C} = \sum_{\substack{p \in \Gamma \backslash \mathfrak{H} \\ p \neq i, \rho}} v_p(f)$$

for r small enough, and $\oint_C = \int_{C_1} + \int_{C_2} + \cdots + \int_{C_8}$. We now compute the pieces.

(a) C_1 and C_3 . Since $f(z+1) = f(z)$, also $\frac{f'}{f}(z+1) = \frac{f'}{f}(z)$; C_1 is the translate of C_3 by 1 traversed in the opposite direction (with the symmetric indentations of (4)). Hence $\int_{C_1} \frac{f'}{f} + \int_{C_3} \frac{f'}{f} = 0$.

(b) C_2 . Since f is a modular function, $\tilde{f}(q) = f(z) = \sum_{n \geq n_0} a_n q^n$. Along C_2 , $\text{Im } z = M$, so $|q| = e^{-2\pi M}$: as z runs from $\frac{1}{2} + iM$ to $-\frac{1}{2} + iM$, $q = e^{2\pi iz}$ runs once around the circle $|q| = e^{-2\pi M}$ clockwise. Since f has no poles or zeros with $\text{Im } z \geq M$ other than at ∞ , $\tilde{f}$ has none in $|q| \leq e^{-2\pi M}$ except possibly at 0. Also $f'(z) = \tilde{f}'(q) \frac{dq}{dz}$, so $\frac{f'(z)}{f(z)} dz = \frac{\tilde{f}'(q)}{\tilde{f}(q)} dq$, and

$$\frac{1}{2\pi i} \int_{C_2} \frac{f'(z)}{f(z)} dz = \frac{1}{2\pi i} \oint_{|q|=e^{-2\pi M}, \text{ clockwise}} \frac{\tilde{f}'(q)}{\tilde{f}(q)} dq = -v_0(\tilde{f}) = -v_\infty(f).$$

(c) C_4, C_6, C_8 . These are traversed clockwise. As $r \rightarrow 0$, the angle of C_4 (at ρ^2) and of C_8 (at ρ) tends to the interior angle $\frac{\pi}{3}$ of F at its corners, and the angle of C_6 tends to π . By Lemma 4.3.1 (with a minus sign for the clockwise orientation),

$$\frac{1}{2\pi i} \left(\int_{C_4} + \int_{C_6} + \int_{C_8} \right) \frac{f'}{f} dz \rightarrow \frac{1}{2\pi i} \left(-\frac{\pi}{3} i v_{\rho^2}(f) - \pi i v_i(f) - \frac{\pi}{3} i v_\rho(f) \right) = -\frac{1}{3} v_\rho(f) - \frac{1}{2} v_i(f),$$

since $v_{\rho^2}(f) = v_\rho(f)$ ($\rho^2 = T^{-1}\rho$, Lemma 4.1.7).

(d) C_5 and C_7 . Since f is modular, $f(-1/z) = z^k f(z)$. Differentiating,

$$f' \left(-\frac{1}{z} \right) \frac{1}{z^2} = k z^{k-1} f(z) + z^k f'(z) \implies \frac{f'(-1/z)}{z^2 f(-1/z)} = \frac{k z^{k-1} f(z) + z^k f'(z)}{z^k f(z)} = \frac{k}{z} + \frac{f'(z)}{f(z)}, \quad (\star)$$

dividing by $f(-1/z) = z^k f(z)$. Now change variables $w = -\frac{1}{z}$ in the integral over C_7 : $dw = \frac{dz}{z^2}$, and $z \mapsto -\frac{1}{z}$ maps C_7 (from i to ρ) onto C_5 traversed backwards (from i to ρ^2), since $S(i) = i$, $S(\rho) = -\bar{\rho} = \rho^2$ and S preserves the unit circle. Thus

$$\int_{C_7} \frac{f'(w)}{f(w)} dw = - \int_{C_5} \frac{f'(-1/z)}{f(-1/z)} \frac{dz}{z^2} \stackrel{(\star)}{=} - \int_{C_5} \left(\frac{k}{z} + \frac{f'(z)}{f(z)} \right) dz,$$

so

$$\frac{1}{2\pi i} \left(\int_{C_5} + \int_{C_7} \right) \frac{f'}{f} dz = -\frac{1}{2\pi i} \int_{C_5} \frac{k}{z} dz.$$

Parametrise C_5 by $z = e^{i\theta}$, $\frac{dz}{z} = i d\theta$, with θ from (nearly) $\frac{2\pi}{3}$ to (nearly) $\frac{\pi}{2}$:

$$-\frac{k}{2\pi i} \int_{2\pi/3}^{\pi/2} i d\theta = -\frac{k}{2\pi i} \left(\frac{\pi i}{2} - \frac{2\pi i}{3} \right) = -\frac{k}{2\pi i} \left(-\frac{2\pi i}{12} \right) = \frac{k}{12} \quad (r \rightarrow 0).$$

Summing all contours, and letting $r \rightarrow 0$,

$$\sum_{\substack{p \in \Gamma \backslash \mathfrak{H} \\ p \neq i, \rho}} v_p(f) = 0 - v_\infty(f) - \frac{1}{3} v_\rho(f) - \frac{1}{2} v_i(f) + \frac{k}{12},$$

which is the valence formula. ■

Correction. Several small slips in my handwritten proof: “ $\int_{C_1} + \int_{C_2} = 0$ ” should read $\int_{C_1} + \int_{C_3} = 0$; the integral along C_2 equals $-2\pi i v_\infty(f)$ (the factor $\frac{1}{2\pi i}$ was missing); and “ $\text{SL}_2(\mathbb{Z}) \backslash \mathfrak{H}$ is a fundamental domain” should say that the sum runs over representatives of $\text{SL}_2(\mathbb{Z}) \backslash \mathfrak{H}$, e.g. points of F .

4.4 Consequences of the valence formula

Corollary 4.4.1 (1) $M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$.

(2) If $k < 0$, k odd, or $k = 2$, then $M_k(\mathrm{SL}_2(\mathbb{Z})) = \{0\}$.

(3) If $k < 12$ or $k = 14$, then $S_k(\mathrm{SL}_2(\mathbb{Z})) = \{0\}$; $S_{12}(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}\Delta$; and for all $k \geq 12$, $S_k(\mathrm{SL}_2(\mathbb{Z})) = \Delta M_{k-12}(\mathrm{SL}_2(\mathbb{Z}))$.

(4) $M_k(\mathrm{SL}_2(\mathbb{Z})) = S_k(\mathrm{SL}_2(\mathbb{Z})) \oplus \mathbb{C}E_k$ for even $k \geq 4$.

(5) If $k = 4, 6, 8, 10, 14$, then $M_k(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}E_k$.

Moreover, Δ has no zeros in $\mathfrak{H}$; E_4 has a simple zero at ρ (and its Γ -orbit) and no other zeros; E_6 has a simple zero at i (and its orbit) and no other zeros.

Proof. We write $M_k = M_k(\mathrm{SL}_2(\mathbb{Z}))$, $S_k = S_k(\mathrm{SL}_2(\mathbb{Z}))$. For a nonzero modular form all $v_p \geq 0$.

(1) $\mathbb{C} \subseteq M_0$ is clear. Let $0 \neq f \in M_0$ be arbitrary, and let $c = f(\infty)$ be the value of f at the cusp. Since the constant c is a modular form of weight 0, also $g(z) = f(z) - c \in M_0$. Then $g(\infty) = f(\infty) - c = c - c = 0$, so g has a zero of order at least 1 at ∞ , i.e. $v_\infty(g) \geq 1$. Suppose, by way of contradiction, that $g \neq 0$. Then by the valence formula

$$v_\infty(g) + \frac{1}{2}v_i(g) + \frac{1}{3}v_\rho(g) + \sum v_p(g) = \frac{0}{12} = 0 \implies 1 \leq 0,$$

a contradiction. So $g \equiv 0$, i.e. $f = c$, and $M_0 \subseteq \mathbb{C}$.

(2) For odd k see the remark after Lemma 4.1.3. If $k < 0$, then $\frac{k}{12} < 0$, while the left-hand side of the valence formula is nonnegative for a nonzero modular form; so no $f \neq 0$ exists. For $k = 2$ the right-hand side is $\frac{1}{6}$; multiplying by 6,

$$6v_\infty(f) + 3v_i(f) + 2v_\rho(f) + 6 \sum v_p(f) = 1.$$

Since all v 's are nonnegative integers, the left-hand side is 0 or ≥ 2 ; no sum adds up to 1. Hence $M_2 = \{0\}$.

(3) Let $0 \neq f \in S_k$; then f vanishes at ∞ , i.e. $v_\infty(f) \geq 1$.

- If $k < 12$, then $\frac{k}{12} < 1$, contradicting that the left-hand side is at least 1. So $S_k = \{0\}$.
- If $k = 14$, the right-hand side is $\frac{7}{6}$; multiplying by 6: $6v_\infty + 3v_i + 2v_\rho + 6 \sum v_p = 7$ with $v_\infty \geq 1$, i.e. $3v_i + 2v_\rho + 6(v_\infty - 1) + 6 \sum v_p = 1$; again there is no solution. So $S_{14} = \{0\}$.

Δ has no zeros in $\mathfrak{H}$: $\Delta \in S_{12}$ with $v_\infty(\Delta) = 1$ (Proposition 4.2.10), and the valence formula $1 + \frac{1}{2}v_i + \frac{1}{3}v_\rho + \sum v_p = 1$ forces all other $v_p(\Delta) = 0$.

$S_k = \Delta M_{k-12}$ for $k \geq 12$ (this includes $S_{12} = \mathbb{C}\Delta$, since $M_0 = \mathbb{C}$). “ $\supseteq$ ”: for $g \in M_{k-12}$, Δg is weakly modular of weight $12 + (k - 12) = k$, holomorphic on $\mathfrak{H}$ (both factors are), and $v_\infty(\Delta g) = v_\infty(\Delta) + v_\infty(g) \geq 1 + 0 = 1$. Thus $\Delta g \in S_k$. “ $\subseteq$ ”: for $f \in S_k$ define $g = f/\Delta$. Since Δ has no zeros in $\mathfrak{H}$, g is well defined and holomorphic on $\mathfrak{H}$. As $v_\infty(\Delta) = 1$ and $v_\infty(f) \geq 1$, $v_\infty(g) = v_\infty(f) - v_\infty(\Delta) \geq 1 - 1 = 0$, so g is holomorphic at ∞ . And

$$g(\gamma z) = \frac{(cz + d)^k f(z)}{(cz + d)^{12} \Delta(z)} = (cz + d)^{k-12} \frac{f(z)}{\Delta(z)} = (cz + d)^{k-12} g(z).$$

So $g \in M_{k-12}$ and $f = \Delta g \in \Delta M_{k-12}$.

(4) Let $f \in M_k$, $f(z) = a_0 + a_1 q + a_2 q^2 + \dots$, and define $g = f - a_0 E_k$. Since $E_k = 1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n$, we obtain

$$g = (a_0 + a_1 q + \dots) - a_0 \left(1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n \right) = \sum_{n \geq 1} \left(a_n + a_0 \frac{2k}{B_k} \sigma_{k-1}(n) \right) q^n,$$

and g is holomorphic (on $\mathfrak{H}$ and at ∞), so $g \in S_k$ and $f = g + a_0 E_k$. Thus $M_k \subseteq S_k + \mathbb{C}E_k$; the reverse inclusion is clear since $S_k \subseteq M_k$ and $E_k \in M_k$. The sum is direct: clearly $\{0\} \subseteq S_k \cap \mathbb{C}E_k$; and if $h \in S_k \cap \mathbb{C}E_k$, then $h = cE_k$ and $h \in S_k$ implies that the constant term of h vanishes, i.e. $0 = c \cdot 1$, so $c = 0$ and $h = 0$.

(5) By (3) and (4), $M_k = S_k \oplus \mathbb{C}E_k = \{0\} \oplus \mathbb{C}E_k$ for $k = 4, 6, 8, 10, 14$. (My notes argue via the valence formula, as follows; the argument also gives the zeros of E_4, E_6 .) For $k = 4$: for $0 \neq f \in M_4$, $v_\infty(f) + \frac{1}{2}v_i(f) + \frac{1}{3}v_\rho(f) + \sum v_p(f) = \frac{1}{3}$ forces $v_\rho(f) = 1$ and everything else 0. In particular E_4 has a simple zero exactly at (the orbit of) ρ , and f, E_4 have the same zeros, so $f/E_4 \in M_0 = \mathbb{C}$ by the claim below. Similarly: $k = 6$, right-hand side $\frac{1}{2}$: $v_i(f) = 1$; $k = 8$, $\frac{2}{3}$: $v_\rho(f) = 2$; $k = 10$, $\frac{5}{6}$: $v_\rho(f) = v_i(f) = 1$; $k = 14$, $\frac{7}{6}$: $v_i(f) = 1, v_\rho(f) = 2$.

Claim. If $f_1, f_2 \in M_k$ with $v_p(f_1) = v_p(f_2)$ for all $p \in \mathfrak{H} \cup \{\infty\}$, $f_2 \neq 0$, then $f_1/f_2 \in M_0 = \mathbb{C}$. Indeed, let $g = f_1/f_2$.

- (i) *Weak modularity*: for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, $g(\gamma z) = \frac{f_1(\gamma z)}{f_2(\gamma z)} = \frac{(cz+d)^k f_1(z)}{(cz+d)^k f_2(z)} = \frac{f_1(z)}{f_2(z)} = (cz+d)^0 g(z)$.
- (ii) *Holomorphy on $\mathfrak{H}$* : at each p , the zero of f_2 is cancelled by the zero of f_1 of the same order.
- (iii) *At ∞* : with $n = v_\infty(f_1) = v_\infty(f_2)$,

$$\tilde{g}(q) = \frac{a_n q^n + a_{n+1} q^{n+1} + \cdots}{b_n q^n + b_{n+1} q^{n+1} + \cdots} = \frac{q^n (a_n + a_{n+1} q + \cdots)}{q^n (b_n + b_{n+1} q + \cdots)} = \frac{a_n + a_{n+1} q + \cdots}{b_n + b_{n+1} q + \cdots} \xrightarrow{q \rightarrow 0} \frac{a_n}{b_n},$$

with $b_n \neq 0$; so g is holomorphic at ∞ .

Hence $g \in M_0 = \mathbb{C}$. ■

Proposition 4.4.2

$E_4(\rho) = 0$ and $E_6(i) = 0$.

Proof. This follows from Corollary 4.4.1, but there is also a direct proof from the transformation law $E_k(-1/z) = z^k E_k(z)$. For E_6 at $z = i$: $-1/i = i$, so $E_6(i) = i^6 E_6(i) = -E_6(i)$, hence $E_6(i) = 0$. For E_4 : note that $\rho = \frac{1}{2} + \frac{\sqrt{3}}{2}i = e^{\pi i/3}$, so $\rho^3 = -1$, $\rho^4 = -\rho$, and ρ satisfies $\rho^2 - \rho + 1 = 0$. Also $-\frac{1}{\rho} = -\bar{\rho} = \rho^2 = \rho - 1$ (as $|\rho| = 1$). Then

$$E_4\left(-\frac{1}{\rho}\right) = \rho^4 E_4(\rho) \implies E_4(\rho - 1) = -\rho E_4(\rho) \implies E_4(\rho) = -\rho E_4(\rho),$$

using $E_4(\rho - 1) = E_4(\rho)$ (periodicity). Since $-\rho \neq 1$, $E_4(\rho) = 0$. ■

Notes. In my handwritten notes this was written as $E_4(e^{\pi i}/e^{\pi i/3}) = E_4(\rho^2)$ and then $\rho^2 = -\rho + 1$; the correct relation is $\rho^2 = \rho - 1$ (from $\rho^2 - \rho + 1 = 0$), i.e. $\rho^2 = T^{-1}\rho$. The conclusion is the same.

4.5 Dimension formula and generators

Proposition 4.5.1

For every integer $j \geq 0$,

$$\dim M_{2j}(\Gamma) = \begin{cases} \lfloor \frac{j}{6} \rfloor, & j \equiv 1 \pmod{6}, \\ \lfloor \frac{j}{6} \rfloor + 1, & j \not\equiv 1 \pmod{6}. \end{cases}$$

Proof. We use induction on $j \geq 0$.

- (1) $j = 0$: $M_0(\Gamma) = \mathbb{C}$ and $0 \not\equiv 1 \pmod{6}$; indeed $\dim M_0 = 1 = \lfloor \frac{0}{6} \rfloor + 1$.
- (2) $j = 1$: $M_2(\Gamma) = \{0\}$ and $1 \equiv 1 \pmod{6}$; indeed $\dim M_2 = 0 = \lfloor \frac{1}{6} \rfloor$.
- (3) $j = 2, 3, 4, 5$: by Corollary 4.4.1(4) and (3), $M_{2j} = S_{2j} \oplus \mathbb{C}E_{2j}$ with $S_{2j} = \{0\}$, so $\dim M_{2j} = 0 + 1 = 1 = \lfloor \frac{j}{6} \rfloor + 1$.
- (4) $j = 6$: $M_{12} = S_{12} \oplus \mathbb{C}E_{12}$ with $S_{12} = \mathbb{C}\Delta$, so $\dim M_{12} = 1 + 1 = 2 = \lfloor \frac{6}{6} \rfloor + 1$.
- (5) $j = 7$: $M_{14} = S_{14} \oplus \mathbb{C}E_{14}$, $S_{14} = \{0\}$, so $\dim M_{14} = 1 = \lfloor \frac{7}{6} \rfloor$, and $7 \equiv 1 \pmod{6}$.
- (6) Assume $j \geq 8$ and that the formula holds for all smaller values. Since $2j > 14$, Corollary 4.4.1 gives $M_{2j} = S_{2j} \oplus \mathbb{C}E_{2j}$ and $S_{2j} = \Delta M_{2j-12} = \Delta M_{2(j-6)}$, and $g \mapsto \Delta g$ is an isomorphism $M_{2(j-6)} \rightarrow S_{2j}$ (injective as $\Delta \neq 0$). Hence $\dim M_{2j} = \dim M_{2(j-6)} + 1$, and $j - 6 \equiv j \pmod{6}$:
 - if $j \equiv 1 \pmod{6}$: $\dim M_{2j} = \lfloor \frac{j-6}{6} \rfloor + 1 = \lfloor \frac{j}{6} \rfloor - 1 + 1 = \lfloor \frac{j}{6} \rfloor$;
 - if $j \not\equiv 1 \pmod{6}$: $\dim M_{2j} = \lfloor \frac{j-6}{6} \rfloor + 1 + 1 = \lfloor \frac{j}{6} \rfloor + 1$.

By induction, the statement holds for all $j \geq 0$. ■

Correction. In my handwritten notes the case $j = 6$ ($k = 12$) was included among $j = 2, \dots, 6$ with $\dim S_{2j} = 0$. But $S_{12} = \mathbb{C}\Delta$ has dimension 1, so $\dim M_{12} = 2$, which agrees with the formula $\lfloor \frac{6}{6} \rfloor + 1 = 2$. Also “ $\lfloor \frac{2j}{6} \rfloor$ ” there should read $\lfloor \frac{j}{6} \rfloor$.

Proposition 4.5.2

$E_4E_6 = E_{10}$, $E_4^2 = E_8$, and $E_6E_8 = E_4E_{10} = E_{14}$.

Proof. Since $E_4E_6 \in M_{10}$ and $\dim M_{10} = \lfloor \frac{5}{6} \rfloor + 1 = 1$, there exists $c \in \mathbb{C}$ such that $E_4E_6 = cE_{10}$. Comparing constant terms ($1 \cdot 1 = c \cdot 1$) gives $c = 1$. As a check, using Example 4.2.9,

$$\begin{aligned} E_4E_6 &= (1 + 240q + 2160q^2 + 6720q^3 + \dots)(1 - 504q - 16632q^2 - 122976q^3 - \dots) \\ &= 1 + (240 - 504)q + (2160 - 240 \cdot 504 - 16632)q^2 + \dots \\ &= 1 - 264q - 135432q^2 - \dots = E_{10}. \end{aligned}$$

The other identities are proved in the same way, using $\dim M_8 = \dim M_{14} = 1$. ■

Proposition 4.5.3

Every $f \in M_k(\Gamma)$ can be written as $f = \sum_{4i+6j=k} c_{ij} E_4^i E_6^j$ with $c_{ij} \in \mathbb{C}$ (sum over integers $i, j \geq 0$).

Proof. We use strong induction on even $k \geq 0$ (for odd k , $k < 0$ and $k = 2$ we have $M_k = \{0\}$, an empty sum).

- (1) $k = 0$: $M_0 = \mathbb{C}$, and $E_4^0 E_6^0 = 1$.
- (2) For $k = 4, 6, 8, 10, 14$ we have $E_4, E_6, E_4^2, E_4 E_6, E_4^2 E_6 \in M_k(\Gamma)$ respectively, and these are nonzero, so by Corollary 4.4.1(5) ($M_k = \mathbb{C}E_k$, one-dimensional) they are generators of $M_4, M_6, M_8, M_{10}, M_{14}$. Thus for $f \in M_k$: $f = c_{1,0}E_4, c_{0,1}E_6, c_{2,0}E_4^2, c_{1,1}E_4 E_6, c_{2,1}E_4^2 E_6$ respectively.
- (3) Let $k = 12$ or $k > 14$, and assume the statement for all even $k' < k$. Since $k \geq 4$ is even, there are integers $i, j \geq 0$ with $4i + 6j = k$ (e.g. $j = 0$ if $4 \mid k$, and $j = 1$ otherwise), and $E_4^i E_6^j \in M_k(\Gamma)$ has constant term 1. Let $f \in M_k(\Gamma)$ with constant term a_0 , and put $g = f - a_0 E_4^i E_6^j$. Then $g \in S_k(\Gamma) = \Delta M_{k-12}(\Gamma)$, so $g = \Delta f_1$ with $f_1 \in M_{k-12}$. We obtain

$$f = a_0 E_4^i E_6^j + \Delta f_1 = a_0 E_4^i E_6^j + \frac{(2\pi)^{12}}{1728} (E_4^3 - E_6^2) f_1.$$

Since $f_1 \in M_{k-12}$ and $k - 12 < k$, by assumption $f_1 = \sum_{4p+6q'=k-12} d_{p,q'} E_4^p E_6^{q'}$. Substituting,

$$f = a_0 E_4^i E_6^j + \frac{(2\pi)^{12}}{1728} \sum_{4p+6q'=k-12} d_{p,q'} (E_4^{p+3} E_6^{q'} - E_4^p E_6^{q'+2}).$$

Observe that the weights are $4(p+3) + 6q' = (4p+6q') + 12 = (k-12) + 12 = k$ and $4p+6(q'+2) = k$, as well as $4i+6j = k$. Thus the statement also holds for k .

By induction, the statement is true for all even $k \geq 0$. ■

Correction. In my handwritten proof g was defined as $f - E_4^i E_6^j$; to obtain a cusp form one must subtract $a_0 E_4^i E_6^j$, where $a_0 = f(\infty)$. Also, “ $\dim(E_4^{p+3} E_6^q)$ ” there means the weight.

4.6 The j -invariant

Definition 4.6.1 (j -invariant)

$$j(z) = \frac{1728 g_2(z)^3}{\Delta(z)}, \quad z \in \mathfrak{H}.$$

By Proposition 4.2.10,

$$j(z) = \frac{1728 \left(\frac{4}{3} \pi^4 E_4(z)\right)^3}{\frac{(2\pi)^{12}}{1728} (E_4(z)^3 - E_6(z)^2)} = \frac{1728^2 \cdot 4^3 \pi^{12}}{3^3 \cdot 2^{12} \pi^{12}} \cdot \frac{E_4(z)^3}{E_4(z)^3 - E_6(z)^2} = 1728 \frac{E_4(z)^3}{E_4(z)^3 - E_6(z)^2},$$

since $\frac{1728 \cdot 64}{27 \cdot 4096} = 1$. This is a modular function of weight 0: it is holomorphic on $\mathfrak{H}$ (as Δ has no zeros there), and for $\gamma \in \Gamma$,

$$j(\gamma z) = 1728 \frac{E_4(\gamma z)^3}{E_4(\gamma z)^3 - E_6(\gamma z)^2} = 1728 \frac{(cz+d)^{12} E_4(z)^3}{(cz+d)^{12} (E_4(z)^3 - E_6(z)^2)} = j(z).$$

Its q -expansion is $j = \frac{1728(1+720q+\dots)}{1728(q-24q^2+\dots)} = \frac{1}{q} + 744 + \dots$, so $v_\infty(j) = -1$.

Proposition 4.6.2

The function j induces a bijection $\Gamma \backslash \mathfrak{H} \rightarrow \mathbb{C}$.

Proof. Since $j(\gamma z) = j(z)$, the map is well defined. Let $c \in \mathbb{C}$. We show that $j(z) = c$ has exactly one solution modulo Γ . Consider $f_c = 1728E_4^3 - c(E_4^3 - E_6^2) \in M_{12}(\Gamma)$. Since $E_4^3 - E_6^2$ has no zeros in $\mathfrak{H}$, for $z \in \mathfrak{H}$: $j(z) = c \iff f_c(z) = 0$. The constant term of f_c is $1728 \neq 0$, so $f_c \neq 0$ and $v_\infty(f_c) = 0$. The valence formula for f_c (weight 12) gives

$$\frac{1}{2}v_i(f_c) + \frac{1}{3}v_\rho(f_c) + \sum_{p \neq i, \rho} v_p(f_c) = 1.$$

With nonnegative integers, the only solutions are: $v_p(f_c) = 1$ for exactly one orbit $p \neq i, \rho$ and all others 0; or $v_i(f_c) = 2$ and all others 0; or $v_\rho(f_c) = 3$ and all others 0. In each case f_c vanishes on exactly one Γ -orbit. ■

Proposition 4.6.3

The modular functions of weight 0 for Γ are precisely the rational functions of j .

Proof. Rational functions of j are clearly modular functions of weight 0. Conversely, we first show: if $g \in M_{12N}(\Gamma)$, $N \geq 0$, then g/Δ^N is a polynomial in j of degree $\leq N$. For $N = 0$, $g \in M_0 = \mathbb{C}$. For $N \geq 1$, $E_4^{3N} \in M_{12N}$ has constant term 1, so $g - a_0 E_4^{3N} \in S_{12N} = \Delta M_{12N-12}$, i.e. $g = a_0 E_4^{3N} + \Delta g_1$ with $g_1 \in M_{12(N-1)}$. Then

$$\frac{g}{\Delta^N} = a_0 \left(\frac{E_4^3}{\Delta} \right)^N + \frac{g_1}{\Delta^{N-1}}, \quad \frac{E_4^3}{\Delta} = \frac{1728E_4^3}{(2\pi)^{12}(E_4^3 - E_6^2)} = \frac{j}{(2\pi)^{12}},$$

and by induction g_1/Δ^{N-1} is a polynomial in j of degree $\leq N-1$.

Now let $f \neq 0$ be a modular function of weight 0. Its poles in $\mathfrak{H}$ lie in finitely many Γ -orbits (Lemma 4.1.7), say those of $p_1, \dots, p_r$, with pole orders $n_1, \dots, n_r$ at $p_1, \dots, p_r$. Since $j - j(p_l)$ is holomorphic on $\mathfrak{H}$ and vanishes at p_l (hence on its orbit), the function $h = f \cdot \prod_{l=1}^r (j - j(p_l))^{n_l}$ is a modular function of weight 0 which is holomorphic on $\mathfrak{H}$. Let $N = \max(0, -v_\infty(h))$. Then $\Delta^N h$ is weakly modular of weight $12N$, holomorphic on $\mathfrak{H}$, and $v_\infty(\Delta^N h) = N + v_\infty(h) \geq 0$; so $\Delta^N h \in M_{12N}$ and $h = \Delta^N h / \Delta^N \in \mathbb{C}[j]$. Therefore $f = h / \prod_l (j - j(p_l))^{n_l} \in \mathbb{C}(j)$. ■

Proposition 4.6.4 (Uniformisation)

For any $A, B \in \mathbb{C}$ with $A^3 - 27B^2 \neq 0$ there exists a lattice $L = \lambda L_z$ ($\lambda \in \mathbb{C}^\times$, $z \in \mathfrak{H}$) such that $g_2(L) = A$ and $g_3(L) = B$. Consequently every elliptic curve $y^2 = 4x^3 - Ax - B$ over $\mathbb{C}$ is isomorphic to a torus $\mathbb{C}/L$ (Theorem 2.5.2).

Proof. For a lattice $L = \{\alpha\omega_1 + b\omega_2\}$ we may (swapping ω_1, ω_2 if necessary) factor out ω_2 : let $\lambda = \omega_2$ and $z = \omega_1/\omega_2 \in \mathfrak{H}$; then $L = \lambda L_z = \lambda\{az + b\}$. By definition of $G_k(L)$,

$$g_2(\lambda L_z) = 60G_4(\lambda L_z) = 60 \sum_{0 \neq \ell \in L_z} \frac{1}{(\lambda \ell)^4} = \lambda^{-4} (60G_4(L_z)) = \lambda^{-4} g_2(z),$$

and similarly $g_3(\lambda L_z) = \lambda^{-6} g_3(z)$. We already know $g_2(z) = \frac{4}{3}\pi^4 E_4(z)$ and $g_3(z) = \frac{8}{27}\pi^6 E_6(z)$. Thus the statement can be restated as: there exist $\lambda \in \mathbb{C}^\times$ and $z \in \mathfrak{H}$ with

$$E_4(z) = \lambda^4 \left(\frac{3}{4\pi^4} A \right) = \lambda^4 a, \quad E_6(z) = \lambda^6 \left(\frac{27}{8\pi^6} B \right) = \lambda^6 b,$$

where $a = \frac{3A}{4\pi^4}$ and $b = \frac{27B}{8\pi^6}$. Then $a^3 - b^2 = \frac{27A^3}{64\pi^{12}} - \frac{729B^2}{64\pi^{12}} = \frac{27}{64\pi^{12}} (A^3 - 27B^2)$, so $A^3 \neq 27B^2$ becomes $a^3 \neq b^2$. We distinguish three cases.

- $a = 0$: then $b \neq 0$. Take $z = \rho$: $E_4(\rho) = 0 = \lambda^4 a$, and $E_6(\rho) \neq 0$ (the only zeros of E_6 are on the orbit of i , Corollary 4.4.1); choose λ with $\lambda^6 = E_6(\rho)/b$.

- $b = 0$: then $a \neq 0$. Take $z = i$: $E_6(i) = 0$, $E_4(i) \neq 0$; choose λ with $\lambda^4 = E_4(i)/a$.
- $a, b \neq 0$: by Proposition 4.6.2 there is $z \in \mathfrak{H}$ with $j(z) = \frac{1728a^3}{a^3 - b^2}$ (note $a^3 \neq b^2$). Since $j = 1728 \frac{E_4^3}{E_4^3 - E_6^2}$, this means $E_4(z)^3(a^3 - b^2) = a^3(E_4(z)^3 - E_6(z)^2)$, i.e.

$$b^2 E_4(z)^3 = a^3 E_6(z)^2, \quad \text{equivalently} \quad \frac{E_6(z)^2}{E_4(z)^3} = \frac{b^2}{a^3} = 1 - \frac{1728}{j(z)}.$$

In particular $E_4(z) \neq 0$ (otherwise $E_6(z) = 0$ too, impossible as $E_4^3 - E_6^2 = \frac{1728}{(2\pi)^{12}} \Delta \neq 0$). Choose λ with $\lambda^4 = E_4(z)/a$. Then $\lambda^{12} = E_4(z)^3/a^3$ and $(\lambda^6 b)^2 = \lambda^{12} b^2 = E_4(z)^3 b^2/a^3 = E_6(z)^2$, so $E_6(z) = \pm \lambda^6 b$. If the sign is $-$, replace λ by $i\lambda$ (which does not change λ^4 and changes the sign of λ^6).

In all cases $E_4(z) = \lambda^4 a$ and $E_6(z) = \lambda^6 b$, hence $g_2(\lambda L_z) = A$ and $g_3(\lambda L_z) = B$. ■

Correction. In my handwritten proof the quotient was written as $\frac{E_6^3}{E_4^2}$ instead of $\frac{E_6(z)^2}{E_4(z)^3}$, the condition $A^3 \neq 27B^2$ as “ $A^3 \neq 27B$ ”, and the constants as $\frac{3}{4}\pi^4 A$, $\frac{27}{8}\pi^6 B$ instead of $\frac{3A}{4\pi^4}$, $\frac{27B}{8\pi^6}$ (because $g_2(\lambda L_z) = \lambda^{-4}g_2(z)$). The cases $a = 0$ or $b = 0$ (where $j(z) = 0$ resp. 1728) and the choice of λ were not treated. The correspondence “lattices $\lambda L_z \leftrightarrow \mathrm{SL}_2(\mathbb{Z}) \backslash \mathfrak{H} \leftrightarrow$ elliptic curves”, which my notes indicate with a pointer to Diamond–Shurman, is the starting point of the modular interpretation of modular curves [1, §1.5].

4.7 The Eisenstein series of weight 2

Motivation. For $k = 2$ the Eisenstein series $\sum_{(m,n) \neq (0,0)} \frac{1}{(mz+n)^2}$ is not absolutely convergent, so we cannot ensure the modularity property under $z \mapsto -\frac{1}{z}$ as for $k \geq 4$:

$$\sum_{(m,n) \neq (0,0)} \frac{1}{(m(-\frac{1}{z}) + n)^2} = \sum_{(\frac{1}{z})^2(-m+nz)^2} \frac{1}{(\frac{1}{z})^2(-m+nz)^2} = z^2 \sum_{(m,n) \neq (0,0)} \frac{1}{(nz + (-m))^2},$$

but we cannot rename $m \rightarrow n$, $n \rightarrow -m$, because rearranging a non-absolutely convergent double series may change its value. Still, we can get G_2 “almost” modular.

Definition 4.7.1

The Eisenstein series of weight 2 is defined by the iterated sum

$$G_2(z) = \sum_{m \in \mathbb{Z}} \sum_{\substack{n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} \frac{1}{(mz+n)^2} = 2\zeta(2) + 2 \sum_{m \geq 1} \sum_{n \in \mathbb{Z}} \frac{1}{(mz+n)^2} = 2\zeta(2) - 8\pi^2 \sum_{n \geq 1} \sigma_1(n)q^n,$$

$$\text{and } E_2 = \frac{G_2}{2\zeta(2)} = \frac{3}{\pi^2} G_2 = 1 - 24 \sum_{n \geq 1} \sigma_1(n)q^n.$$

The last expression for G_2 follows exactly as in Theorem 4.2.8, using Lemma 4.2.5(ii) with $k = 2$: $\sum_n (mz+n)^{-2} = (2\pi i)^2 \sum_{r \geq 1} r q^{mr} = -4\pi^2 \sum_r r q^{mr}$; this shows that the iterated sum converges.

Correction. In my handwritten notes the coefficient was written as $-4\pi^2$; the factor 2 in front of $\sum_{m \geq 1}$ gives $-8\pi^2$. This is consistent with $E_2 = 1 - 24 \sum \sigma_1(n)q^n$ and with the computation of η'/η below.

Theorem 4.7.2

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$:

$$G_2(\gamma z) = (cz + d)^2 G_2(z) - 2\pi i c(cz + d).$$

The proof uses *Hecke's trick*: for $s > 0$ consider

$$G_{2,s}(z) = \sum_{(m,n) \neq (0,0)} \frac{1}{(mz + n)^2 |mz + n|^{2s}}.$$

Lemma 4.7.3

For $s > 0$, $G_{2,s}(z)$ converges absolutely, and $G_{2,s}(\gamma z) = (cz + d)^2 |cz + d|^{2s} G_{2,s}(z)$ for $\gamma \in \Gamma$.

Proof. Since $\left| \frac{1}{(mz+n)^2 |mz+n|^{2s}} \right| = \frac{1}{|mz+n|^2 |mz+n|^{2s}} = \frac{1}{|mz+n|^{2+2s}}$ and $2 + 2s > 2$, the series converges absolutely by Lemma 2.3.3 (for the lattice L_z). For $\gamma \in \Gamma$, exactly as in Proposition 4.2.3, $m(\gamma z) + n = \frac{m'z + n'}{cz + d}$ with $(m', n') = (m, n)\gamma$ running over $\mathbb{Z}^2 \setminus \{0\}$ exactly once, so

$$G_{2,s}(\gamma z) = \sum_{(m,n) \neq 0} \frac{(cz + d)^2 |cz + d|^{2s}}{(m'z + n')^2 |m'z + n'|^{2s}} = (cz + d)^2 |cz + d|^{2s} G_{2,s}(z),$$

the rearrangement being allowed by absolute convergence. ■

We need a version of the Poisson summation formula for functions which are not in the Schwartz space.

Lemma 4.7.4 (Poisson summation, C^2 version)

Let $g : \mathbb{R} \rightarrow \mathbb{C}$ be of class C^2 with $|g^{(j)}(x)| \leq \frac{C}{1+x^2}$ for $j = 0, 1, 2$. Then $\sum_{n \in \mathbb{Z}} g(n) = \sum_{n \in \mathbb{Z}} \hat{g}(n)$, both series converging absolutely.

Proof. The proof of Theorem 1.5.1 applies verbatim: $h(x) = \sum_k g(x+k)$ and its first two derivatives converge uniformly on $[0, 1]$ by the M -test, so h is 1-periodic of class C^2 and equals its absolutely convergent Fourier series; the coefficients are $c_m = \hat{g}(m)$ by the same computation. ■

Proof of Theorem 4.7.2. Write $z = u + iv$, $v > 0$, and for $s \geq 0$ put

$$\varphi_s(x) = \frac{1}{(x+z)^2 |x+z|^{2s}}, \quad f_{m,s}(x) = \frac{1}{(x+mz)^2 |x+mz|^{2s}} \quad (m \geq 1).$$

Step 1: the decomposition. Splitting off $m = 0$ and pairing (m, n) with $(-m, -n)$ (allowed by absolute convergence for $s > 0$),

$$G_{2,s}(z) = 2\zeta(2+2s) + 2 \sum_{m \geq 1} \sum_{n \in \mathbb{Z}} f_{m,s}(n).$$

Each $f_{m,s}$ is smooth (as $x + mz \neq 0$ for real x) and $f_{m,s}^{(j)}(x) = O(|x|^{-2-2s-j})$, so Lemma 4.7.4 gives $\sum_n f_{m,s}(n) = \sum_n \hat{f}_{m,s}(n)$. Substituting $x = mt$, $dx = m dt$, $|m| = m$:

$$\hat{f}_{m,s}(y) = \int_{-\infty}^{\infty} \frac{e^{-2\pi i m t y} m dt}{m^2(t+z)^2 m^{2s} |t+z|^{2s}} = \frac{1}{m^{1+2s}} \hat{\varphi}_s(my).$$

Hence

$$G_{2,s}(z) = \underbrace{2\zeta(2+2s)}_A + 2 \underbrace{\sum_{m \geq 1} \frac{\hat{\varphi}_s(0)}{m^{1+2s}}}_B + 2 \underbrace{\sum_{m \geq 1} \frac{1}{m^{1+2s}} \sum_{n \neq 0} \hat{\varphi}_s(mn)}_C.$$

Part A tends to $2\zeta(2)$ as $s \rightarrow 0^+$.

Step 2: part B. Substituting $t = x + u$ and then $t = vw$,

$$\begin{aligned} \hat{\varphi}_s(0) &= \int_{-\infty}^{\infty} \frac{dx}{((x+u)+iv)^2 |(x+u)+iv|^{2s}} = \int_{-\infty}^{\infty} \frac{v dw}{(vw+iv)^2 |vw+iv|^{2s}} \\ &= \frac{1}{v^{1+2s}} \int_{-\infty}^{\infty} \frac{dw}{(w+i)^2 (w^2+1)^s} = \frac{I(s)}{v^{1+2s}}, \end{aligned}$$

using $|w+i|^{2s} = (w^2+1)^s$. So $B = 2\zeta(1+2s) v^{-1-2s} I(s) = [2s \zeta(1+2s)] \cdot v^{-1-2s} \cdot \frac{I(s)}{s}$.

- $2s \zeta(1+2s) \rightarrow 1$ as $s \rightarrow 0^+$, since ζ has a simple pole at 1 with residue 1 (Theorem 1.6.4).
- $I(0) = \int_{-\infty}^{\infty} \frac{dw}{(w+i)^2} = \left[-\frac{1}{w+i} \right]_{-\infty}^{\infty} = 0$, so $\frac{I(s)}{s} \rightarrow I'(0)$ (definition of the derivative). By the Leibniz rule (differentiation under the integral, justified since $\left| \partial_s \frac{1}{(w+i)^2 (w^2+1)^s} \right| = \frac{\log(w^2+1)}{(w^2+1)^{1+s}} \leq \frac{\log(w^2+1)}{w^2+1}$, which is integrable), using $\frac{d}{ds} a^{-s} = -a^{-s} \log a$,

$$I'(0) = \int_{-\infty}^{\infty} \frac{-\log(w^2+1)}{(w+i)^2} dw.$$

Integrate by parts with $U = -\log(w^2+1)$, $dV = (w+i)^{-2} dw$, so $dU = \frac{-2w}{w^2+1} dw$, $V = -\frac{1}{w+i}$:

$$I'(0) = \left[\frac{\log(w^2+1)}{w+i} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-\frac{1}{w+i} \right) \left(\frac{-2w}{w^2+1} \right) dw = 0 - \int_{-\infty}^{\infty} \frac{2w}{(w+i)(w^2+1)} dw,$$

since $w+i$ grows faster than $\log(w^2+1)$. Now $\frac{2w}{(w+i)(w^2+1)} = \frac{2w}{(w+i)^2(w-i)}$; integrate over the boundary Γ_R of the upper half-disc of radius R (the segment $[-R, R]$ and the arc C_R). On C_R , $w = Re^{i\theta}$, $|w \pm i| \geq R-1$, so $\left| \frac{2w}{(w+i)^2(w-i)} \right| \leq \frac{2R}{(R-1)^3}$, and by the ML-inequality $\left| \int_{C_R} \right| \leq \frac{2R}{(R-1)^3} \cdot \pi R = \frac{2\pi R^2}{(R-1)^3} \rightarrow 0$. Inside Γ_R the only pole is the simple pole $w=i$, so by the residue theorem

$$\int_{-\infty}^{\infty} \frac{2w dw}{(w+i)^2(w-i)} = 2\pi i \lim_{w \rightarrow i} (w-i) \frac{2w}{(w+i)^2(w-i)} = 2\pi i \cdot \frac{2i}{(2i)^2} = 2\pi i \left(-\frac{i}{2} \right) = \pi.$$

Hence $I'(0) = -\pi$.

Therefore $B \rightarrow 1 \cdot v^{-1} \cdot (-\pi) = -\frac{\pi}{\text{Im } z}$.

Step 3: part C. First we compute $\hat{\varphi}_0(y) = \int_{-\infty}^{\infty} \frac{e^{-2\pi i xy}}{(x+z)^2} dx$ for $y \neq 0$, by contour integration of $f(w) = \frac{e^{-2\pi i y w}}{(w+z)^2}$, which has a double pole at $w = -z$ in the lower half-plane.

- $y > 0$: close the contour by the lower semicircle C_R^- (clockwise contour $\Gamma_R^- = [-R, R] \cup C_R^-$). For $w = \alpha + i\beta$ with $\beta \leq 0$, $\left| e^{-2\pi i y(\alpha+i\beta)} \right| = e^{2\pi y\beta} \leq 1$, and $|w+z| \geq R-|z|$, so $|f(w)| \leq \frac{1}{(R-|z|)^2}$ on C_R^- and $\left| \int_{C_R^-} f \right| \leq \frac{\pi R}{(R-|z|)^2} \rightarrow 0$. Since Γ_R^- is clockwise,

$$\begin{aligned} \hat{\varphi}_0(y) &= -2\pi i \text{Res}_{w=-z} f = -2\pi i \lim_{w \rightarrow -z} \frac{d}{dw} \left[(w+z)^2 \frac{e^{-2\pi i y w}}{(w+z)^2} \right] \\ &= -2\pi i (-2\pi i y e^{2\pi i y z}) = -4\pi^2 y e^{2\pi i y z}. \end{aligned}$$

- $y < 0$: close by the upper semicircle (counterclockwise); now $|e^{-2\pi i y w}| \leq 1$ for $\text{Im } w \geq 0$, the arc integral tends to 0 in the same way, and there is no pole inside. So $\hat{\varphi}_0(y) = 0$.

Next, for $y \neq 0$ and $0 \leq s \leq \frac{1}{2}$, integrating by parts twice (the boundary terms vanish), $\hat{\varphi}_s(y) = \frac{1}{(2\pi i y)^2} \int e^{-2\pi i x y} \varphi_s''(x) dx$, and $|\varphi_s''(x)| \leq C_z(1+x^2)^{-2}$ uniformly in s ; so $|\hat{\varphi}_s(y)| \leq \frac{C'_z}{y^2}$. Hence the terms of C are bounded by $\frac{2C'_z}{m^{1+2s}m^2n^2} \leq \frac{2C'_z}{m^3n^2}$, which is summable over $m \geq 1, n \neq 0$. Since $\hat{\varphi}_s(y) \rightarrow \hat{\varphi}_0(y)$ as $s \rightarrow 0$ (dominated convergence in the integral), dominated convergence for the double series gives

$$\begin{aligned} \lim_{s \rightarrow 0^+} C &= 2 \sum_{m \geq 1} \frac{1}{m} \sum_{n \neq 0} \hat{\varphi}_0(mn) = 2 \sum_{m \geq 1} \frac{1}{m} \sum_{n \geq 1} (-4\pi^2 mn e^{2\pi i mn z}) \\ &= -8\pi^2 \sum_{m \geq 1} \sum_{n \geq 1} n q^{mn} = -8\pi^2 \sum_{N \geq 1} \sigma_1(N) q^N. \end{aligned}$$

Step 4: conclusion. Combining A, B, C :

$$\lim_{s \rightarrow 0^+} G_{2,s}(z) = 2\zeta(2) - \frac{\pi}{\text{Im } z} - 8\pi^2 \sum_{N \geq 1} \sigma_1(N) q^N = G_2(z) - \frac{\pi}{\text{Im } z} =: G_2^*(z).$$

Letting $s \rightarrow 0^+$ in Lemma 4.7.3 (with $|cz + d|^{2s} \rightarrow 1$) gives $G_2^*(\gamma z) = (cz + d)^2 G_2^*(z)$. Now, using $\text{Im}(\gamma z) = \frac{\text{Im } z}{|cz + d|^2}$,

$$\begin{aligned} G_2(\gamma z) &= \frac{\pi}{\text{Im}(\gamma z)} + G_2^*(\gamma z) = \frac{\pi}{\text{Im } z} |cz + d|^2 + (cz + d)^2 \left(-\frac{\pi}{\text{Im } z} + G_2(z) \right) \\ &= \frac{\pi}{\text{Im } z} (|cz + d|^2 - (cz + d)^2) + (cz + d)^2 G_2(z). \end{aligned}$$

Writing $W = cz + d$, we have $|W|^2 - W^2 = W(\overline{W} - W) = -2i W \text{Im } W$ and $\text{Im } W = c \text{Im } z$. Hence

$$G_2(\gamma z) = \frac{\pi}{\text{Im } z} (cz + d)(-2ic \text{Im } z) + (cz + d)^2 G_2(z) = (cz + d)^2 G_2(z) - 2\pi ic(cz + d). \quad \blacksquare$$

Corollary 4.7.5

$$G_2(-1/z) = z^2 G_2(z) - 2\pi iz \text{ and } E_2(-1/z) = z^2 E_2(z) + \frac{6z}{\pi i}.$$

Proof. Take $\gamma = S$, so $c = 1, d = 0$. For $E_2 = \frac{3}{\pi^2} G_2$: $E_2(-1/z) = z^2 E_2(z) - \frac{3}{\pi^2} 2\pi iz = z^2 E_2(z) - \frac{6iz}{\pi} = z^2 E_2(z) + \frac{6z}{\pi i}$, as $\frac{1}{i} = -i$. ■

Notes. Points in my handwritten proof that needed care: (i) f_m is *not* a Schwartz function ($|x|^N f_m(x) \not\rightarrow 0$ for $N \geq 2$), so Poisson summation is used in the C^2 form of Lemma 4.7.4; (ii) the limit used for part B is $2s\zeta(1+2s) \rightarrow 1$ (not $2\zeta(1+2s) \rightarrow 1$, which I had marked with “?”), combined with $I(s)/s \rightarrow I'(0)$ because $I(0) = 0$ (also marked “?”); (iii) the vanishing of the arc integrals only needs the ML-inequality, Jordan’s lemma is not required; (iv) the interchange of $\lim_{s \rightarrow 0}$ with the double sum in part C is justified by the uniform bound $|\hat{\varphi}_s(y)| \leq C'/y^2$.

4.8 The Dedekind η -function and the product formula for Δ

Definition 4.8.1 (Dedekind eta function)

For $z \in \mathfrak{H}$ and $q = e^{2\pi iz}$,

$$\eta(z) = e^{2\pi iz/24} \prod_{n=1}^{\infty} (1 - e^{2\pi inz}) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n),$$

where $q^{1/24} := e^{2\pi iz/24}$.

Proposition 4.8.2

The product $\prod_{n \geq 1} (1 - q^n)$ converges absolutely and uniformly on compact subsets of $\mathfrak{H}$. Hence η is holomorphic on $\mathfrak{H}$, and $\eta(z) \neq 0$ for all $z \in \mathfrak{H}$.

Proof. Since $q = e^{2\pi iz}$ and $z \in \mathfrak{H}$, $|q| < 1$, so $\sum |q|^n$ converges and the product converges absolutely. For uniform convergence, let $K \subseteq \mathfrak{H}$ be an arbitrary compact subset. The function $z = x + iy \mapsto \text{Im}(z) = y$ is continuous on K , so it has a minimum $y_0 = \min_{z \in K} \text{Im } z$; since $K \subseteq \mathfrak{H}$, $y_0 > 0$. Define $r = e^{-2\pi y_0}$; then $0 < r < 1$, and for all $z \in K$, $|q| = e^{-2\pi \text{Im } z} \leq e^{-2\pi y_0} = r$, so $|q|^n \leq r^n =: M_n$. Since $\sum M_n = \sum r^n < \infty$, by the Weierstrass M -test $\sum |q^n|$ converges uniformly on K , and therefore the product $\prod (1 - q^n)$ converges uniformly on K [6, Ch. 5, Prop. 3.2]; so does $q^{1/24} \prod (1 - q^n)$. A uniformly convergent product of holomorphic functions is holomorphic, and an absolutely convergent product of nonzero factors is nonzero; since $|q^n| < 1$, no factor $1 - q^n$ vanishes. ■

Theorem 4.8.3

$\eta(-1/z) = (-iz)^{1/2} \eta(z)$ for $z \in \mathfrak{H}$, where $(-iz)^{1/2}$ is the principal branch (note $\text{Re}(-iz) = \text{Im } z > 0$).

Proof. Since η is nonzero, we can consider its logarithmic derivative. Logarithmic derivatives of uniformly convergent products of nonvanishing functions may be computed termwise [6, Ch. 5, Prop. 3.2]; with $\frac{dq}{dz} = 2\pi iq$,

$$\frac{\eta'(z)}{\eta(z)} = \frac{2\pi i}{24} + \sum_{n \geq 1} \frac{-2\pi i n q^n}{1 - q^n} = 2\pi i \left(\frac{1}{24} - \sum_{n \geq 1} \sum_{m \geq 1} n q^{nm} \right) = \frac{2\pi i}{24} \left(1 - 24 \sum_{N \geq 1} \sigma_1(N) q^N \right) = \frac{2\pi i}{24} E_2(z),$$

using $\frac{q^n}{1 - q^n} = \sum_{m \geq 1} q^{nm}$. (In my notes this was computed by expanding $\log(1 - q^n) = -\sum_m \frac{q^{nm}}{m}$.) If we apply $z \mapsto -\frac{1}{z}$ and use the transformation of E_2 ,

$$\frac{\eta'(-1/z)}{\eta(-1/z)} = \frac{2\pi i}{24} E_2\left(-\frac{1}{z}\right) = \frac{2\pi i}{24} \left(z^2 E_2(z) + \frac{6z}{\pi i} \right) = z^2 \frac{\eta'(z)}{\eta(z)} + \frac{z}{2}. \quad (\star\star)$$

Now let $h(z) = \frac{\eta(-1/z)}{(-iz)^{1/2} \eta(z)}$, holomorphic and nonzero on $\mathfrak{H}$. Since $\frac{d}{dz} \log(-iz)^{1/2} = \frac{1}{2z}$,

$$\frac{h'(z)}{h(z)} = \frac{1}{z^2} \frac{\eta'(-1/z)}{\eta(-1/z)} - \frac{1}{2z} - \frac{\eta'(z)}{\eta(z)} \stackrel{(\star\star)}{=} \frac{1}{z^2} \left[z^2 \frac{\eta'(z)}{\eta(z)} + \frac{z}{2} \right] - \frac{1}{2z} - \frac{\eta'(z)}{\eta(z)} = 0.$$

Since $\mathfrak{H}$ is connected, h is a constant C . Substituting $z = i$ (where $-1/i = i$ and $(-i \cdot i)^{1/2} = 1$): $C = h(i) = \frac{\eta(i)}{1 \cdot \eta(i)} = 1$. Therefore $\eta(-1/z) = (-iz)^{1/2} \eta(z)$. ■

Correction. In my handwritten proof the logarithm $\log \frac{\eta(-1/z)}{z^{1/2} \eta(z)}$ was used without specifying

ing a branch, and the constant was found as $C = \frac{1}{\sqrt{i}}$, giving “ $\eta(-1/z) = (z/i)^{1/2}\eta(z) = (iz)^{1/2}\eta(z)$ ”. With the principal branch, $(z/i)^{1/2} = (-iz)^{1/2}$, which is the statement of the theorem; the final “ $(iz)^{1/2}$ ” was a slip. Working with h'/h instead of $\log h$ avoids the choice of branch for $\log \eta$.

Theorem 4.8.4

$\eta^{24} \in S_{12}(\Gamma)$, and $\eta^{24} = (2\pi)^{-12}\Delta$. Consequently

$$\Delta(z) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24} = (2\pi)^{12} (q - 24q^2 + 252q^3 - 1472q^4 + \cdots).$$

Proof. Since η is holomorphic on $\mathfrak{H}$, so is η^{24} . Also

$$\eta(z+1) = e^{2\pi i(z+1)/24} \prod_{n \geq 1} (1 - e^{2\pi i n(z+1)}) = e^{2\pi i/24} e^{2\pi i z/24} \prod_{n \geq 1} (1 - e^{2\pi i n z} \underbrace{e^{2\pi i n}}_{=1}) = e^{2\pi i/24} \eta(z),$$

so $\eta^{24}(z+1) = e^{2\pi i} \eta^{24}(z) = \eta^{24}(z)$; and by Theorem 4.8.3, $\eta^{24}(-1/z) = (-iz)^{12} \eta^{24}(z) = (-i)^{12} z^{12} \eta^{24}(z) = z^{12} \eta^{24}(z)$. By Lemma 4.1.3, η^{24} is weakly modular of weight 12. Finally, as $z \rightarrow i\infty$, $q \rightarrow 0$ and

$$\eta^{24}(z) = q \prod_{n \geq 1} (1 - q^n)^{24} = q(1 - q)^{24} (1 - q^2)^{24} \cdots = q - 24q^2 + \cdots,$$

so η^{24} is holomorphic at ∞ and its constant term vanishes: $\eta^{24} \in S_{12}(\Gamma)$. Since $S_{12}(\Gamma) = \mathbb{C}\Delta$ (Corollary 4.4.1), $\eta^{24} = c\Delta$; comparing the coefficients of q (1 and $(2\pi)^{12}$) gives $\eta^{24} = (2\pi)^{-12}\Delta$. Expanding the product gives the coefficients 252 and -1472 . ■

4.9 Derivatives of modular forms

Proposition 4.9.1

Let $f \in M_k(\Gamma)$ with $f(z) = \sum_{n \geq 0} a_n q^n$, and define $f' = \frac{1}{2\pi i} \frac{df}{dz}$. Then

$$f'(z) = \sum_{n \geq 1} n a_n q^n \quad \text{and} \quad f'(\gamma z) = (cz + d)^{k+2} f'(z) + \frac{kc}{2\pi i} (cz + d)^{k+1} f(z)$$

for $\gamma \in \Gamma$. In particular f' is *not* a modular form (unless $k = 0$ or $f = 0$).

Proof. Differentiating the q -expansion termwise (a power series in q , locally uniformly convergent):

$$f'(z) = \frac{1}{2\pi i} \frac{d}{dz} \sum_{n \geq 0} a_n e^{2\pi i n z} = \frac{1}{2\pi i} \sum_{n \geq 0} a_n 2\pi i n e^{2\pi i n z} = \sum_{n \geq 1} n a_n q^n.$$

For the transformation law, differentiate $f(\frac{az+b}{cz+d}) = (cz+d)^k f(z)$ with respect to z . Since

$$\frac{d}{dz} \frac{az+b}{cz+d} = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{ad-bc}{(cz+d)^2} = \frac{1}{(cz+d)^2},$$

the chain rule gives (after dividing by $2\pi i$)

$$\begin{aligned} \frac{f'(\gamma z)}{(cz+d)^2} &= \frac{1}{2\pi i} \left(kc(cz+d)^{k-1} f(z) \right) + (cz+d)^k f'(z) \\ \implies f'(\gamma z) &= (cz+d)^{k+2} f'(z) + \frac{kc}{2\pi i} (cz+d)^{k+1} f(z). \end{aligned}$$

■

Remark 4.9.2. Combining this with Theorem 4.7.2 one checks that $f' - \frac{k}{12}E_2f$ is a modular form of weight $k + 2$ (the *Serre derivative*); for example $E'_4 = \frac{1}{3}(E_2E_4 - E_6)$. My handwritten Note 4 ends with the heading “The growth of f ”, which is the topic of the next chapter.

Chapter 5

Growth of Fourier Coefficients and Cusp Forms

In this short chapter we compare the growth of the Fourier coefficients of modular forms of level 1: those of a non-cusp form grow like n^{k-1} , while those of a cusp form are $O(n^{k/2})$ (Hecke's bound). We follow [2, §III.2, Props. 8–9] and [4, Ch. VII, §4.3]. Throughout, $k \geq 4$ is even and $q = e^{2\pi iz}$.

Definition 5.0.1 (Big- O)

For functions f, g on the positive integers, $f(n) = O(g(n))$ as $n \rightarrow \infty$ if there are $M \in \mathbb{R}$ and $n_0 \in \mathbb{Z}$ such that $|f(n)| \leq M|g(n)|$ for all $n \geq n_0$.

Theorem 5.0.2

Let $f(z) = \sum_{n \geq 0} a_n q^n \in M_k(\mathrm{SL}_2(\mathbb{Z}))$.

- (1) If f is not a cusp form, then $a_n = O(n^{k-1})$, and this is sharp: there are constants $c_1, c_2 > 0$ with $c_1 n^{k-1} \leq |a_n| \leq c_2 n^{k-1}$ for all large n .
- (2) (Hecke) If f is a cusp form, then $a_n = O(n^{k/2})$.

Since $M_k = \mathbb{C}E_k \oplus S_k$ for all even $k \geq 4$ (Corollary 4.4.1(4)), for (1) it suffices to check the growth of the coefficients of E_k , and then to compare with (2).

Eisenstein series

Recall that for even $k \geq 4$,

$$E_k(z) = 1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n = \sum_{n \geq 0} a_n q^n, \quad a_n = -\frac{2k}{B_k} \sigma_{k-1}(n) \quad (n \geq 1).$$

Proposition 5.0.3

For all $n \geq 1$: $n^{k-1} \leq \sigma_{k-1}(n) \leq \zeta(k-1) n^{k-1}$. In particular the coefficients of E_k satisfy $|a_n| \asymp n^{k-1}$, i.e. they grow on the order of n^{k-1} .

Proof. Lower bound. Since $n \mid n$, the sum $\sigma_{k-1}(n) = \sum_{d \mid n} d^{k-1}$ contains the term n^{k-1} , and all terms are positive; so $\sigma_{k-1}(n) \geq n^{k-1}$.

Upper bound. If $d \mid n$, then $n = d'd$ for some positive integer d' , and $d' \mid n$; the map $d \mapsto d' = n/d$ is a bijection of the set of positive divisors of n . Thus

$$\sigma_{k-1}(n) = \sum_{d \mid n} d^{k-1} = \sum_{d' \mid n} \left(\frac{n}{d'}\right)^{k-1} = n^{k-1} \sum_{d' \mid n} \frac{1}{(d')^{k-1}} \leq n^{k-1} \sum_{m=1}^{\infty} \frac{1}{m^{k-1}} = n^{k-1} \zeta(k-1),$$

where the series converges since $k \geq 4$ gives $k-1 \geq 3 > 1$.

Hence $n^{k-1} \leq \sigma_{k-1}(n) \leq \zeta(k-1) n^{k-1}$, and since $\zeta(k-1)$ is a constant, $\sigma_{k-1}(n) = O(n^{k-1})$. Therefore $a_n = -\frac{2k}{B_k} \sigma_{k-1}(n)$ satisfies $\left| \frac{2k}{B_k} \right| n^{k-1} \leq |a_n| \leq \left| \frac{2k}{B_k} \right| \zeta(k-1) n^{k-1}$. ■

Cusp forms

Theorem 5.0.4 (Hecke)

If $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$, $f(z) = \sum_{n \geq 1} a_n q^n$, then $a_n = O(n^{k/2})$.

Proof. Step 1: decay at ∞ . As $z \rightarrow i\infty$, $q \rightarrow 0$, and

$$|f(z)| = \left| \sum_{n \geq 1} a_n q^n \right| = |a_1 q + a_2 q^2 + \cdots| = |a_1 + a_2 q + a_3 q^2 + \cdots| |q|.$$

The power series $a_1 + a_2 q + \cdots$ is bounded for $|q| \leq \frac{1}{2}$, say, so $|f(z)| = O(|q|) = O(e^{-2\pi \mathrm{Im} z})$ as $\mathrm{Im} z \rightarrow \infty$.

Step 2: an invariant function. Define $\phi_f(z) = |f(z)| (\mathrm{Im} z)^{k/2}$ for $z \in \mathfrak{H}$. We claim that ϕ_f is continuous and invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$.

(1) ϕ_f is continuous, since f is holomorphic and $\mathrm{Im} z$ is continuous.

(2) ϕ_f is invariant: for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, using $\mathrm{Im}(\gamma z) = \frac{\mathrm{Im} z}{|cz + d|^2}$,

$$\begin{aligned} \phi_f(\gamma z) &= |f(\gamma z)| \mathrm{Im}(\gamma z)^{k/2} = |(cz + d)^k f(z)| \left(\frac{\mathrm{Im} z}{|cz + d|^2} \right)^{k/2} \\ &= |cz + d|^k |f(z)| \frac{(\mathrm{Im} z)^{k/2}}{|cz + d|^k} = |f(z)| (\mathrm{Im} z)^{k/2} = \phi_f(z). \end{aligned}$$

Step 3: ϕ_f is bounded. By Step 1,

$$\lim_{\mathrm{Im} z \rightarrow \infty} \phi_f(z) = \lim_{\mathrm{Im} z \rightarrow \infty} |f(z)| (\mathrm{Im} z)^{k/2} = \lim_{\mathrm{Im} z \rightarrow \infty} O(e^{-2\pi \mathrm{Im} z} (\mathrm{Im} z)^{k/2}) = 0,$$

since the exponential decays faster than any power grows. So there is Y with $\phi_f \leq 1$ on $F \cap \{\mathrm{Im} z \geq Y\}$, and on the compact set $F \cap \{\mathrm{Im} z \leq Y\}$ the continuous function ϕ_f is bounded. Hence ϕ_f is bounded on F . Since every $z \in \mathfrak{H}$ is $\mathrm{SL}_2(\mathbb{Z})$ -equivalent to a point of F (Theorem 3.3.2) and ϕ_f is invariant, $\sup_{\mathfrak{H}} \phi_f = \sup_F \phi_f < \infty$. So there is $M > 0$ with

$$\phi_f(z) = |f(z)| (\mathrm{Im} z)^{k/2} \leq M \implies |f(z)| \leq M (\mathrm{Im} z)^{-k/2} \quad (z \in \mathfrak{H}).$$

Step 4: the coefficients. Fix $y > 0$. Since $|q| < 1$, Cauchy's formula for the coefficients of $\tilde{f}(q) = f(z)$ on the circle $|q| = R = e^{-2\pi y}$ gives

$$a_n = \frac{1}{2\pi i} \oint_{|q|=R} \frac{\tilde{f}(q)}{q^{n+1}} dq.$$

Parametrise the circle by $q = e^{2\pi i(x+iy)}$, $0 \leq x \leq 1$ (as x moves from 0 to 1, $\arg q = 2\pi x$ runs once around, while $|q| = e^{-2\pi y}$ stays constant); then $dq = 2\pi i q dx$, and

$$\begin{aligned} a_n &= \frac{1}{2\pi i} \int_0^1 \frac{f(x+iy)}{q \cdot q^n} 2\pi i q dx = \int_0^1 f(x+iy) q^{-n} dx \\ &= \int_0^1 f(x+iy) e^{-2\pi i(x+iy)n} dx = \int_0^1 f(x+iy) e^{-2\pi i x n} e^{2\pi y n} dx. \end{aligned}$$

Hence, by Step 3,

$$|a_n| \leq \int_0^1 |f(x+iy)| |e^{-2\pi i x n}| e^{2\pi y n} dx \leq \int_0^1 M y^{-k/2} \cdot 1 \cdot e^{2\pi y n} dx = M y^{-k/2} e^{2\pi y n}.$$

This holds for every $y > 0$. Choosing $y = \frac{1}{n}$, we obtain $|a_n| \leq M n^{k/2} e^{2\pi}$, i.e. $a_n = O(n^{k/2})$. ■

Correction. In my handwritten proof the invariance computation used $(\frac{\operatorname{Im} z}{(cz+d)^2})^{k/2}$; the correct factor is $\frac{\operatorname{Im} z}{|cz+d|^2}$ (absolute value), which is what makes the factors $|cz+d|^k$ cancel. Also “ $\max_S \phi_f = \max_F \phi_f$ ” should be read as a statement about suprema; the existence of the supremum is Step 3. In the computation of a_n a factor $\frac{1}{2\pi i}$ appeared once too often in an intermediate line; it cancels against $dq = 2\pi i q dx$.

Proof of Theorem 5.0.2. (2) is Theorem 5.0.4. For (1), write $f = a_0 E_k + g$ with $a_0 \neq 0$ and $g \in S_k$ (Corollary 4.4.1(4)). By Proposition 5.0.3 and Theorem 5.0.4, the n -th coefficient of f is $-a_0 \frac{2k}{B_k} \sigma_{k-1}(n) + O(n^{k/2})$, and $|a_0 \frac{2k}{B_k}| n^{k-1} \leq |a_0 \frac{2k}{B_k} \sigma_{k-1}(n)| \leq |a_0 \frac{2k}{B_k}| \zeta(k-1) n^{k-1}$. Since $k-1 > \frac{k}{2}$ for $k > 2$, the term $O(n^{k/2})$ is eventually smaller than half of the lower bound, which gives constants $c_1, c_2 > 0$ as claimed. ■

Remark 5.0.5. Thus the growth of the coefficients distinguishes cusp forms from non-cusp forms: if $a_n = O(n^{k/2})$, then f is a cusp form. The bound $O(n^{k/2})$ is far from optimal: Deligne proved $a_n = O(n^{(k-1)/2+\varepsilon})$ for every $\varepsilon > 0$ (the Ramanujan–Petersson conjecture), e.g. $|\tau(p)| \leq 2p^{11/2}$ for the coefficients of $\Delta/(2\pi)^{12}$.

Roadmap. On the first page about cusp forms my handwritten notes contain a small diagram of where the project is heading: modular forms of level 1 and their Eisenstein series are related to elliptic curves (Chapters 2–4); Eisenstein series of higher level are related to theta functions, which lead to the four squares theorem (Chapters 6–8).

Chapter 6

Modular Forms of Congruence Subgroups

Notes. In my handwritten notes this chapter (and the next) was titled “Modular Forms of Level N ”. Strictly speaking, “modular forms of level N ” refers to modular forms for the specific congruence subgroups $\Gamma_0(N)$ or $\Gamma_1(N)$. Since here Γ' is an arbitrary congruence subgroup, we speak of *modular forms of congruence subgroups* instead. The phrase “level N ” is kept only where $\Gamma_0(N)$ or $\Gamma_1(N)$ is meant.

In this chapter we extend the definitions of Chapter 4 from $\mathrm{SL}_2(\mathbb{Z})$ to a congruence subgroup Γ' . The new phenomena are that Γ' has several cusps, each with its own Fourier expansion and its own *width*, and that for odd weight a cusp can be *irregular*. We then state the valence formula for Γ' , deduce a dimension bound, and give the analytic characterisations of modular forms and cusp forms by their growth. We follow [2, §III.3] and [1, Ch. 1–3]. Throughout, $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, $q_N = e^{2\pi iz/N}$, and $\mathcal{O}(\mathfrak{H})$ denotes the space of meromorphic functions on $\mathfrak{H}$.

6.1 The slash operator

Fix $k \in \mathbb{Z}$. The group $\mathrm{GL}_2^+(\mathbb{R})$ acts on $\mathfrak{H}$ (Proposition 3.1.4), and for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R})$ we write $j(\gamma, z) = cz + d$ (the *automorphy factor*).

Definition 6.1.1 (Slash operator)

For $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$ and $f \in \mathcal{O}(\mathfrak{H})$ define $f|_k\gamma : \mathfrak{H} \rightarrow \mathbb{C}$ by

$$(f|_k\gamma)(z) = (\det \gamma)^{k/2} j(\gamma, z)^{-k} f(\gamma z).$$

For $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ this is $(cz + d)^{-k} f(\gamma z)$, so “ f is weakly modular of weight k ” means exactly $f|_k\gamma = f$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Proposition 6.1.2

For $\gamma_1, \gamma_2 \in \mathrm{GL}_2^+(\mathbb{R})$:

- (i) $j(\gamma_1\gamma_2, z) = j(\gamma_1, \gamma_2 z) j(\gamma_2, z)$;
- (ii) $f|_k(\gamma_1\gamma_2) = (f|_k\gamma_1)|_k\gamma_2$.

So the slash operator defines a (right) action of $\mathrm{GL}_2^+(\mathbb{R})$ on $\mathcal{O}(\mathfrak{H})$.

Proof. (i) Write $\gamma_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$; the bottom row of $\gamma_1\gamma_2$ is $(c_1a_2 + d_1c_2, c_1b_2 + d_1d_2)$. Then

$$\begin{aligned} j(\gamma_1\gamma_2, z) &= (c_1a_2 + d_1c_2)z + (c_1b_2 + d_1d_2) = c_1a_2z + d_1c_2z + c_1b_2 + d_1d_2 \\ &= (c_1a_2z + c_1b_2 + d_1c_2z + d_1d_2) \cdot \frac{c_2z + d_2}{c_2z + d_2} = \left(\frac{c_1(a_2z + b_2)}{c_2z + d_2} + \frac{d_1(c_2z + d_2)}{c_2z + d_2} \right) (c_2z + d_2) \\ &= \left(c_1 \frac{a_2z + b_2}{c_2z + d_2} + d_1 \right) (c_2z + d_2) = j(\gamma_1, \gamma_2 z) j(\gamma_2, z). \end{aligned}$$

(ii) Using (i), $\det(\gamma_1\gamma_2) = \det \gamma_1 \det \gamma_2$ (both positive, so the real powers split), and $\gamma_1\gamma_2 z = \gamma_1(\gamma_2 z)$:

$$\begin{aligned} (f|_k \gamma_1\gamma_2)(z) &= (\det \gamma_1\gamma_2)^{k/2} j(\gamma_1\gamma_2, z)^{-k} f(\gamma_1\gamma_2 z) \\ &= (\det \gamma_1)^{k/2} (\det \gamma_2)^{k/2} j(\gamma_1, \gamma_2 z)^{-k} j(\gamma_2, z)^{-k} f(\gamma_1(\gamma_2 z)) \\ &= (\det \gamma_2)^{k/2} j(\gamma_2, z)^{-k} \left[(\det \gamma_1)^{k/2} j(\gamma_1, \gamma_2 z)^{-k} f(\gamma_1(\gamma_2 z)) \right] \\ &= (\det \gamma_2)^{k/2} j(\gamma_2, z)^{-k} [(f|_k \gamma_1)(\gamma_2 z)] = ((f|_k \gamma_1)|_k \gamma_2)(z). \quad \blacksquare \end{aligned}$$

Remark 6.1.3. For a scalar matrix λI ($\lambda \in \mathbb{R}^\times$) we have $f|_k(\lambda I) = (\lambda^2)^{k/2} \lambda^{-k} f = \text{sign}(\lambda)^k f$. In particular $f|_k(-I) = (-1)^k f$, and positive scalar matrices act trivially.

Example 6.1.4 (Theta function). Let $\theta(z) = \sum_{n \in \mathbb{Z}} e^{2\pi i z n^2} = \sum_n q^{n^2}$ for $z \in \mathfrak{H}$. We will see in Theorem 7.6.2 that $\theta(z) = \left(\frac{2z}{i}\right)^{-1/2} \theta\left(-\frac{1}{4z}\right)$, i.e. $\theta\left(-\frac{1}{4z}\right) = \left(\frac{2z}{i}\right)^{1/2} \theta(z)$. Squaring, $\theta^2\left(-\frac{1}{4z}\right) = \frac{2z}{i} \theta^2(z)$. For $\gamma = \begin{pmatrix} 0 & -1 \\ 4 & 0 \end{pmatrix} \in \text{GL}_2^+(\mathbb{R})$ (so $\det \gamma = 4$, $j(\gamma, z) = 4z$, $\gamma z = -\frac{1}{4z}$),

$$(\theta^2|_1 \gamma)(z) = 4^{1/2} (4z)^{-1} \theta^2\left(-\frac{1}{4z}\right) = 2 (4z)^{-1} \frac{2z}{i} \theta^2(z) = \frac{1}{2z} \cdot \frac{2z}{i} \theta^2(z) = -i \theta^2(z).$$

6.2 Modular forms for congruence subgroups

Definition 6.2.1

Let $\Gamma' \subseteq \Gamma$ be a congruence subgroup of level N , $k \in \mathbb{Z}$, and $f \in \mathcal{O}(\mathfrak{H})$. We call f a *modular function of weight k for Γ'* if

- (1) $f|_k \gamma = f$ for all $\gamma \in \Gamma'$, and
- (2) for every $\gamma_0 \in \Gamma = \text{SL}_2(\mathbb{Z})$, $(f|_k \gamma_0)(z) = \sum_{n \in \mathbb{Z}} a_n q_N^n$ with only finitely many nonzero a_n with $n < 0$ (“ f is meromorphic at all cusps”).

If moreover f is holomorphic on $\mathfrak{H}$ and, for every $\gamma_0 \in \Gamma$, $a_n = 0$ for all $n < 0$, then f is a *modular form of weight k for Γ'* ; the space of these is $M_k(\Gamma')$. If in addition $a_0 = 0$ for every γ_0 , f is a *cuspidal form*; the space of these is $S_k(\Gamma')$.

Correction. In my handwritten definition of a modular function, the condition “with $a_0 = 0$ ” was attached to (2); it belongs to the definition of a cuspidal form only.

The expansion in (2) always exists:

Proposition 6.2.2 (Existence of the q_N -expansion)

Let f be meromorphic on $\mathfrak{H}$ with $f|_k \gamma = f$ for all $\gamma \in \Gamma'$, where $\Gamma' \supseteq \Gamma(N)$. Then for every

$\gamma_0 \in \Gamma$ the function $g = f|_k \gamma_0$ satisfies $g(z + N) = g(z)$, hence (if g is holomorphic for large $\text{Im } z$) $g(z) = \sum_{n \in \mathbb{Z}} a_n q_N^n$.

Proof. Let $g(z) = (f|_k \gamma_0)(z)$. Since f is invariant under Γ' , for all $\gamma \in \Gamma'$, using Proposition 6.1.2,

$$g|_k(\gamma_0^{-1} \gamma \gamma_0) = ((f|_k \gamma_0)|_k \gamma_0^{-1} \gamma \gamma_0) = f|_k(\gamma_0 \gamma_0^{-1} \gamma \gamma_0) = f|_k \gamma \gamma_0 = (f|_k \gamma)|_k \gamma_0 = f|_k \gamma_0 = g.$$

So g is invariant under $\gamma_0^{-1} \Gamma' \gamma_0$. Since $\Gamma(N) \leq \Gamma$ (Proposition 3.2.4), $\gamma_0 \Gamma(N) \gamma_0^{-1} = \Gamma(N) \subseteq \Gamma'$, hence $\Gamma(N) \subseteq \gamma_0^{-1} \Gamma' \gamma_0$. We also know $T^N = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in \Gamma(N)$, so g is invariant under T^N : $(g|_k T^N)(z) = 1^{k/2} \cdot 1^{-k} g(z + N) = g(z)$, i.e. $g(z + N) = g(z)$. This N -periodicity implies that $g(z) = G(q_N)$ for a function G on the punctured disc, meromorphic there, and so g has a Laurent expansion $g(z) = \sum_n a_n q_N^n$ (valid for $\text{Im } z$ large, where g has no poles). ■

Definition 6.2.3 (Width)

Let $\gamma_0 \in \Gamma$. The set $\{h \in \mathbb{Z} : T^h \in \pm \gamma_0^{-1} \Gamma' \gamma_0\}$ is a subgroup of $\mathbb{Z}$ containing N ; its positive generator h is called the *width* of the cusp $s = \gamma_0 \infty$. It divides N , and $g = f|_k \gamma_0$ satisfies $g(z + h) = \pm g(z)$; if $g(z + h) = g(z)$, then $g(z) = \sum_n a_n q_h^n$ with $q_h = e^{2\pi i z/h} = q_N^{N/h}$, i.e. $\sum_n a_n q_h^n = \sum_n a_n (q_N^{N/h})^n$.

Example 6.2.4 (The cusps of $\Gamma_0(p)$). Let p be a prime; $\Gamma_0(p)$ has the two cusps ∞ and 0 (Problem 3.4.7, Problem 3.4.9).

- *The cusp ∞ :* $\gamma_0 = I$ gives $\infty = \gamma_0 \infty$, and $\gamma_0 T^h \gamma_0^{-1} = I \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} I^{-1} = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \in \Gamma_0(p)$ for every h ; so the smallest h is $h = 1$, and $(f|_k I)(z) = \sum_n a_n e^{2\pi i n z} = \sum a_n q^n$ is the Fourier expansion of f at ∞ .
- *The cusp 0 :* $\gamma_0 = S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ gives $0 = S \infty$. Using $S^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$,

$$\gamma_0 T^h \gamma_0^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & h \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -h & 1 \end{pmatrix}.$$

In $\Gamma_0(p)$ every lower left entry is $\equiv 0 \pmod{p}$, so the smallest h with $-h \equiv 0 \pmod{p}$ is $h = p$. Hence $(f|_k S)(z) = \sum_n a_n e^{2\pi i n z/p} = \sum a_n q_p^n$ is the Fourier expansion of f at 0 .

The same computation shows that for every N the cusp ∞ of $\Gamma_0(N)$ has width 1 and the cusp 0 has width N ; equivalently $S^{-1} \Gamma_0(N) S = \Gamma^0(N) = \{b \equiv 0 \pmod{N}\}$, which contains T^h iff $N \mid h$.

Notes. Why the *smallest* h ? Any h' with $T^{h'} \in \gamma_0^{-1} \Gamma' \gamma_0$ gives an expansion of g in powers of $q_{h'}$, but if $h' = rh$ then $q_h = q_{h'}^r$ and the expansion in $q_{h'}$ only uses exponents divisible by r . Measuring orders of vanishing at a cusp (as in the valence formula) must be done in the local parameter q_h of the smallest width. In my handwritten notes I asked (with “?”) whether $\gamma_0^{-1} \Gamma' \gamma_0$ may contain translations T^h with $h < N$: yes, and then $h \mid N$, since the admissible h form a subgroup $h\mathbb{Z}$ of $\mathbb{Z}$ containing N .

Cusps of Γ' and independence of the representative

Proposition 6.2.5

Every cusp $s \in \mathbb{Q} \cup \{\infty\}$ can be written as $s = \sigma^{-1}\infty$ for some $\sigma \in \Gamma$. Moreover, if $\gamma \in \Gamma$, $\gamma \neq \pm I$, fixes s (a *parabolic* element), then $\sigma\gamma\sigma^{-1} = \pm T^h$ for some $h \neq 0$.

Proof. The first statement is Proposition 3.1.5 (applied to s , giving $\tau \in \Gamma$ with $\tau\infty = s$; put $\sigma = \tau^{-1}$). For the second, $\sigma\gamma\sigma^{-1}(\infty) = \sigma\gamma(s) = \sigma s = \infty$, so $\sigma\gamma\sigma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ fixes ∞ . Then $\frac{a}{c} = \infty$ forces $c = 0$; since $ad - bc = ad = 1$ with $a, d \in \mathbb{Z}$, $a = d = \pm 1$, and $\sigma\gamma\sigma^{-1} = \pm \begin{pmatrix} 1 & \pm b \\ 0 & 1 \end{pmatrix} = \pm T^h$, with $h \neq 0$ since $\gamma \neq \pm I$. ■

Notes. In my handwritten notes this was approached via eigenvalues: a parabolic $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ has $\mathrm{tr}(\gamma)^2 = 4 \det \gamma = 4$, so its characteristic polynomial is $\lambda^2 \mp 2\lambda + 1 = (\lambda \mp 1)^2$, with the double eigenvalue ± 1 ; and the conjugacy to $\pm T^h$ was then explained (with the help of an AI chat) via the transitivity of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R} \cup \{\infty\}$. The direct argument above, using only Proposition 3.1.5 and the stabiliser of ∞ in $\mathrm{SL}_2(\mathbb{Z})$, suffices for our purposes.

Proposition 6.2.6 (★)

Let f be invariant under Γ' (weight k), and let $\gamma_1, \gamma_2 \in \Gamma$ with $\gamma_1\infty = \gamma'\gamma_2\infty$ for some $\gamma' \in \Gamma'$ (i.e. the cusps $\gamma_1\infty$ and $\gamma_2\infty$ are Γ' -equivalent). Then the q_N -expansions of $f|_k\gamma_1$ and $f|_k\gamma_2$ differ only by multiplying each coefficient a_n by a nonzero constant (a root of unity). In particular, the smallest power of q_N occurring is the same.

Proof. Suppose $\gamma_1\infty = \gamma'\gamma_2\infty$ for some $\gamma' \in \Gamma'$. Then $\gamma_1^{-1}\gamma'\gamma_2\infty = \infty$, i.e. ∞ is fixed by $\gamma_1^{-1}\gamma'\gamma_2 \in \Gamma$. Any matrix in $\mathrm{SL}_2(\mathbb{Z})$ fixing ∞ is $\pm \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} = \pm T^h$ (proof of Proposition 6.2.5). Then

$$\gamma_1^{-1}\gamma'\gamma_2 = \pm T^h \implies \gamma_2 = \pm(\gamma')^{-1}\gamma_1 T^h.$$

Now let $g(z) = (f|_k\gamma_1)(z) = \sum_n a_n e^{2\pi i n z/N} = \sum_n a_n q_N^n$ be the expansion at the first representative. For the second representative, by Proposition 6.1.2,

$$f|_k\gamma_2 = f|_k(\pm I \cdot (\gamma')^{-1}\gamma_1 T^h) = \left(\left((f|_k(\pm I))|_k(\gamma')^{-1} \right) |_k\gamma_1 \right) |_k T^h.$$

Since $f|_k(\pm I) = (\pm 1)^k f$ (Remark 6.1.3) and $f|_k(\gamma')^{-1} = f$ (f is invariant under Γ' , and $(\gamma')^{-1} \in \Gamma'$),

$$\begin{aligned} (f|_k\gamma_2)(z) &= (\pm 1)^k ((f|_k\gamma_1)|_k T^h)(z) = (\pm 1)^k (g|_k T^h)(z) = (\pm 1)^k g(z+h) \\ &= (\pm 1)^k \sum_n a_n e^{2\pi i (z+h)n/N} = (\pm 1)^k \sum_n a_n e^{2\pi i h n/N} e^{2\pi i z n/N} = (\pm 1)^k \sum_n (a_n e^{2\pi i h n/N}) q_N^n. \end{aligned}$$

Since $e^{2\pi i h n/N}$ is a root of unity (a nonzero complex constant), the q_N -expansion at γ_2 differs from that at γ_1 by the nonzero factors $(\pm 1)^k e^{2\pi i h n/N}$. ■

Definition 6.2.7 (Behaviour at a cusp)

Let $f \in \mathcal{O}(\mathfrak{H})$ be invariant under Γ' , and let $s \in \mathbb{Q} \cup \{\infty\}$ with $s = \gamma_0\infty$, $\gamma_0 \in \Gamma$. We say that f is

- (1) *meromorphic at the cusp s* if $(f|_k\gamma_0)(z) = \sum_{n \geq -M} a_n q_N^n$ for some M ;
- (2) *holomorphic at the cusp s* if $(f|_k\gamma_0)(z) = \sum_{n \geq 0} a_n q_N^n$;
- (3) *vanishes at the cusp s* if $(f|_k\gamma_0)(z) = \sum_{n \geq 1} a_n q_N^n$.

By Proposition 6.2.6 these properties depend only on the Γ' -equivalence class of s , not on the choice of γ_0 . Thus f is a modular function / modular form / cusp form for Γ' iff it satisfies (1) / (2) / (3) at a set of representatives of the (finitely many, Problem 3.4.9) cusps of Γ' , in addition to being invariant (and holomorphic on $\mathfrak{H}$ in the last two cases).

6.3 Stabilisers of cusps and odd weight

If $-I \in \Gamma'$, there are no nonzero modular functions of odd weight k : since $-I \in \Gamma'$,

$$f(z) = (f|_k(-I))(z) = (-1)^{-k} f\left(\frac{-z}{-1}\right) = -f(z),$$

so $f = 0$. (For $\Gamma' = \mathrm{SL}_2(\mathbb{Z})$ this was observed in Chapter 4.) But if $-I \notin \Gamma'$, odd weights are possible.

Correction. In my handwritten notes, $\Gamma(2)$ was given as an example of a group not containing $-I$. In fact $-I \in \Gamma(2)$, since $-1 \equiv 1 \pmod{2}$. Examples with $-I \notin \Gamma'$ are $\Gamma_1(N)$ and $\Gamma(N)$ for $N \geq 3$ (as $-1 \not\equiv 1 \pmod{N}$).

Proposition 6.3.1 (Stabiliser of ∞)

Let $\Gamma' \leq \Gamma$ be a congruence subgroup and Γ'_∞ the stabiliser of ∞ in Γ' . Then $\Gamma'_\infty \cdot \{\pm I\} = \langle \pm T^h \rangle = \{\pm T^{mh} : m \in \mathbb{Z}\}$ for some $h \geq 1$. More generally, for a cusp $x = \sigma^{-1}\infty$ ($\sigma \in \Gamma$), $\sigma\Gamma'_x\sigma^{-1} \cdot \{\pm I\} = \langle \pm T^h \rangle$, where h is the width of x .

Proof. Let $\gamma \in \Gamma'_\infty$; then $\gamma(\infty) = \frac{a\infty+b}{c\infty+d} = \frac{a}{c} = \infty$, so $c = 0$. Since $ad - bc = 1$, $ad = 1$, so $a = d = 1$ or $a = d = -1$, and $\gamma = \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \pm T^b$, $b \in \mathbb{Z}$. Thus every element of Γ'_∞ is of the form $\pm T^b$. The set $U = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{Z} \right\}$ is a group under multiplication, and $\phi : \mathbb{Z} \rightarrow U$, $b \mapsto T^b$, is an isomorphism $(\mathbb{Z}, +) \cong (U, \cdot)$. Every subgroup of $\mathbb{Z}$ is cyclic, generated by some $h \geq 0$; hence so is every subgroup of U , generated by T^h . The image of $\Gamma'_\infty \cdot \{\pm I\}$ in U (forgetting the sign) is such a subgroup, and it is nontrivial since $T^N \in \Gamma(N) \subseteq \Gamma'$. Hence $\Gamma'_\infty \cdot \{\pm I\} = \langle \pm T^h \rangle$ with $h \geq 1$.

For a cusp $x = \sigma^{-1}\infty$: $\gamma \in \Gamma'_x$ iff $\gamma x = x$ iff $\sigma\gamma\sigma^{-1}\infty = \infty$. So $\sigma\Gamma'_x\sigma^{-1}$ is the stabiliser of ∞ in the congruence subgroup $\sigma\Gamma'\sigma^{-1}$ (which contains $\Gamma(N)$), and the first part applies. ■

Definition 6.3.2 (Regular and irregular cusps)

Let $x = \sigma^{-1}\infty$ be a cusp of Γ' of width h .

- If $-I \in \Gamma'$, then $\sigma\Gamma'_x\sigma^{-1} = \langle -I, T^h \rangle$.
- If $-I \notin \Gamma'$, then exactly one of T^h , $-T^h$ lies in $\sigma\Gamma'_x\sigma^{-1}$: if both did, then $-I = (-T^h)(T^h)^{-1} \in \sigma\Gamma'\sigma^{-1}$, hence $-I \in \Gamma'$. If $\sigma\Gamma'_x\sigma^{-1}$ is generated by T^h , the cusp x is called *regular*; if it is generated by $-T^h$, it is called *irregular*.

Proposition 6.3.3

Let f be invariant under Γ' of odd weight k , $-I \notin \Gamma'$, and let $x = \sigma^{-1}\infty$ be a cusp of width h . Put $g = f|_k\sigma^{-1}$. Then

$$g(z) = \begin{cases} \sum_{n \text{ even}} c_n e^{\pi i n z / h} = \sum_m c_{2m} q_h^m, & x \text{ regular,} \\ \sum_{n \text{ odd}} c_n e^{\pi i n z / h}, & x \text{ irregular.} \end{cases}$$

Proof. First, g is invariant under $\sigma\Gamma'\sigma^{-1}$: for all $\gamma \in \Gamma'$,

$$g|_k(\sigma\gamma\sigma^{-1}) = ((f|_k\sigma^{-1})|_k\sigma\gamma\sigma^{-1}) = f|_k(\sigma^{-1}\sigma\gamma\sigma^{-1}) = f|_k\gamma\sigma^{-1} = (f|_k\gamma)|_k\sigma^{-1} = f|_k\sigma^{-1} = g.$$

- (1) If x is regular, $T^h \in \sigma\Gamma'\sigma^{-1}$, so $g|_k T^h = g$, i.e. $g(z+h) = g(z)$, and $g(z) = \sum_{m \in \mathbb{Z}} \tilde{c}_m e^{2\pi i m z / h} = \sum_{n \text{ even}} c_n e^{\pi i n z / h}$.
- (2) If x is irregular, $-T^h \in \sigma\Gamma'\sigma^{-1}$, so $g|_k(-T^h) = g$, i.e. $(-1)^k g(z+h) = g(z)$, and since k is odd, $g(z+h) = -g(z)$. Then $g(z+2h) = -g(z+h) = g(z)$, so g is $2h$ -periodic and $g(z) = \sum_{n \in \mathbb{Z}} c_n e^{2\pi i n z / 2h} = \sum_{n \text{ odd}} c_n e^{\pi i n z / h}$. In addition, from $g(z+h) = -g(z)$:

$$\sum_n c_n e^{\pi i n (z+h) / h} = - \sum_n c_n e^{\pi i n z / h} \implies \sum_n c_n e^{\pi i n z / h} e^{\pi i n} = - \sum_n c_n e^{\pi i n z / h}.$$

Since $e^{\pi i n} = (e^{\pi i})^n = (-1)^n$, comparing coefficients gives $(-1)^n c_n = -c_n$; if n is even, $c_n = -c_n$, so $c_n = 0$. Hence only odd n occur. ■

6.4 The valence formula and a dimension bound

For $x \in \mathfrak{H}$ let $\bar{\Gamma}'_x$ be the stabiliser of x in $\bar{\Gamma}'$. If x is not Γ -equivalent to i or ρ , then $\bar{\Gamma}'_x$ is trivial (Theorem 3.3.2(3)); if x is equivalent to i or ρ , then $\#\bar{\Gamma}'_x \in \{1, 2\}$ resp. $\{1, 3\}$ (generated by a conjugate of S resp. ST , if any). Points x with $\#\bar{\Gamma}'_x > 1$ are called *elliptic points* of Γ' ; there are only finitely many modulo Γ' . At a cusp s of width h , the order $v_s(f)$ is measured in the local parameter q_h (and in $e^{\pi i z / h}$ at an irregular cusp in odd weight).

Theorem 6.4.1 (Valence formula for Γ')

Let $\Gamma' \subseteq \text{SL}_2(\mathbb{Z})$ be a congruence subgroup and $f \neq 0$ a modular function of weight k for Γ' . Then

$$\sum_{x \in \Gamma' \backslash \mathfrak{H}} \frac{v_x(f)}{\#\bar{\Gamma}'_x} + \sum_{s \in \text{cusps}(\Gamma')} v_s(f) = \frac{k}{12} [\bar{\Gamma} : \bar{\Gamma}'],$$

where $[\bar{\Gamma} : \bar{\Gamma}']$ is the number of cosets of $\bar{\Gamma}'$ in $\bar{\Gamma}$.

Proof sketch. Let $\bar{\Gamma} = \bigsqcup_{i=1}^n \bar{\Gamma}' \alpha_i$ and consider $F = \prod_{i=1}^n f|_k \alpha_i$. Replacing α_i by $\gamma' \alpha_i$ does not change $f|_k \alpha_i$ up to a sign, and right multiplication by $\gamma \in \Gamma$ permutes the cosets; hence F^2 is a modular function of weight $2kn$ for $\text{SL}_2(\mathbb{Z})$. Applying Theorem 4.3.2 to F^2 and computing the orders of F^2 : a point $x \in \mathfrak{H}$ not equivalent to i, ρ has n preimages $\alpha_i^{-1}x$ modulo Γ' , counted with the multiplicity $\#\bar{\Gamma}'_x$ in the elliptic cases; and the cusp ∞ of $\text{SL}_2(\mathbb{Z})$ splits into the cusps of Γ' , a cusp of width h accounting for h of the n cosets. Dividing by 2 gives the formula; see [1, Thm. 3.5.1 and §3.1] for the details. ■

Correction. In my handwritten statement the stabiliser was taken in $\bar{\Gamma}$ (“ $\bar{\Gamma}_x$ the stabilizer of x in $\bar{\Gamma}$ ”); it must be the stabiliser in $\bar{\Gamma}'$. We work in $\bar{\Gamma}$ because γ and $-\gamma$ act in the same way on $\mathfrak{H}$, so I and $-I$ are the same transformation.

Proposition 6.4.2 (1) If $k < 0$, then $M_k(\Gamma') = \{0\}$.

(2) $M_0(\Gamma') = \mathbb{C}$.

Proof. Let $f \in M_k(\Gamma')$, $f \neq 0$. Then f is holomorphic on $\mathfrak{H}$ and at the cusps, so $v_x(f) \geq 0$ for all $x \in \mathfrak{H} \cup \text{cusps}(\Gamma')$, and the left-hand side of the valence formula is nonnegative.

- (1) If $k < 0$, the right-hand side is negative; so no such f exists, i.e. $M_k(\Gamma') = \{0\}$.
- (2) It is clear that $\mathbb{C} \subseteq M_0(\Gamma')$. Now let $f \in M_0(\Gamma')$, and let $c = f(x_1)$ be the value of f at some cusp x_1 (the constant term of its expansion there). Since $c \in M_0(\Gamma')$, $g = f - c \in M_0(\Gamma')$, and $g(x_1) = f(x_1) - c = c - c = 0$ implies that g has a zero at x_1 , i.e. $v_{x_1}(g) \geq 1$. If $g \neq 0$, this contradicts the valence formula (whose right-hand side is 0). So $g = 0$, i.e. $f = c$, and $M_0(\Gamma') \subseteq \mathbb{C}$. ■

Proposition 6.4.3 (Dimension bound)

$\dim M_k(\Gamma') \leq \frac{k}{12} [\bar{\Gamma} : \bar{\Gamma}'] + 1$.

Proof. Let $M = \frac{k}{12} [\bar{\Gamma} : \bar{\Gamma}']$ be the right-hand side of the valence formula, and let $R = \lfloor M \rfloor + 1$ be the unique integer with $M < R \leq M + 1$. Since Γ' has only finitely many elliptic points and cusps modulo Γ' , we can choose R distinct non-elliptic points of $\Gamma' \backslash \mathfrak{H}$:

$$p_1, p_2, \dots, p_R \in \Gamma' \backslash \mathfrak{H}.$$

Suppose, by way of contradiction, that $\dim M_k(\Gamma') \geq R + 1$. Then we can choose $R + 1$ linearly independent modular forms $f_1, \dots, f_{R+1} \in M_k(\Gamma')$; let $V = \text{span}\{f_1, \dots, f_{R+1}\}$, so $f \in V$ means $f = \alpha_1 f_1 + \dots + \alpha_{R+1} f_{R+1}$. Construct the linear map

$$\mathcal{T} : V \rightarrow \mathbb{C}^R, \quad \mathcal{T}(f) = (f(p_1), f(p_2), \dots, f(p_R))^T$$

(choosing a representative of each p_i in $\mathfrak{H}$). By the rank–nullity theorem, $R + 1 = \dim V = \text{rank}(\mathcal{T}) + \text{null}(\mathcal{T}) \leq R + \text{null}(\mathcal{T})$, so $\text{null}(\mathcal{T}) \geq 1$. Thus there is a nonzero $f \in V$ with $\mathcal{T}(f) = 0$, i.e. $f(p_1) = f(p_2) = \dots = f(p_R) = 0$. This implies that f has R distinct zeros at the p_i , so $v_{p_i}(f) \geq 1$. Because $\#\bar{\Gamma}'_{p_i} = 1$ for all non-elliptic points p_i ,

$$\text{LHS} \geq \sum_{i=1}^R \frac{v_{p_i}(f)}{\#\bar{\Gamma}'_{p_i}} \geq \sum_{i=1}^R 1 = R > M = \text{RHS},$$

a contradiction. So $\dim M_k(\Gamma') \leq R \leq M + 1$. ■

6.5 Growth conditions and cusp forms

We would like to characterise holomorphy at the cusps by a polynomial bound.

Lemma 6.5.1

Let $\tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ with $c \neq 0$, and let $h > 0$. There are constants $C_1, C_2 > 0$ such that for all $z = x + iy$ with $|x| \leq h$ and $y \geq 1$:

$$c^2 y^2 \leq |cz + d|^2 \leq C_1 y^2 \quad \text{and} \quad \mathrm{Im}(\tau z) = \frac{y}{|cz + d|^2} \leq \frac{1}{c^2 y} \leq 1.$$

Proof. $|cz + d|^2 = (cx + d)^2 + c^2 y^2 \geq c^2 y^2$, and $(cx + d)^2 \leq (|c|h + |d|)^2 \leq (|c|h + |d|)^2 y^2$ for $y \geq 1$, so we can take $C_1 = (|c|h + |d|)^2 + c^2$. The second claim follows from $\mathrm{Im}(\tau z) = \frac{y}{|cz + d|^2}$ and $c^2 \geq 1$. ■

Proposition 6.5.2

Let f be holomorphic on $\mathfrak{H}$ and invariant under Γ' in weight k . Suppose there are $C, \nu \geq 0$ such that

$$|f(x + iy)| \leq C y^{-\nu} \quad \text{for all } x \in \mathbb{R} \text{ and } 0 < y \leq 1.$$

Then f is holomorphic at all cusps, i.e. $f \in M_k(\Gamma')$. If moreover $\nu < k$, then $f \in S_k(\Gamma')$.

Proof. Let s be a cusp of Γ' . We may assume $s \neq \infty$: indeed $\begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix} \in \Gamma(N) \subseteq \Gamma'$ maps ∞ to $\frac{1}{N}$, so ∞ is Γ' -equivalent to the finite cusp $\frac{1}{N}$, and holomorphy/vanishing only depend on the class (Proposition 6.2.6). Choose $\sigma \in \Gamma$ with $\sigma s = \infty$, and write $\sigma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Claim: $c \neq 0$. Suppose, by way of contradiction, $c = 0$. Then $\sigma^{-1} = \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, so $s = \sigma^{-1}\infty = \infty$, a contradiction.

Let $g = f|_k \sigma^{-1}$; it is invariant under $\sigma\Gamma'\sigma^{-1}$ (proof of Proposition 6.3.3), so $g(z + H) = g(z)$ for $H = h$ or $H = 2h$ (h the width), and $g(z) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n z / H}$. Since g is holomorphic, for every $y > 0$ (changing the contour integral over $|q_H| = e^{-2\pi y / H}$ into an integral over a horizontal segment, as in Theorem 5.0.4),

$$a_n = \frac{1}{H} \int_0^H \frac{g(x + iy)}{e^{2\pi i n(x + iy)/H}} dx.$$

Now estimate g for $|x| \leq H$, $y \geq 1$. By Lemma 6.5.1 (with $\tau = \sigma^{-1}$), $0 < \mathrm{Im}(\sigma^{-1}z) \leq 1$, so by assumption

$$|f(\sigma^{-1}z)| \leq C \mathrm{Im}(\sigma^{-1}z)^{-\nu} = C \left(\frac{|cz + d|^2}{y} \right)^{\nu} \leq C C_1^{\nu} y^{\nu},$$

and $|j(\sigma^{-1}, z)|^{-k} = |cz + d|^{-k} \leq |c|^{-k} y^{-k}$ (for $k \geq 0$; for $k < 0$ use the upper bound instead). Hence

$$|g(z)| = |(\det \sigma^{-1})^{k/2} j(\sigma^{-1}, z)^{-k} f(\sigma^{-1}z)| \leq C' y^{\nu-k} \quad (|x| \leq H, y \geq 1),$$

uniformly in x . Substituting into a_n and using $|e^{-2\pi i n(x + iy)/H}| = e^{2\pi n y / H}$:

$$|a_n| \leq \frac{1}{H} \int_0^H C' y^{\nu-k} e^{2\pi n y / H} dx = C' y^{\nu-k} e^{2\pi n y / H} \quad (y \geq 1).$$

- If $n < 0$, then $e^{2\pi n y / H}$ decays faster than $y^{\nu-k}$ grows as $y \rightarrow \infty$; so $|a_n| \leq 0$, i.e. $a_n = 0$. Hence g has no negative powers: f is holomorphic at s , and $f \in M_k(\Gamma')$.

- If $n = 0$, then $e^{2\pi ny/H} = 1$, so $|a_0| \leq C'y^{\nu-k}$. Since $\nu < k$, $y^{\nu-k} \rightarrow 0$ as $y \rightarrow \infty$, so $a_0 = 0$, i.e. f vanishes at s , and $f \in S_k(\Gamma')$. ■

Correction. In my handwritten statement the hypothesis was “ $f(z) = O(\text{Im}(z)^{-\nu})$ as $\text{Im}(z) \rightarrow \infty$ ”. The proof (and Koblitz’s statement) uses the behaviour as $\text{Im}(z) \rightarrow 0$, uniformly in $\text{Re } z$: the point is that $\sigma^{-1}z$ approaches the real axis when $z \rightarrow i\infty$. Also, the case of the cusp ∞ (where $\sigma = I$ and the argument does not apply directly) is handled by moving to an equivalent finite cusp; my notes simply assumed $x \neq \infty$. Finally, my notes reduced odd k to even k by passing to f^2 ; working with the period $H = 2h$ avoids this.

Theorem 6.5.3 (Characterisation of cusp forms)

Let f be holomorphic on $\mathfrak{H}$ and invariant under Γ' in weight $k > 0$. Then

$$f \in S_k(\Gamma') \iff \phi(z) := |f(z)| \text{Im}(z)^{k/2} \text{ is bounded on } \mathfrak{H}.$$

Proof. ($\Leftarrow$) If $\phi(z)$ is bounded on $\mathfrak{H}$ by some M , then $|f(z)| \leq M \text{Im}(z)^{-k/2}$. Since $\frac{k}{2} < k$, by Proposition 6.5.2 (with $\nu = \frac{k}{2}$), $f \in S_k(\Gamma')$.

($\Rightarrow$) Let $f \in S_k(\Gamma')$. Note that ϕ is invariant under Γ' : for $\gamma \in \Gamma'$,

$$\phi(\gamma z) = |f(\gamma z)| \text{Im}(\gamma z)^{k/2} = |(cz + d)^k f(z)| \left(\frac{\text{Im } z}{|cz + d|^2} \right)^{k/2} = |f(z)| \text{Im}(z)^{k/2} = \phi(z),$$

and ϕ is continuous on $\mathfrak{H}$. More generally, for $\tau \in \text{SL}_2(\mathbb{Z})$ the same computation gives

$$\phi(\tau z) = |(f|_k \tau)(z)| \text{Im}(z)^{k/2},$$

since $|f(\tau z)| = |j(\tau, z)|^k |(f|_k \tau)(z)|$ and $\text{Im}(\tau z)^{k/2} = \text{Im}(z)^{k/2} / |j(\tau, z)|^k$.

Write $\bar{\Gamma} = \bigsqcup_{i=1}^n \alpha_i \bar{\Gamma}'$ and $F' = \bigcup_i \alpha_i^{-1} F$ (Proposition 3.3.3); every $z \in \mathfrak{H}$ is Γ' -equivalent to a point of F' , so $\sup_{\mathfrak{H}} \phi = \max_i \sup_{w \in F} \phi(\alpha_i^{-1} w)$. For each i , $\phi(\alpha_i^{-1} w) = |(f|_k \alpha_i^{-1})(w)| \text{Im}(w)^{k/2}$. Since f vanishes at every cusp, $f|_k \alpha_i^{-1} = \sum_{n \geq 1} a_n e^{2\pi i n w / H}$; for $\text{Im } w \geq Y$,

$$\begin{aligned} |(f|_k \alpha_i^{-1})(w)| &\leq \sum_{n \geq 1} |a_n| e^{-2\pi n \text{Im}(w)/H} \\ &= e^{-2\pi \text{Im}(w)/H} \sum_{m \geq 0} |a_{m+1}| e^{-2\pi m \text{Im}(w)/H} \leq M_i e^{-2\pi \text{Im}(w)/H}, \end{aligned}$$

the last sum being bounded for $\text{Im } w \geq Y$. Hence $\phi(\alpha_i^{-1} w) \leq M_i e^{-2\pi \text{Im}(w)/H} \text{Im}(w)^{k/2}$, which tends to 0 as $\text{Im } w \rightarrow \infty$ since the exponential decays faster than the polynomial grows. On the compact part $F \cap \{\text{Im } w \leq Y\}$ the continuous function $w \mapsto \phi(\alpha_i^{-1} w)$ is bounded. Thus ϕ is bounded on $\mathfrak{H}$. ■

Correction. In my handwritten proof the invariance computation ended with “ $|f(z)| \text{Im}(z)$ ” (the exponent $\frac{k}{2}$ was missing), and the formula for $\phi(\sigma^{-1}z)$ should contain $|(f|_k \sigma^{-1})(z)|$ with absolute value. The boundedness was shown “in a neighbourhood of each cusp”; the decomposition $F' = \bigcup \alpha_i^{-1} F$ makes the passage to all of $\mathfrak{H}$ precise.

Theorem 6.5.4 (Hecke bound for Γ')

If $f \in S_k(\Gamma')$ and $f(z) = \sum_{n \geq 1} a_n q_h^n$ is its expansion at ∞ (h the width of ∞), then $a_n = O(n^{k/2})$. The same holds for the expansion at any cusp.

Proof. By Theorem 6.5.3 there is M with $|f(z)| \leq M \operatorname{Im}(z)^{-k/2}$ on $\mathfrak{H}$. As in Theorem 5.0.4, $a_n = \frac{1}{h} \int_0^h f(x + iy) e^{-2\pi i n(x+iy)/h} dx$ for every $y > 0$, so $|a_n| \leq M y^{-k/2} e^{2\pi n y/h}$, and choosing $y = \frac{1}{n}$ gives $|a_n| \leq M n^{k/2} e^{2\pi/h}$. For a cusp $s = \sigma^{-1}\infty$ we apply this to $f|_k \sigma^{-1}$, which lies in $S_k(\sigma\Gamma'\sigma^{-1})$: it is invariant under $\sigma\Gamma'\sigma^{-1}$ (proof of Proposition 6.3.3), the group $\sigma\Gamma'\sigma^{-1}$ contains $\sigma\Gamma(N)\sigma^{-1} = \Gamma(N)$, and for $\gamma_0 \in \Gamma$, $(f|_k \sigma^{-1})|_k \gamma_0 = f|_k(\sigma^{-1}\gamma_0)$ with $\sigma^{-1}\gamma_0 \in \Gamma$, so it vanishes at all cusps. ■

Notes. My handwritten Note 6 ends with the statement of the Hecke bound and the beginning of the claim $f \in S_k(\Gamma') \Rightarrow f|_k \sigma^{-1} \in S_k(\sigma\Gamma'\sigma^{-1})$; the proofs above complete them. An example in my notes on the cusps of $\Gamma_0(p)$ (deciding which $\frac{u}{v}$ are equivalent to ∞ or 0) was crossed out; it is covered by Problems 3.4.7 and 3.4.9.

Chapter 7

Properties of Modular Forms of Congruence Subgroups

Let Γ' be a congruence subgroup, i.e. $\Gamma(N) \subseteq \Gamma' \subseteq \mathrm{SL}_2(\mathbb{Z})$ for some N . We usually restrict ourselves to the groups $\Gamma_0(N)$ and $\Gamma_1(N)$. In this chapter we compare spaces of modular forms for different groups, introduce Dirichlet characters and the decomposition of $M_k(\Gamma_1(N))$ according to the *Nebentypus*, discuss oldforms and the Atkin–Lehner involution, and construct the first concrete examples: Eisenstein series for $\Gamma_0(p)$ and powers of the Jacobi theta function. We follow [2, §III.3] and [1, §§1.2, 4.2–4.3].

7.1 Changing the group

Proposition 7.1.1

If $\Gamma'' \subseteq \Gamma'$ are congruence subgroups, then $M_k(\Gamma') \subseteq M_k(\Gamma'')$ and $S_k(\Gamma') \subseteq S_k(\Gamma'')$.

Proof. 1. *Modularity.* For all $f \in M_k(\Gamma')$, $f|_k\gamma = f$ for all $\gamma \in \Gamma'$. So, as every $\gamma'' \in \Gamma''$ lies in Γ' , it also holds that $f|_k\gamma'' = f$.

2. *Behaviour at the cusps.* The cusps of Γ' and of Γ'' are the same set $\mathbb{Q} \cup \{\infty\}$, but the equivalence classes might differ: two cusps x, y are Γ' -equivalent if $\gamma x = y$ for some $\gamma \in \Gamma'$, and since Γ'' is smaller, γ might not lie in Γ'' . (My notes also observe that a cusp x fixed by a parabolic $\gamma \in \Gamma'$ is fixed by some power $\gamma^{a-b} \in \Gamma''$: by the pigeonhole principle two of the powers $\gamma, \gamma^2, \dots$ lie in the same coset of the finite index subgroup Γ'' .) Let x be a cusp and $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ with $\sigma x = \infty$. Let h be the width of x for Γ' , so $\sigma\Gamma'_x\sigma^{-1} \cdot \{\pm I\} = \langle \pm T^h \rangle$ (Proposition 6.3.1). Then $\sigma\Gamma''_x\sigma^{-1} \leq \sigma\Gamma'_x\sigma^{-1}$, and as any subgroup of the cyclic group of translations is generated by a multiple of the generator, $\sigma\Gamma''_x\sigma^{-1} \cdot \{\pm I\} = \langle \pm T^{rh} \rangle$ for some $r \geq 1$. Thus, since $f \in M_k(\Gamma')$ has the expansion

$$f|_k\sigma^{-1} = \sum_{n \geq 0} a_n e^{2\pi i n z / h} = \sum_{n \geq 0} a_n e^{2\pi i (nr) z / (rh)} \quad (\text{at the cusp } x \text{ of } \Gamma'),$$

which is the expansion of f at x as a cusp of Γ'' (in the parameter q_{rh} , with exponents $nr \geq 0$, resp. ≥ 1 if $a_0 = 0$), f is holomorphic (resp. vanishes) at all cusps of Γ'' . So $f \in M_k(\Gamma'')$ (resp. $S_k(\Gamma'')$). ■

In particular, since $\Gamma(N) \subseteq \Gamma'$, we have $M_k(\Gamma') \subseteq M_k(\Gamma(N))$.

Proposition 7.1.2

Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and $f \in M_k(\Gamma')$. Then $g = f|_k\alpha$ is invariant under $\alpha^{-1}\Gamma'\alpha$ and holomorphic at all cusps; hence $g \in M_k(\Gamma'')$ for every congruence subgroup $\Gamma'' \subseteq \alpha^{-1}\Gamma'\alpha$ (and S_k is preserved likewise). If $\alpha^{-1}\Gamma'\alpha$ is itself a congruence subgroup, then $\phi_\alpha : M_k(\Gamma') \rightarrow M_k(\alpha^{-1}\Gamma'\alpha)$, $f \mapsto f|_k\alpha$, is an isomorphism with inverse $\phi_{\alpha^{-1}}$.

Proof. 1. *Modularity.* Let $g = f|_k\alpha$; we would like to show $g|_k\gamma' = g$ for $\gamma' \in \alpha^{-1}\Gamma'\alpha$. Any such γ' can be written as $\gamma' = \alpha^{-1}\gamma\alpha$ for some $\gamma \in \Gamma'$. Then

$$g|_k\gamma' = ((f|_k\alpha)|_k\alpha^{-1}\gamma\alpha) = f|_k\gamma\alpha = ((f|_k\gamma)|_k\alpha) = f|_k\alpha = g,$$

since $f|_k\gamma = f$.

2. *Bijectivity.* Note that $(\phi_{\alpha^{-1}} \circ \phi_{\alpha})(f) = \phi_{\alpha^{-1}}(f|_k \alpha) = (f|_k \alpha)|_k \alpha^{-1} = f|_k I = f$. If $\phi_{\alpha}(f_1) = \phi_{\alpha}(f_2)$, then $(\phi_{\alpha^{-1}} \circ \phi_{\alpha})(f_1) = (\phi_{\alpha^{-1}} \circ \phi_{\alpha})(f_2)$, so $f_1 = f_2$ (injective). In a similar way $(\phi_{\alpha} \circ \phi_{\alpha^{-1}})(g) = g$; so for any $g \in M_k(\alpha^{-1}\Gamma'\alpha)$ there is $f = g|_k \alpha^{-1}$ with $\phi_{\alpha}(f) = g$ (surjective; that $f \in M_k(\Gamma')$ follows from parts 1 and 3 applied to α^{-1}).

3. *Holomorphy at the cusps is preserved.* Let y be an arbitrary cusp and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma\infty = y$; we need to check that $g|_k \gamma = f|_k \alpha \gamma$ is holomorphic at ∞ . Since α has rational entries, $x := \alpha y = (\alpha \gamma)(\infty) \in \mathbb{Q} \cup \{\infty\}$ is a cusp. Choose $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ with $\sigma\infty = x$. Since both $\alpha \gamma$ and σ map ∞ to x , the matrix $P = \sigma^{-1} \alpha \gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ fixes ∞ , so $P = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$ with $AD = \det P > 0$; replacing P by $-P$ (which changes $f|_k P$ only by the sign $(-1)^k$) we may assume $A, D > 0$. Then $\alpha \gamma = \sigma P$ and

$$g|_k \gamma = f|_k \alpha \gamma = f|_k \sigma P = (f|_k \sigma)|_k P.$$

Since f is holomorphic at x , $(f|_k \sigma)(z) = \sum_{n \geq 0} a_n e^{2\pi i n z/h}$. As $j(P, z) = 0 \cdot z + D = D$ and $Pz = \frac{Az+B}{D}$,

$$(g|_k \gamma)(z) = (AD)^{k/2} D^{-k} (f|_k \sigma)\left(\frac{Az+B}{D}\right) = (AD)^{k/2} D^{-k} \sum_{n \geq 0} \left(a_n e^{2\pi i n B/(Dh)}\right) e^{2\pi i n (A/D)z/h}.$$

Since $\frac{A}{D} > 0$, each $e^{2\pi i n (A/D)z/h}$ with $n \geq 0$ stays bounded as $z \rightarrow i\infty$ (and tends to 0 for $n \geq 1$): this is an expansion in nonnegative powers of $e^{2\pi i (A/D)z/h}$. So g is holomorphic at y (and vanishes there if $a_0 = 0$). ■

Correction. In my handwritten notes the proposition was stated for $\alpha \in \mathrm{GL}_2^+(\mathbb{R})$. For general real α the group $\alpha^{-1}\Gamma'\alpha$ need not be contained in $\mathrm{SL}_2(\mathbb{Z})$ and α need not map cusps to cusps; one needs $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Also the factor $(\det P)^{k/2}$ was omitted in the computation of $(f|_k \sigma)|_k P$ (it is a harmless nonzero constant), and the step “ $\sigma \gamma = \sigma P$ ” should read $\alpha \gamma = \sigma P$.

Proposition 7.1.3

If $f \in M_k(\Gamma(N))$, then $f(Nz) \in M_k(\Gamma_1(N^2))$.

Proof. Let $\alpha = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{Q})$. Note that

$$\begin{aligned} (f|_k \alpha)(z) &= (\det \alpha)^{k/2} j(\alpha, z)^{-k} f\left(\frac{Nz+0}{0z+1}\right) = N^{k/2} (0z+1)^{-k} f(Nz) = N^{k/2} f(Nz) \\ \implies f(Nz) &= N^{-k/2} (f|_k \alpha)(z). \end{aligned}$$

By Proposition 7.1.2, $f|_k \alpha$ is invariant under $\alpha^{-1}\Gamma(N)\alpha$ and holomorphic at all cusps. To show $f(Nz) \in M_k(\Gamma_1(N^2))$ it suffices to show $\Gamma_1(N^2) \subseteq \alpha^{-1}\Gamma(N)\alpha$, which is equivalent to

$\alpha\Gamma_1(N^2)\alpha^{-1} \subseteq \Gamma(N)$. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_1(N^2)$ be arbitrary. Then $a \equiv d \equiv 1 \pmod{N^2}$ and

$c \equiv 0 \pmod{N^2}$. Conjugating with α (where $\alpha^{-1} = \begin{pmatrix} 1/N & 0 \\ 0 & 1 \end{pmatrix}$):

$$\alpha \gamma \alpha^{-1} = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \frac{1}{N} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Na & Nb \\ c & d \end{pmatrix} \begin{pmatrix} \frac{1}{N} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & Nb \\ c/N & d \end{pmatrix}.$$

Since $c \equiv 0 \pmod{N^2}$, $c/N \in \mathbb{Z}$ and $c/N \equiv 0 \pmod{N}$. Since $a \equiv d \equiv 1 \pmod{N^2}$, also $a \equiv d \equiv 1 \pmod{N}$. Also $Nb \equiv 0 \pmod{N}$. Thus $\alpha \gamma \alpha^{-1} \in \Gamma(N)$. Done. ■

Correction. In my handwritten notes $\alpha = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}$ was said to lie in $\Gamma(N)$; it has determinant N , so $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, which is why Proposition 7.1.2 is needed in that generality.

Notes. Why $f(Nz)$? Because we are going to compare coefficients of expansions in powers of $e^{2\pi iz}$, not of $e^{2\pi iz/N}$. Every $f \in M_k(\Gamma(N))$ has $f(z) = \sum_{n \geq 0} a_n e^{2\pi in z/N}$, and then

$$f(Nz) = \sum_{n \geq 0} a_n e^{2\pi in(Nz)/N} = \sum_{n \geq 0} a_n e^{2\pi in z}$$

has the “same” Fourier coefficients. (Every cusp of $\Gamma(N)$ has width N , because $\Gamma(N) \trianglelefteq \mathrm{SL}_2(\mathbb{Z})$ gives $\Gamma(N)_x = \sigma^{-1} \langle \pm T^N \rangle \sigma \cap \Gamma(N)$; so $z \mapsto Nz$ normalises the expansions at every cusp.) So it suffices to study the spaces $M_k(\Gamma_1(M))$, $M = N^2$.

7.2 Dirichlet characters

Note that $(\mathbb{Z}/N\mathbb{Z})^\times = \{\bar{a} \in \mathbb{Z}/N\mathbb{Z} : \gcd(a, N) = 1\}$ is a finite abelian group under multiplication.

Definition 7.2.1 (Characters)

A *character* of a finite abelian group G is a group homomorphism $\varphi : G \rightarrow \mathbb{C}^\times$ (so $\varphi(ab) = \varphi(a)\varphi(b)$). The *trivial character* maps every element to 1. Since $g^{|G|} = 1$, every value $\varphi(g)$ satisfies $\varphi(g)^{|G|} = \varphi(g^{|G|}) = 1$, i.e. is a root of unity.

Definition 7.2.2 (Dirichlet character)

Let $\varphi : (\mathbb{Z}/N\mathbb{Z})^\times \rightarrow \mathbb{C}^\times$ be a character. The associated *Dirichlet character modulo N* is the function $\chi : \mathbb{Z} \rightarrow \mathbb{C}$,

$$\chi(r) = \begin{cases} \varphi(r \bmod N), & \gcd(r, N) = 1, \\ 0, & \gcd(r, N) > 1. \end{cases}$$

Proposition 7.2.3

A Dirichlet character χ modulo N satisfies:

- (1) $\chi(r + N) = \chi(r)$ for all $r \in \mathbb{Z}$;
- (2) $\chi(rs) = \chi(r)\chi(s)$ for all $r, s \in \mathbb{Z}$;
- (3) $\chi(1) = 1$.

Proof. (1) Note $\gcd(r + N, N) = \gcd(r, N)$ ($r + N$ and N share exactly the common factors of r and N).

- If $\gcd(r + N, N) = 1$, then $\gcd(r, N) = 1$, and $(r + N) \bmod N = r \bmod N$, so $\chi(r + N) = \varphi((r + N) \bmod N) = \varphi(r \bmod N) = \chi(r)$.
- If $\gcd(r + N, N) > 1$, then $\chi(r + N) = 0$; also $\gcd(r, N) > 1$, so $\chi(r) = 0$. So $\chi(r + N) = \chi(r)$.

(2) If $\gcd(rs, N) = 1$, then both $\gcd(r, N) = 1$ and $\gcd(s, N) = 1$. Thus, as φ is a homomorphism,

$$\chi(rs) = \varphi(rs \bmod N) = \varphi((r \bmod N)(s \bmod N)) = \varphi(r \bmod N)\varphi(s \bmod N) = \chi(r)\chi(s).$$

If $\gcd(rs, N) > 1$, then $\chi(rs) = 0$. Also, $\gcd(rs, N) > 1$ means that one of r or s shares a prime factor with N ; WLOG $\gcd(r, N) > 1$, so $\chi(r) = 0$. Thus $\chi(rs) = 0 = 0 \cdot \chi(s) = \chi(r)\chi(s)$.

(3) $\gcd(1, N) = 1$, and $\chi(1) = \varphi(\bar{1}) = 1$ since a homomorphism maps the identity to the identity. ■

The *trivial character* modulo N is $\chi_{\text{triv}}(r) = 1$ if $\gcd(r, N) = 1$ and 0 otherwise.

Definition 7.2.4 (Lifting; conductor)

If $M \mid N$ and χ is a Dirichlet character modulo M , it can be *lifted* to a character χ' modulo N :

$$\chi'(r) = \begin{cases} \chi(r \bmod M), & \gcd(r, N) = 1, \\ 0, & \text{else.} \end{cases}$$

The *conductor* of a Dirichlet character χ modulo N is the smallest positive integer $c \mid N$ such that there is a character $\chi_0 : (\mathbb{Z}/c\mathbb{Z})^\times \rightarrow \mathbb{C}^\times$ with $\chi(r) = \chi_0(r \bmod c)$ for all r with $\gcd(r, N) = 1$, i.e. χ is the lift of a character modulo c . If the conductor is N , χ is called *primitive*.

Correction. In my handwritten notes “ $M \mid N$ ” was explained as “ $M = kN$, so M smaller than N ”; it should be $N = kM$.

Example 7.2.5. (1) The conductor of χ_{triv} modulo N is 1. Note that $c = 1$ divides N . Since $(\mathbb{Z}/1\mathbb{Z})^\times = \{\bar{0}\}$, there is exactly one character $\chi_0 : (\mathbb{Z}/1\mathbb{Z})^\times \rightarrow \mathbb{C}^\times$, with $\chi_0(r \bmod 1) = 1$ for all r . Now, for all r with $\gcd(r, N) = 1$, by definition of χ_{triv} , $\chi_{\text{triv}}(r) = 1 = \chi_0(r \bmod 1)$.

(2) There are precisely two Dirichlet characters modulo 4, since $(\mathbb{Z}/4\mathbb{Z})^\times = \{\bar{1}, \bar{3}\}$ is cyclic of order 2: the trivial one, and

$$\chi_{-4}(r) = \left(\frac{-1}{r}\right) = \begin{cases} 1, & r \equiv 1 \pmod{4}, \\ -1, & r \equiv 3 \pmod{4}, \\ 0, & r \text{ even.} \end{cases}$$

The character χ_{-4} is primitive (conductor 4), since the only characters modulo 1 and 2 are trivial, while $\chi_{-4}(3) = -1$. It is *odd*: $\chi_{-4}(-1) = -1$.

Proposition 7.2.6

Let χ be a Dirichlet character modulo N . Then $\tilde{\chi} : \Gamma_0(N) \rightarrow \mathbb{C}^\times$, $\tilde{\chi} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \chi(d)$, is a group character (homomorphism) of $\Gamma_0(N)$, trivial on $\Gamma_1(N)$.

Proof. We need $\tilde{\chi}(\gamma_1\gamma_2) = \tilde{\chi}(\gamma_1)\tilde{\chi}(\gamma_2)$ for all $\gamma_1, \gamma_2 \in \Gamma_0(N)$. Let $\gamma_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \in \Gamma_0(N)$. Thus $a_id_i - b_ic_i = 1$, and since $c_i \equiv 0 \pmod{N}$, $a_id_i \equiv 1 \pmod{N}$, so $\gcd(d_i, N) = 1$. Thus $\chi(d_1), \chi(d_2)$ are both nonzero (roots of unity), so $\tilde{\chi}$ takes values in $\mathbb{C}^\times$. Furthermore

$$\gamma_1\gamma_2 = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix},$$

and since $c_1 \equiv 0 \pmod{N}$ we observe that the lower right entry satisfies $c_1b_2 + d_1d_2 \equiv d_1d_2 \pmod{N}$. Because χ is a Dirichlet character modulo N (periodic and multiplicative),

$$\tilde{\chi}(\gamma_1\gamma_2) = \chi(c_1b_2 + d_1d_2) = \chi(d_1d_2) = \chi(d_1)\chi(d_2) = \tilde{\chi}(\gamma_1)\tilde{\chi}(\gamma_2).$$

For $\gamma \in \Gamma_1(N)$, $d \equiv 1$, so $\tilde{\chi}(\gamma) = \chi(1) = 1$. ■

7.3 Decomposition by the Nebentypus

Definition 7.3.1 (Nebentypus)

For a Dirichlet character χ modulo N , let

$$M_k(N, \chi) = \left\{ f \in M_k(\Gamma_1(N)) \mid f|_k \gamma = \chi(d)f \text{ for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \right\},$$

the space of *modular forms of level N and Nebentypus (character) χ* : the subspace on which $\Gamma_0(N)$ acts by the character χ . Define $S_k(N, \chi) = M_k(N, \chi) \cap S_k(\Gamma_1(N))$.

Lemma 7.3.2

$\Gamma_0(N)$ acts on $M_k(\Gamma_1(N))$ by $\gamma \cdot f = f|_k \gamma$ (a right action), $\Gamma_1(N)$ acts trivially, and hence $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ acts: for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ put $\langle d \rangle f = f|_k \gamma_d$, where $\gamma_d \in \Gamma_0(N)$ is any matrix with lower right entry $\equiv d \pmod{N}$.

Proof. Well defined. Let $f \in M_k(\Gamma_1(N))$ and $\gamma \in \Gamma_0(N)$. For any $\gamma_1 \in \Gamma_1(N)$, since $\Gamma_1(N) \trianglelefteq \Gamma_0(N)$ (Proposition 3.2.4), $\gamma\gamma_1\gamma^{-1} = \gamma_2 \in \Gamma_1(N)$. Thus, applying $|_k \gamma_1$ to $f|_k \gamma$,

$$(f|_k \gamma)|_k \gamma_1 = f|_k \gamma \gamma_1 = f|_k (\gamma \gamma_1 \gamma^{-1} \gamma) = f|_k \gamma_2 \gamma = (f|_k \gamma_2)|_k \gamma = f|_k \gamma,$$

since $f|_k \gamma_2 = f$. Hence $f|_k \gamma$ is invariant under $\Gamma_1(N)$; it is holomorphic at the cusps by Proposition 7.1.2 (with $\alpha = \gamma$). So $f|_k \gamma \in M_k(\Gamma_1(N))$, and Proposition 6.1.2 shows that this is an action. Next, $\Gamma_1(N)$ acts trivially, by $f|_k \gamma = f$. Because $\Gamma_1(N) \trianglelefteq \Gamma_0(N)$ and $\Gamma_1(N)$ acts trivially, the quotient $\Gamma_0(N)/\Gamma_1(N)$ acts as well; and $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ via $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto d \pmod{N}$ (proof of Proposition 3.2.10(3)). Concretely, if $\gamma, \gamma' \in \Gamma_0(N)$ have the same $d \pmod{N}$, then $\gamma'\gamma^{-1} \in \Gamma_1(N)$ and $f|_k \gamma' = (f|_k \gamma'\gamma^{-1})|_k \gamma = f|_k \gamma$. Since $(\mathbb{Z}/N\mathbb{Z})^\times$ is abelian, the order of composition does not matter. ■

Theorem 7.3.3 (Decomposition)

$$M_k(\Gamma_1(N)) = \bigoplus_{\chi} M_k(N, \chi), \quad S_k(\Gamma_1(N)) = \bigoplus_{\chi} S_k(N, \chi),$$

the sums running over all Dirichlet characters χ modulo N .

Proof. We use the following facts about the characters of the finite abelian group $G = (\mathbb{Z}/N\mathbb{Z})^\times$, of order $\varphi(N)$ [4, Ch. VI, §1]: there are exactly $\varphi(N)$ characters, and

$$(O1) \sum_{d \in G} \chi(d) \overline{\psi(d)} = \begin{cases} \varphi(N), & \chi = \psi, \\ 0, & \chi \neq \psi, \end{cases} \quad (O2) \sum_{\chi} \chi(d) = \begin{cases} \varphi(N), & d = 1, \\ 0, & d \neq 1. \end{cases}$$

For each χ define the linear operator on $M_k(\Gamma_1(N))$

$$e_{\chi} f = \frac{1}{\varphi(N)} \sum_{d \in G} \overline{\chi(d)} \langle d \rangle f.$$

$e_{\chi} f \in M_k(N, \chi)$: for $\gamma \in \Gamma_0(N)$ with lower right entry d_0 , using that $\langle \cdot \rangle$ is an action of the abelian group G ,

$$(e_{\chi} f)|_k \gamma = \frac{1}{\varphi(N)} \sum_d \overline{\chi(d)} \langle dd_0 \rangle f = \frac{1}{\varphi(N)} \sum_{d'} \overline{\chi(d' d_0^{-1})} \langle d' \rangle f = \chi(d_0) e_{\chi} f,$$

substituting $d' = dd_0$ and using $\overline{\chi(d_0^{-1})} = \chi(d_0)$ ($\chi(d_0)$ is a root of unity).

$\sum_{\chi} e_{\chi} f = f$: by (O2), $\sum_{\chi} e_{\chi} f = \frac{1}{\varphi(N)} \sum_d \left(\sum_{\chi} \overline{\chi(d)} \right) \langle d \rangle f = \frac{1}{\varphi(N)} \varphi(N) \langle 1 \rangle f = f$ (note $\overline{\chi(d)} = \chi(d^{-1})$ and $d^{-1} = 1$ iff $d = 1$).

The sum is direct: if $f \in M_k(N, \psi)$, then $\langle d \rangle f = \psi(d)f$, so by (O1) $e_{\chi} f = \frac{1}{\varphi(N)} \sum_d \overline{\chi(d)} \psi(d) f$ equals f if $\chi = \psi$ and 0 otherwise. Hence if $\sum_{\psi} f_{\psi} = 0$ with $f_{\psi} \in M_k(N, \psi)$, applying e_{χ} gives $f_{\chi} = 0$.

The same argument applies to $S_k(\Gamma_1(N))$, since $\langle d \rangle$ preserves cusp forms (Proposition 7.1.2). ■

Notes. In my handwritten notes the proof is “KOBLITZ” together with a diagram: the action of $\Gamma_0(N)$ on $M_k(\Gamma_1(N))$, the trivial action of $\Gamma_1(N)$, the induced action of $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^{\times}$, and the fact that a finite-dimensional representation (here $M_k(\Gamma_1(N))$ is finite dimensional by Proposition 6.4.3) of a finite abelian group decomposes into one-dimensional irreducible pieces, i.e. into eigenspaces for the characters. The projectors e_{χ} above make this decomposition explicit.

Proposition 7.3.4 (1) $M_k(\Gamma_0(N)) = M_k(N, \chi_{\text{triv}})$.

(2) Every $f \in M_k(N, \chi)$ has a Fourier expansion at ∞ of the form $f(z) = \sum_{n \geq 0} a_n q^n$, $q = e^{2\pi iz}$.

(3) If $\chi(-1) \neq (-1)^k$, then $M_k(N, \chi) = \{0\}$.

Proof. (1) By definition, $M_k(N, \chi_{\text{triv}}) = \{f \in M_k(\Gamma_1(N)) : f|_k \gamma = \chi_{\text{triv}}(d)f \ \forall \gamma \in \Gamma_0(N)\}$.

“ $\subseteq$ ”: Let $f \in M_k(\Gamma_0(N))$; since $\Gamma_1(N) \subseteq \Gamma_0(N)$, $M_k(\Gamma_0(N)) \subseteq M_k(\Gamma_1(N))$ (Proposition 7.1.1), so $f \in M_k(\Gamma_1(N))$. For all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ we have $ad - bc = 1 \Rightarrow ad \equiv 1 \pmod{N} \Rightarrow \gcd(d, N) = 1$, so $\chi_{\text{triv}}(d) = 1$; and since $f \in M_k(\Gamma_0(N))$, $f|_k \gamma = f = 1 \cdot f = \chi_{\text{triv}}(d)f$. Therefore $f \in M_k(N, \chi_{\text{triv}})$.

“ $\supseteq$ ”: Let $f \in M_k(N, \chi_{\text{triv}})$; then $f \in M_k(\Gamma_1(N))$ and $f|_k \gamma = \chi_{\text{triv}}(d)f$ for all $\gamma \in \Gamma_0(N)$. Again $\chi_{\text{triv}}(d) = 1$, so $f|_k \gamma = f$, and $f \in M_k(\Gamma_0(N))$.

(2) Since $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_1(N)$, the stabiliser of ∞ in $\Gamma_1(N)$ is the same as in $\text{SL}_2(\mathbb{Z})$ up to $\pm I$, the width of ∞ is 1, and $f(z+1) = f(z)$. Thus $f(z) = \sum_{n \geq 0} a_n e^{2\pi i n z}$.

(3) Note that $-I \in \Gamma_0(N)$ (with $d = -1$). So for all $f \in M_k(N, \chi)$,

$$(f|_k(-I))(z) = \chi(-1)f(z) \implies (-1)^k f(z) = \chi(-1)f(z) \implies ((-1)^k - \chi(-1))f(z) = 0.$$

If $\chi(-1) \neq (-1)^k$, the only solution is $f = 0$. (So if k is odd and $\chi(-1) = 1$ (χ even), or k is even and $\chi(-1) = -1$ (χ odd), then $M_k(N, \chi) = \{0\}$.) ■

7.4 Oldforms and the Atkin–Lehner involution

Oldforms are modular forms of higher level that come from lower levels (divisors of N). There are two ways to raise the level.

Proposition 7.4.1 (Method I: inclusion)

If $N \mid M$, then $M_k(\Gamma_0(N)) \subseteq M_k(\Gamma_0(M))$.

Proof. If $N \mid M$, then $M = kN$ for some integer k . Thus for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(M)$: $c \equiv 0 \pmod{M}$, i.e. $c = lM = (lk)N$, so $c \equiv 0 \pmod{N}$. So $\Gamma_0(M) \subseteq \Gamma_0(N)$, and by Proposition 7.1.1, $M_k(\Gamma_0(N)) \subseteq M_k(\Gamma_0(M))$: every modular form of level N is of level M if $N \mid M$. ■

Proposition 7.4.2 (Method II: the shift operator)

If $f \in M_k(\Gamma_0(N))$ and $\ell \geq 1$ is an integer, then $f(\ell z) \in M_k(\Gamma_0(N\ell))$. The map $V_\ell : M_k(\Gamma_0(N)) \rightarrow M_k(\Gamma_0(N\ell))$, $f \mapsto f(\ell z)$, is called the *shift operator*.

Proof. Let $f \in M_k(\Gamma_0(N))$ and define $g(z) = f(\ell z)$. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N\ell)$. Then $c \equiv 0 \pmod{N\ell}$, so $c = kN\ell$, $c/\ell = kN$, and $c/\ell \equiv 0 \pmod{N}$. Note that

$$\begin{aligned} (g|_k \gamma)(z) &= j(\gamma, z)^{-k} g(\gamma z) = j(\gamma, z)^{-k} f(\ell(\gamma z)) = j(\gamma, z)^{-k} f\left(\ell \frac{az + b}{cz + d}\right) \\ &= j(\gamma, z)^{-k} f\left(\frac{a(\ell z) + b\ell}{(c/\ell)(\ell z) + d}\right). \end{aligned}$$

Let $z' = \ell z$ and $\gamma' = \begin{pmatrix} a & b\ell \\ c/\ell & d \end{pmatrix}$; this lies in $\Gamma_0(N)$, since $ad - (b\ell)(c/\ell) = ad - bc = 1$ and $c/\ell \equiv 0 \pmod{N}$. Moreover $j(\gamma', z') = (c/\ell)(\ell z) + d = cz + d = j(\gamma, z)$. We can rewrite

$$(g|_k \gamma)(z) = j(\gamma', z')^{-k} f(\gamma' z') = (f|_k \gamma')(z') = f(z') = f(\ell z) = g(z),$$

as $\gamma' \in \Gamma_0(N)$ already. Holomorphy at the cusps follows from Proposition 7.1.2, since $g = \ell^{-k/2} f|_k \begin{pmatrix} \ell & 0 \\ 0 & 1 \end{pmatrix}$. So $f(\ell z) \in M_k(\Gamma_0(N\ell))$. ■

Report Problem 7.4.3 (Atkin–Lehner involution)

Let $f \in M_k(N, \chi)$ and $w_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$. Show that w_N normalises $\Gamma_0(N)$ and that $f|_k w_N \in M_k(N, \bar{\chi})$.

Solution. w_N normalises $\Gamma_0(N)$. We have $w_N^{-1} = \frac{1}{N} \begin{pmatrix} 0 & 1 \\ -N & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{N} \\ -1 & 0 \end{pmatrix}$. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, write $c = Nc'$. Then

$$w_N \gamma w_N^{-1} = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{N} \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -c & -d \\ Na & Nb \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{N} \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} d & -c' \\ -Nb & a \end{pmatrix}.$$

This has integer entries, determinant $ad - bc = 1$, and lower left entry $-Nb \equiv 0 \pmod{N}$; so $w_N \gamma w_N^{-1} \in \Gamma_0(N)$, with lower right entry a . In the same way $w_N^{-1} \gamma w_N \in \Gamma_0(N)$. Hence $w_N \Gamma_0(N) w_N^{-1} = \Gamma_0(N)$. Note that the lower right entry of $w_N \gamma w_N^{-1}$ is $a \equiv d^{-1} \pmod{N}$.

$f|_k w_N \in M_k(N, \bar{\chi})$. Let $g = f|_k w_N$ and $\gamma \in \Gamma_0(N)$ as above. Then $\gamma = w_N^{-1} \gamma' w_N$ with $\gamma' = w_N \gamma w_N^{-1}$, whose lower right entry is a , and

$$g|_k \gamma = f|_k w_N \gamma = f|_k \gamma' w_N = (f|_k \gamma')|_k w_N = \chi(a) f|_k w_N = \chi(a) g.$$

Since $ad \equiv 1 \pmod{N}$, $\chi(a)\chi(d) = \chi(ad) = \chi(1) = 1$, so $\chi(a) = \chi(d)^{-1} = \overline{\chi(d)}$ ($\chi(d)$ is a root of unity). Hence $g|_k \gamma = \overline{\chi(d)}g$ for all $\gamma \in \Gamma_0(N)$; in particular g is invariant under $\Gamma_1(N)$. Since $w_N \in \mathrm{GL}_2^+(\mathbb{Q})$, g is holomorphic at all cusps by Proposition 7.1.2. Therefore $f|_k w_N \in M_k(N, \overline{\chi})$.

Finally, $w_N^2 = \begin{pmatrix} -N & 0 \\ 0 & -N \end{pmatrix} = -NI$, so $(f|_k w_N)|_k w_N = f|_k(-NI) = (-1)^k f$ (Remark 6.1.3); thus $|_k w_N$ is an involution up to sign. ■

7.5 Eisenstein series for $\Gamma_0(N)$

Definition 7.5.1

For even $k > 2$ define

$$E_{k,N}(z) = \frac{1}{2} \sum_{\substack{\gcd(c,d)=1 \\ c \equiv 0 \pmod{N}}} \frac{1}{(cz+d)^k}, \quad G_{k,N}(z) = \sum_{\substack{(c,d) \neq (0,0) \\ c \equiv 0 \pmod{N}}} \frac{1}{(cz+d)^k}.$$

The factor $\frac{1}{2}$ accounts for the pairs (c, d) and $(-c, -d)$, which both appear since $\gcd(c, d) = \gcd(-c, -d) = 1$ ($-I \in \Gamma_0(N)$). Both series converge absolutely and uniformly on compact subsets of $\mathfrak{H}$ (they are subseries of G_k , Proposition 4.2.3), so they are holomorphic.

Proposition 7.5.2

Let p be a prime and $k \geq 4$ even. Then

$$G_{k,p}(z) = 2\left(1 - \frac{1}{p^k}\right)\zeta(k) E_{k,p}(z) + \frac{1}{p^k} G_k(z) \quad \text{and} \quad G_{k,p}(z) = G_k(pz),$$

and consequently

$$E_{k,p}(z) = \left(1 - \frac{1}{p^k}\right)^{-1} \left(E_k(pz) - \frac{1}{p^k} E_k(z)\right) = 1 - \frac{2k}{B_k} \left(1 - \frac{1}{p^k}\right)^{-1} \sum_{n \geq 1} \left(\sigma_{k-1}\left(\frac{n}{p}\right) - \frac{1}{p^k} \sigma_{k-1}(n)\right) q^n,$$

where $\sigma_{k-1}(n/p) = 0$ if $p \nmid n$. In particular $E_{k,p} \in M_k(\Gamma_0(p))$.

Proof. First formula. Split the pairs $(c, d) \neq (0, 0)$ with $c \equiv 0 \pmod{p}$ according to whether $p \nmid d$ or $p \mid d$:

$$G_{k,p}(z) = \underbrace{\sum_{\substack{(c,d) \neq (0,0), \\ \gcd(d,p)=1 \\ c \equiv 0 \pmod{p}}} \frac{1}{(cz+d)^k}}_{(1)} + \underbrace{\sum_{\substack{(c,d) \neq (0,0) \\ c \equiv 0 \pmod{p}, d \equiv 0 \pmod{p}}} \frac{1}{(cz+d)^k}}_{(2)}.$$

We partition the sum over all pairs (c, d) by extracting their greatest common divisor $n = \gcd(c, d) \geq 1$: $c = nc'$, $d = nd'$ with $\gcd(c', d') = 1$. Since p is prime, the condition $c = nc' \equiv 0 \pmod{p}$ has two cases, $p \nmid n$ or $p \mid n$.

(1) If $p \nmid n$, i.e. $\gcd(n, p) = 1$, then $p \mid c'$; and then $p \nmid d'$ (as $\gcd(c', d') = 1$), so $p \nmid d$. Conversely

$p \nmid d$ forces $p \nmid n$. So part (1) is

$$\begin{aligned} \sum_{\substack{n \geq 1 \\ \gcd(n,p)=1}} \sum_{\substack{\gcd(c',d')=1 \\ c' \equiv 0 \pmod{p}}} \frac{1}{(nc'z + nd')^k} &= \left(\sum_{\substack{n \geq 1 \\ \gcd(n,p)=1}} \frac{1}{n^k} \right) \left(\sum_{\substack{\gcd(c',d')=1 \\ c' \equiv 0 \pmod{p}}} \frac{1}{(c'z + d')^k} \right) \\ &= \left(\sum_{\substack{n \geq 1 \\ \gcd(n,p)=1}} \frac{1}{n^k} \right) 2E_{k,p}(z). \end{aligned}$$

Note that $\sum_{n \geq 1} \frac{1}{n^k} = \zeta(k)$ and $\sum_{m \geq 1} \frac{1}{(mp)^k} = \frac{1}{p^k} \sum_m \frac{1}{m^k} = \frac{1}{p^k} \zeta(k)$, so $\sum_{\gcd(n,p)=1} \frac{1}{n^k} = \zeta(k) - \frac{1}{p^k} \zeta(k) = (1 - \frac{1}{p^k}) \zeta(k)$. Therefore part (1) equals $(1 - \frac{1}{p^k}) \zeta(k) 2E_{k,p}(z)$.

(2) If $p \mid c$ and $p \mid d$, then $c = pC$ and $d = pD$ with $(C, D) \neq (0, 0)$ arbitrary. So

$$\sum_{\substack{(c,d) \neq (0,0) \\ c \equiv d \equiv 0 \pmod{p}}} \frac{1}{(cz + d)^k} = \frac{1}{p^k} \sum_{(C,D) \neq (0,0)} \frac{1}{(Cz + D)^k} = \frac{1}{p^k} G_k(z).$$

Therefore $G_{k,p}(z) = 2(1 - \frac{1}{p^k}) \zeta(k) E_{k,p}(z) + \frac{1}{p^k} G_k(z)$.

Second formula. On the other hand, splitting off $c = 0$ and writing $c = c'p$,

$$G_{k,p}(z) = \sum_{d \neq 0} \frac{1}{d^k} + \sum_{c' \neq 0} \sum_{d \in \mathbb{Z}} \frac{1}{(c'pz + d)^k} = \sum_{(c',d) \neq (0,0)} \frac{1}{(c'(pz) + d)^k} = G_k(pz).$$

Conclusion. Substituting $G_k(z) = 2\zeta(k)E_k(z)$ and $G_k(pz) = 2\zeta(k)E_k(pz)$ (Theorem 4.2.8) and dividing by $2\zeta(k)$:

$$E_k(pz) = \left(1 - \frac{1}{p^k}\right) E_{k,p}(z) + \frac{1}{p^k} E_k(z) \implies E_{k,p}(z) = \left(1 - \frac{1}{p^k}\right)^{-1} \left(E_k(pz) - \frac{1}{p^k} E_k(z)\right).$$

Since $E_k(pz) = 1 - \frac{2k}{B_k} \sum_{m \geq 1} \sigma_{k-1}(m) q^{pm} = 1 - \frac{2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n/p) q^n$, we get $E_k(pz) - \frac{1}{p^k} E_k(z) = (1 - \frac{1}{p^k}) - \frac{2k}{B_k} \sum_n (\sigma_{k-1}(n/p) - \frac{1}{p^k} \sigma_{k-1}(n)) q^n$, which gives the q -expansion. Finally $E_k(z) \in M_k(\Gamma_0(1)) \subseteq M_k(\Gamma_0(p))$ and $E_k(pz) \in M_k(\Gamma_0(p))$ by Proposition 7.4.2, so $E_{k,p} \in M_k(\Gamma_0(p))$. ■

Correction. In my handwritten notes the first formula was written with the factor $2(1 - p^k)$; it should be $2(1 - p^{-k})$, as in the derivation.

Proposition 7.5.3

For a prime p , $E_{2,p}(z) := E_2(z) - p E_2(pz) \in M_2(\Gamma_0(p))$.

Proof. It is holomorphic on $\mathfrak{H}$, and its q -expansion $1 - p - 24 \sum_n (\sigma_1(n) - p\sigma_1(n/p)) q^n$ shows that it is holomorphic at ∞ . At the cusp 0, using $E_2(-1/z) = z^2 E_2(z) + \frac{6z}{\pi i}$ for z and for z/p (note $-\frac{p}{z} = -\frac{1}{z/p}$),

$$(E_{2,p}|_2 S)(z) = z^{-2} \left[z^2 E_2(z) + \frac{6z}{\pi i} - p \left(\frac{z^2}{p^2} E_2\left(\frac{z}{p}\right) + \frac{6z/p}{\pi i} \right) \right] = E_2(z) - \frac{1}{p} E_2\left(\frac{z}{p}\right),$$

an expansion in nonnegative powers of $q_p = e^{2\pi i z/p}$; so $E_{2,p}$ is holomorphic at 0 (and $\Gamma_0(p)$ has only the cusps $\infty, 0$). To show $E_{2,p} \in M_2(\Gamma_0(p))$ it therefore suffices to show the invariance under $\Gamma_0(p)$. We know (Theorem 4.7.2, with $E_2 = \frac{3}{\pi^2} G_2$) that for all $\gamma \in \text{SL}_2(\mathbb{Z})$, in particular $\gamma \in \Gamma_0(p)$,

$$E_2(\gamma z) = (cz + d)^2 E_2(z) + \frac{12c(cz + d)}{2\pi i}. \quad (\star)$$

Also, $p(\gamma z) = p \frac{az+b}{cz+d} = \frac{a(pz)+bp}{(c/p)(pz)+d}$. Let $z' = pz$ and $\gamma' = \begin{pmatrix} a & bp \\ c/p & d \end{pmatrix}$; then $\gamma' \in \mathrm{SL}_2(\mathbb{Z})$, since $c \equiv 0 \pmod{p}$ gives $c/p \in \mathbb{Z}$, and $ad - bp(c/p) = ad - bc = 1$; and $p(\gamma z) = \gamma' z'$. Thus, by $(\star)$ applied to γ' and z' ,

$$E_2(\gamma' z') = \left(\frac{c}{p} z' + d\right)^2 E_2(z') + \frac{12(c/p)(\frac{c}{p} z' + d)}{2\pi i} = (cz + d)^2 E_2(pz) + \frac{12(c/p)(cz + d)}{2\pi i},$$

substituting back $z' = pz$. Multiplying both sides by p :

$$pE_2(p\gamma z) = p(cz + d)^2 E_2(pz) + \frac{12c(cz + d)}{2\pi i}.$$

Hence

$$\begin{aligned} E_{2,p}(\gamma z) &= E_2(\gamma z) - pE_2(p\gamma z) \\ &= \left((cz + d)^2 E_2(z) + \frac{12c(cz + d)}{2\pi i}\right) - \left(p(cz + d)^2 E_2(pz) + \frac{12c(cz + d)}{2\pi i}\right) \\ &= (cz + d)^2 (E_2(z) - pE_2(pz)) = (cz + d)^2 E_{2,p}(z). \end{aligned}$$

Therefore $E_{2,p}(z) \in M_2(\Gamma_0(p))$. ■

7.6 Jacobi's theta function

Definition 7.6.1

$$\theta(z) = \sum_{n=-\infty}^{\infty} q^{n^2} = \sum_{n \in \mathbb{Z}} e^{2\pi i n^2 z}, \quad z \in \mathfrak{H}.$$

This series converges absolutely and uniformly on compact subsets of $\mathfrak{H}$, since there $|q| \leq r < 1$ and $\sum_n r^{n^2} < \infty$; so it defines a holomorphic function on $\mathfrak{H}$. Notice that

$$\theta(Tz) = \theta(z + 1) = \sum_n e^{2\pi i n^2 (z+1)} = \sum_n e^{2\pi i n^2 z} \underbrace{e^{2\pi i n^2}}_{=1} = \theta(z),$$

and as $z \rightarrow i\infty$, $\theta(z) \rightarrow 1$ (only $n = 0$ survives).

Theorem 7.6.2

For $z \in \mathfrak{H}$, with the principal branch of the square root,

$$\theta(z) = \left(\frac{1}{-2iz}\right)^{1/2} \theta\left(-\frac{1}{4z}\right), \quad \text{equivalently} \quad \theta\left(-\frac{1}{4z}\right) = \left(\frac{2z}{i}\right)^{1/2} \theta(z).$$

Proof. Both sides are holomorphic in $z \in \mathfrak{H}$ (note $\mathrm{Re}(-2iz) = 2\mathrm{Im} z > 0$, so $(-2iz)^{\pm 1/2}$ is holomorphic), so by the identity theorem it suffices to prove the identity for $z = it$, $t > 0$. Then $-2iz = 2t > 0$, and

$$\theta(it) = \sum_n e^{-2\pi n^2 t} = \vartheta(2t), \quad \theta\left(-\frac{1}{4it}\right) = \theta\left(\frac{i}{4t}\right) = \sum_n e^{-2\pi n^2/(4t)} = \vartheta\left(\frac{1}{2t}\right),$$

where $\vartheta(u) = \sum_n e^{-\pi u n^2}$ is the theta function of Chapter 1. By Proposition 1.5.3, $\vartheta(2t) = \frac{1}{\sqrt{2t}} \vartheta\left(\frac{1}{2t}\right)$, i.e. $\theta(it) = (-2i \cdot it)^{-1/2} \theta\left(-\frac{1}{4it}\right)$. ■

About my handwritten proof. My notes apply Poisson summation directly to $g(x) = e^{2\pi i x^2 z} = f(\sqrt{-2iz}x)$ with $f(x) = e^{-\pi x^2}$ and the scaling rule “ $\hat{g}(y) = \frac{1}{b}\hat{f}(\frac{y}{b})$ for $g(x) = f(bx)$ ”. That rule was proved for real $b > 0$ only, while here $b = \sqrt{-2iz}$ is complex; this is why the proof above first takes $z = it$ (where $b = \sqrt{2t} > 0$) and then uses the identity theorem. My notes also contain an attempt via the substitution $t = -\frac{1}{4z}$, which was abandoned.

Lemma 7.6.3

$\Gamma_0(4)$ is generated by $-I$, T and $T' = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}$.

Proof. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4)$. We have $ad - bc = 1$ and $c \equiv 0 \pmod{4}$; thus $ad \equiv 1 \pmod{4}$, so a, d are both odd (and this persists under the operations below, which stay in $\Gamma_0(4)$). Write

$$\gamma T^n = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & an + b \\ c & cn + d \end{pmatrix}.$$

By the division algorithm we can write $b = aq + r$ with $-\frac{|a|}{2} < r \leq \frac{|a|}{2}$; since a is odd, $r \neq \frac{|a|}{2}$, so $|r| < \frac{|a|}{2}$. Taking $n = -q$, $\gamma T^{-q} = \begin{pmatrix} a & r \\ c & * \end{pmatrix}$; let $b_{\text{new}} = r$, so $|b_{\text{new}}| < \frac{|a|}{2}$.

- If $b_{\text{new}} \neq 0$:

$$\begin{pmatrix} a & b_{\text{new}} \\ c & * \end{pmatrix} T^m = \begin{pmatrix} a & b_{\text{new}} \\ c & * \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 4m & 1 \end{pmatrix} = \begin{pmatrix} a + 4mb_{\text{new}} & b_{\text{new}} \\ * & * \end{pmatrix}.$$

By the division algorithm, dividing a by $4b_{\text{new}}$, there is $m \in \mathbb{Z}$ with $a_{\text{new}} := a + 4mb_{\text{new}}$ satisfying $|a_{\text{new}}| \leq \frac{|4b_{\text{new}}|}{2} = 2|b_{\text{new}}|$; since a is odd and $4b_{\text{new}}$ even, a_{new} is odd, so $|a_{\text{new}}| < 2|b_{\text{new}}|$. Since we have $2|b_{\text{new}}| < |a|$ and $|a_{\text{new}}| < 2|b_{\text{new}}|$, we obtain $|a_{\text{new}}| < 2|b_{\text{new}}| < |a|$. In the next round the new b satisfies $|b'| < \frac{|a_{\text{new}}|}{2} < |b_{\text{new}}|$. So $|b|$ strictly decreases, and we can continue this process until $b_{\text{new}} = 0$.

- If $b_{\text{new}} = 0$: $\gamma' = \begin{pmatrix} a' & 0 \\ c' & d' \end{pmatrix} \in \Gamma_0(4)$, so $a'd' = 1$, $a' = d' = \pm 1$, and $\gamma' = \pm \begin{pmatrix} 1 & 0 \\ c' & 1 \end{pmatrix}$ (up to the sign of c'). Since $\gamma' \in \Gamma_0(4)$, $c' = 4k$ for some k , so $\gamma' = \pm \begin{pmatrix} 1 & 0 \\ 4k & 1 \end{pmatrix} = \pm T'^k \in \langle -I, T, T' \rangle$.

Undoing the multiplications by powers of T and T' shows $\gamma \in \langle -I, T, T' \rangle$. ■

Notes. This is consistent with Problem 3.4.6(d), since $ST^{-4}S = -T'$.

Corollary 7.6.4

Let k be even.

- (1) If $k \equiv 0 \pmod{4}$, then $\theta^k \in M_{k/2}(\Gamma_0(4))$.
- (2) If $k \equiv 2 \pmod{4}$, then $\theta^k \in M_{k/2}(4, \chi_{-4})$, where $\chi_{-4} = \begin{pmatrix} -1 \\ \cdot \end{pmatrix}$ (Example 7.2.5).

In both cases $\theta^k|_{k/2}\gamma = \chi_{-4}(d)^{k/2}\theta^k$ for all $\gamma \in \Gamma_0(4)$.

Proof. By Lemma 7.6.3, and since the set of $\gamma \in \Gamma_0(4)$ with $\theta^k|_{k/2}\gamma = \chi_{-4}(d)^{k/2}\theta^k$ is a subgroup (Propositions 6.1.2 and 7.2.6), it suffices to check the invariance under the generators. Write $w = \frac{k}{2} \in \mathbb{Z}$.

Under T ($d = 1$): since $\theta(z+1) = \theta(z)$, $\theta^k(z+1) = \theta^k(z)$, so $\theta^k|_w T = \theta^k$.

Under T' ($d = 1$): note that $T' = -\frac{1}{4}w_4T^{-1}w_4$, where $w_4 = \begin{pmatrix} 0 & -1 \\ 4 & 0 \end{pmatrix}$ (the Atkin–Lehner involution); indeed $w_4T^{-1}w_4 = \begin{pmatrix} 0 & -1 \\ 4 & -4 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} -4 & 0 \\ -16 & -4 \end{pmatrix}$. Note also the functional equation $\theta(w_4z) = \theta(-\frac{1}{4z}) = (\frac{2z}{i})^{1/2}\theta(z)$ (Theorem 7.6.2). So, since k is even,

$$\begin{aligned} (\theta^k|_w w_4)(z) &= 4^{w/2}(4z)^{-w}\theta^k(w_4z) = 4^{k/4}(4z)^{-k/2}\left(\frac{2z}{i}\right)^{k/2}\theta^k(z) = 4^{k/4-k/2}z^{-k/2}\left(\frac{2z}{i}\right)^{k/2}\theta^k(z) \\ &= 2^{-k/2}z^{-k/2}\left(\frac{2z}{i}\right)^{k/2}\theta^k(z) = \left(\frac{1}{2z}\right)^{k/2}\left(\frac{2z}{i}\right)^{k/2}\theta^k(z) = (-i)^{k/2}\theta^k(z). \end{aligned}$$

Now, using Remark 6.1.3 for the scalar $-\frac{1}{4}$ (which gives the factor $(-1)^w$) and $\theta^k|_w T^{-1} = \theta^k$,

$$\begin{aligned} \theta^k|_w T' &= \theta^k|_w \left(-\frac{1}{4}I \cdot w_4T^{-1}w_4\right) = (-1)^{k/2}\theta^k|_w(w_4T^{-1}w_4) = (-1)^{k/2}(-i)^{k/2}\theta^k|_w(T^{-1}w_4) \\ &= (-1)^{k/2}(-i)^{k/2}\theta^k|_w w_4 = (-1)^{k/2}(-i)^{k/2}(-i)^{k/2}\theta^k = (-1)^{k/2}(-1)^{k/2}\theta^k = \theta^k, \end{aligned}$$

since $(-i)^k = ((-i)^2)^{k/2} = (-1)^{k/2}$.

Under $-I$ ($d = -1$): $\theta^k|_w(-I) = (-1)^w\theta^k = (-1)^{k/2}\theta^k$, and $\chi_{-4}(-1)^{k/2} = (-1)^{k/2}$.

- (1) If $k \equiv 0 \pmod{4}$, then $(-1)^{k/2} = 1$, so $\theta^k|_w(-I) = \theta^k$. Hence θ^k is invariant under all generators, so under $\Gamma_0(4)$.
- (2) If $k \equiv 2 \pmod{4}$, then $(-1)^{k/2} = -1$, so $\theta^k|_w(-I) = -\theta^k$. Note that $\chi_{-4}(d) = 1$ for $d \equiv 1$ and -1 for $d \equiv 3 \pmod{4}$; for $-I$, $d = -1 \equiv 3 \pmod{4}$, so $\chi_{-4}(-1) = -1$. Thus $\theta^k|_w\gamma = \chi_{-4}(d)\theta^k$ for all $\gamma \in \Gamma_0(4)$, i.e. θ^k transforms as an element of $M_{k/2}(4, \chi_{-4})$.

Holomorphy at the cusps. For $0 < y \leq 1$ and all x : $|\theta(x+iy)| \leq \sum_n e^{-2\pi n^2 y} = \vartheta(2y) = (2y)^{-1/2}\vartheta(\frac{1}{2y}) \leq Cy^{-1/2}$ (as $\vartheta(u) \leq \vartheta(\frac{1}{2})$ for $u \geq \frac{1}{2}$). So $|\theta^k| \leq C^k y^{-k/2}$, and by Proposition 6.5.2 (applied to $\Gamma_1(4)$, under which θ^k is invariant, in weight $w = \frac{k}{2}$ with $\nu = \frac{k}{2}$) θ^k is holomorphic at all cusps. ■

Correction. In my handwritten notes the holomorphy of θ^k at the cusps was not checked; the growth criterion gives it immediately. In the computation of $\theta^k|_{T'}$ my notes wrote the second factor as $(-1)^{k/2}$ instead of $(-i)^{k/2}$; the final result θ^k is the same.

Chapter 8

Jacobi's Four Squares Theorem and Hecke Operators of Level 1

In the first part of this chapter we harvest the theory of the previous chapters: since $\theta^4 \in M_2(\Gamma_0(4))$ and this space is two-dimensional, spanned by Eisenstein series, we obtain an exact formula for the number $r_4(n)$ of representations of n as a sum of four squares. In the second part we introduce the Hecke operators T_n on $M_k(\mathrm{SL}_2(\mathbb{Z}))$, via functions on lattices, and study Hecke eigenforms. We follow [2, §III.5] and [4, Ch. VII, §5].

8.1 Proof of Jacobi's four squares theorem

Theorem 8.1.1 (Jacobi)

For every integer $n \geq 1$,

$$r_4(n) := \#\left\{(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4 \mid n = \sum_{i=1}^4 x_i^2\right\} = \begin{cases} 8\sigma_1(n), & 4 \nmid n, \\ 8\sigma_1(n) - 32\sigma_1(n/4), & 4 \mid n. \end{cases}$$

Proof. Recall that the Jacobi theta series $\theta(z) = \sum_{n=-\infty}^{\infty} q^{n^2} = \sum_n e^{2\pi i n^2 z}$, $z \in \mathfrak{H}$, satisfies

- (1) $\theta(-\frac{1}{4z}) = (\frac{2z}{i})^{1/2} \theta(z)$ (Theorem 7.6.2);
- (2) $\theta^k \in M_{k/2}(\Gamma_0(4))$ if $k \equiv 0 \pmod{4}$ and $\theta^k \in M_{k/2}(4, \chi_{-4})$ if $k \equiv 2 \pmod{4}$ (Corollary 7.6.4).

Thus $\theta^4 \in M_2(\Gamma_0(4))$.

Step 1: the q -expansion of θ^4 . Now

$$\theta^4(z) = \left(\sum_{n_1} q^{n_1^2}\right) \left(\sum_{n_2} q^{n_2^2}\right) \left(\sum_{n_3} q^{n_3^2}\right) \left(\sum_{n_4} q^{n_4^2}\right) = \sum_{n_1, n_2, n_3, n_4 = -\infty}^{\infty} q^{n_1^2 + n_2^2 + n_3^2 + n_4^2} = \sum_{m=0}^{\infty} r_4(m) q^m$$

(the product of absolutely convergent series may be multiplied out and regrouped). Indeed, letting $m = n_1^2 + n_2^2 + n_3^2 + n_4^2$, the coefficient of q^m counts $\#\{(n_1, n_2, n_3, n_4) \in \mathbb{Z}^4 : m = n_1^2 + \dots + n_4^2\}$, i.e. how many ways represent m . By manually counting, $r_4(0) = \#\{(0, 0, 0, 0)\} = 1$ and

$$r_4(1) = \#\{(\pm 1, 0, 0, 0), (0, \pm 1, 0, 0), (0, 0, \pm 1, 0), (0, 0, 0, \pm 1)\} = 8,$$

and similarly $r_4(2) = 24$, $r_4(3) = 32$, $r_4(4) = 24$. So

$$\theta^4(z) = 1 + 8q + 24q^2 + 32q^3 + 24q^4 + \dots$$

Step 2: a basis of $M_2(\Gamma_0(4))$. Recall the Eisenstein series $E_{2,2}(z) = E_2(z) - 2E_2(2z)$ (Proposition 7.5.3 with $p = 2$), so $E_{2,2} \in M_2(\Gamma_0(2))$. Also recall:

- (1) if $\Gamma'' \subseteq \Gamma'$, then $M_k(\Gamma') \subseteq M_k(\Gamma'')$ (Proposition 7.1.1);

(2) if $f \in M_k(\Gamma_0(N))$, then $f(\ell z) \in M_k(\Gamma_0(N\ell))$ (Proposition 7.4.2).

Hence, since $\Gamma_0(4) \subseteq \Gamma_0(2)$, from (1) $M_2(\Gamma_0(2)) \subseteq M_2(\Gamma_0(4))$, so $E_{2,2}(z) \in M_2(\Gamma_0(4))$. Also, from (2), $E_{2,2}(2z) \in M_2(\Gamma_0(4))$.

Claim: $E_{2,2}(z)$ and $E_{2,2}(2z)$ form a basis of $M_2(\Gamma_0(4))$. Notice that, with $E_2 = 1 - 24 \sum \sigma_1(n)q^n$,

$$\begin{aligned} E_{2,2}(z) &= E_2(z) - 2E_2(2z) = \left(1 - 24 \sum_{n \geq 1} \sigma_1(n)q^n\right) - 2\left(1 - 24 \sum_{n \geq 1} \sigma_1(n)q^{2n}\right) \\ &= -1 - 24 \sum_{n \geq 1} \sigma_1(n)q^n + 48 \sum_{n \geq 1} \sigma_1(n)q^{2n}, \end{aligned}$$

and so $E_{2,2}(2z) = -1 - 24 \sum_n \sigma_1(n)q^{2n} + 48 \sum_n \sigma_1(n)q^{4n}$. Explicitly,

$$\begin{aligned} E_{2,2}(z) &= -1 - 24q - 72q^2 - 96q^3 - 168q^4 - \dots + 48q^2 + 144q^4 + \dots \\ &= -1 - 24q - 24q^2 - 96q^3 - 24q^4 - \dots, \\ E_{2,2}(2z) &= -1 - 24q^2 - 72q^4 - \dots + 48q^4 + \dots = -1 - 24q^2 - 24q^4 - \dots. \end{aligned}$$

Thus they are linearly independent (the coefficient of q is -24 for the first and 0 for the second, while the constant terms agree). Also, since $[\bar{\Gamma} : \bar{\Gamma}_0(4)] = 6$ (Proposition 3.2.10(4): $4(1 + \frac{1}{2}) = 6$; note $-I \in \Gamma_0(4)$), the dimension bound (Proposition 6.4.3) gives

$$\dim M_2(\Gamma_0(4)) \leq \frac{2}{12} \cdot 6 + 1 = 2.$$

Then $\dim M_2(\Gamma_0(4)) = 2$, and $E_{2,2}(z), E_{2,2}(2z)$ is a basis (a basis is a maximal linearly independent set).

Step 3: comparing coefficients. Since $\theta^4 \in M_2(\Gamma_0(4)) = \langle E_{2,2}(z), E_{2,2}(2z) \rangle$, there exist α, β with $\theta^4(z) = \alpha E_{2,2}(z) + \beta E_{2,2}(2z)$. Then

$$\alpha E_{2,2}(z) + \beta E_{2,2}(2z) = (-\alpha - \beta) + (-24\alpha)q + (-\alpha - \beta)24q^2 + \dots,$$

and comparing with $\theta^4 = 1 + 8q + 24q^2 + \dots$:

$$-\alpha - \beta = 1 \quad \text{and} \quad 8 = -24\alpha \quad \implies \quad \alpha = -\frac{1}{3}, \quad \beta = -\alpha - 1 = \frac{1}{3} - 1 = -\frac{2}{3}.$$

Hence $\theta^4(z) = -\frac{1}{3}E_{2,2}(z) - \frac{2}{3}E_{2,2}(2z)$, i.e.

$$\begin{aligned} \sum_{n \geq 0} r_4(n)q^n &= -\frac{1}{3} \left(-1 - 24 \sum_n \sigma_1(n)q^n + 48 \sum_n \sigma_1(n)q^{2n} \right) \\ &\quad - \frac{2}{3} \left(-1 - 24 \sum_n \sigma_1(n)q^{2n} + 48 \sum_n \sigma_1(n)q^{4n} \right) \\ &= \left(\frac{1}{3} + \frac{2}{3} \right) + 8 \sum_n \sigma_1(n)q^n + (-16 + 16) \sum_n \sigma_1(n)q^{2n} - 32 \sum_n \sigma_1(n)q^{4n} \\ &= 1 + 8 \sum_{n \geq 1} \sigma_1(n)q^n - 32 \sum_{n \geq 1} \sigma_1(n)q^{4n} \\ &= 1 + \sum_{n \geq 1} \left(8\sigma_1(n) - 32\sigma_1\left(\frac{n}{4}\right) \right) q^n, \end{aligned}$$

with the convention $\sigma_1(n/4) = 0$ if $4 \nmid n$. Comparing coefficients of q^n gives the theorem. ■

Corollary 8.1.2 (Lagrange's four squares theorem)

Every positive integer is a sum of four squares; in fact $r_4(n) = 8 \sum_{d|n, 4 \nmid d} d \geq 8$.

Proof. The divisors d of n with $4 \nmid d$ are $d = 4e$ with $e \mid \frac{n}{4}$, so $\sum_{d|n, 4 \nmid d} d = 4\sigma_1(n/4)$, and $8\sigma_1(n) - 32\sigma_1(n/4) = 8(\sigma_1(n) - 4\sigma_1(n/4)) = 8 \sum_{d|n, 4 \nmid d} d \geq 8 \cdot 1$. ■

Correction. In my handwritten notes the coefficient $r_4(3)$ was written as 82; it is 32 (the representations of 3 are $(\pm 1, \pm 1, \pm 1, 0)$ in some order: $4 \cdot 2^3 = 32$). The rest of the computation uses only the coefficients of 1, q , q^2 .

Roadmap. My notes contain at this point a small diagram: the four squares theorem as an example of the connection between modular forms and theta series (quadratic forms), lattices appearing in number theory, physics and computer science, and the Monster group; and the remark “don't do generating functions” (i.e. the point of the proof is the finite dimensionality of $M_2(\Gamma_0(4))$, not a direct manipulation of generating functions).

8.2 Modular forms as functions on lattices

We interpret modular forms on complex lattices. For $z \in \mathfrak{H}$ let $L_z = \langle z, 1 \rangle = \mathbb{Z}z + \mathbb{Z}$. Recall that

$$\{\text{lattices up to homothety}\} \longleftrightarrow \text{SL}_2(\mathbb{Z}) \backslash \mathfrak{H}, \quad L_z \longleftrightarrow z,$$

and $\langle z', 1 \rangle \sim \langle z, 1 \rangle$ (homothetic) iff there is $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z})$ with $z' = \gamma z$; in this case $\langle z', 1 \rangle = (cz + d)^{-1} \langle az + b, cz + d \rangle = (cz + d)^{-1} \langle z, 1 \rangle$, i.e. $L_{z'} = \lambda L_z$ with $\lambda = (cz + d)^{-1}$ (by Lemma 8.2.1 below and Proposition 2.2.6).

Lemma 8.2.1

Let $L = \langle \omega_1, \omega_2 \rangle$ and $L' = \langle \omega'_1, \omega'_2 \rangle$ with $\frac{\omega_1}{\omega_2}, \frac{\omega'_1}{\omega'_2} \in \mathfrak{H}$. Then $L = L'$ if and only if there is $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z})$ with $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = \gamma \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$.

Proof. ($\Rightarrow$) Since $\omega'_1, \omega'_2 \in L' = L$, there are $a, b, c, d \in \mathbb{Z}$ with $\omega'_1 = a\omega_1 + b\omega_2$ and $\omega'_2 = c\omega_1 + d\omega_2$, i.e. $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$; we now check $\det \gamma$. Since $\omega_1, \omega_2 \in L = L'$, the same argument gives $\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \gamma^{-1} \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix}$ with an integer matrix, so $\det \gamma = \pm 1$ (Proposition 2.2.6). Let $\tau = \frac{\omega_1}{\omega_2}$ and $\tau' = \frac{\omega'_1}{\omega'_2}$. Then

$$\tau' = \frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2} = \frac{a\frac{\omega_1}{\omega_2} + b}{c\frac{\omega_1}{\omega_2} + d} = \frac{a\tau + b}{c\tau + d} \implies \text{Im}(\tau') = \frac{\det(\gamma)}{|c\tau + d|^2} \text{Im}(\tau).$$

Since $\tau, \tau' \in \mathfrak{H}$, $\text{Im}(\tau), \text{Im}(\tau') > 0$; thus $\det(\gamma) > 0$, so $\det(\gamma) = 1$. Hence $\gamma \in \text{SL}_2(\mathbb{Z})$.

($\Leftarrow$) Let $\gamma \in \text{SL}_2(\mathbb{Z})$ with $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = \gamma \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$. Since $a, b, c, d \in \mathbb{Z}$, $\omega'_1 = a\omega_1 + b\omega_2$ and $\omega'_2 = c\omega_1 + d\omega_2$ imply $\omega'_1, \omega'_2 \in L$; since $L' = \langle \omega'_1, \omega'_2 \rangle$, $L' \subseteq L$. Also, since $\det(\gamma) = 1$, $\gamma^{-1} \in \text{SL}_2(\mathbb{Z})$, so $\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \gamma^{-1} \begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix}$ gives $\omega_1, \omega_2 \in L'$. Hence $L \subseteq L'$, and $L = L'$. ■

Proposition 8.2.2

Let $\mathcal{L}_k$ be the set of functions $F : \{\text{lattices in } \mathbb{C}\} \rightarrow \mathbb{C}$ which are *homogeneous of weight $-k$* , i.e. $F(\lambda L) = \lambda^{-k} F(L)$ for all $\lambda \in \mathbb{C}^\times$. Then

$$f \mapsto F, \quad F(\langle \omega_1, \omega_2 \rangle) = \omega_2^{-k} f\left(\frac{\omega_1}{\omega_2}\right) \quad \left(\frac{\omega_1}{\omega_2} \in \mathfrak{H}\right),$$

is a bijection between weakly modular functions f of weight k for $\text{SL}_2(\mathbb{Z})$ and $\mathcal{L}_k$, with inverse $F \mapsto f$, $f(z) = F(L_z)$.

Proof. F is well defined. Every lattice has a basis with $\frac{\omega_1}{\omega_2} \in \mathfrak{H}$ (swap the basis if necessary). If $\langle \omega'_1, \omega'_2 \rangle$ is another such basis of the same lattice L , then by Lemma 8.2.1 $\omega'_1 = a\omega_1 + b\omega_2$, $\omega'_2 = c\omega_1 + d\omega_2$ with $\gamma \in \text{SL}_2(\mathbb{Z})$, and

$$\begin{aligned} (\omega'_2)^{-k} f\left(\frac{\omega'_1}{\omega'_2}\right) &= (c\omega_1 + d\omega_2)^{-k} f\left(\frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2} \cdot \frac{1/\omega_2}{1/\omega_2}\right) = \omega_2^{-k} \left(c\frac{\omega_1}{\omega_2} + d\right)^{-k} f\left(\frac{a\frac{\omega_1}{\omega_2} + b}{c\frac{\omega_1}{\omega_2} + d}\right) \\ &= \omega_2^{-k} f\left(\frac{\omega_1}{\omega_2}\right) \end{aligned}$$

since $f(\gamma\tau) = (c\tau + d)^k f(\tau)$. So F is invariant under the change of basis by γ , i.e. well defined. It is homogeneous: $F(\lambda L) = (\lambda\omega_2)^{-k} f\left(\frac{\lambda\omega_1}{\lambda\omega_2}\right) = \lambda^{-k} F(L)$. We use this homogeneity to define F on all lattices: for $L = \langle \omega_1, \omega_2 \rangle$ we have $L = \omega_2 \langle \frac{\omega_1}{\omega_2}, 1 \rangle = \omega_2 L_{\omega_1/\omega_2}$, so $F(L) = \omega_2^{-k} F(L_{\omega_1/\omega_2}) = \omega_2^{-k} f\left(\frac{\omega_1}{\omega_2}\right)$.

Conversely, if $F \in \mathcal{L}_k$ and $f(z) := F(L_z)$, then for $\gamma \in \text{SL}_2(\mathbb{Z})$:

$$\begin{aligned} (f|_k \gamma)(z) &= (cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right) = (cz + d)^{-k} F\left(\left\langle \frac{az + b}{cz + d}, 1 \right\rangle\right) \\ &= (cz + d)^{-k} F((cz + d)^{-1} \langle az + b, cz + d \rangle) \\ &= (cz + d)^{-k} ((cz + d)^{-1})^{-k} F(\langle az + b, cz + d \rangle) \\ &= F(\langle az + b, cz + d \rangle) = F(\langle z, 1 \rangle) = f(z), \end{aligned}$$

using Lemma 8.2.1: since $\begin{pmatrix} az+b \\ cz+d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix}$, we have $\langle az + b, cz + d \rangle = \langle z, 1 \rangle$. Therefore f is weakly modular of weight k . The two constructions are clearly inverse to each other. ■

8.3 Hecke operators on $M_k(\text{SL}_2(\mathbb{Z}))$

We use the above correspondence to define Hecke operators on $\mathcal{L}_k$ and then interpret them on $M_k(\text{SL}_2(\mathbb{Z}))$.

Definition 8.3.1

Fix k and $n \geq 1$. Define $T_n : \mathcal{L}_k \rightarrow \mathcal{L}_k$ by

$$(T_n F)(L) = n^{k-1} \sum_{[L:L']=n} F(L'),$$

the sum running over all sublattices $L' \subseteq L$ of index n . (The operator T_n “averages” F over all sublattices of index n .)

Proposition 8.3.2

A lattice L has only finitely many sublattices of index n ; so the sum above is finite. Moreover $T_n F$ is again homogeneous of weight $-k$.

Proof. Let $L' \subseteq L$ be a sublattice of index n . Since L is an abelian group under $+$ and L' is a subgroup of L , L/L' is a quotient group of order $[L : L'] = n$, which is finite. Thus for all $\ell \in L$, $\ell + L' \in L/L'$, and by Lagrange's theorem $n(\ell + L') = 0 + L'$, i.e. $n\ell + L' = 0 + L'$, so $n\ell \in L'$. Hence $nL \subseteq L' \subseteq L$. The subgroups L' with $nL \subseteq L' \subseteq L$ correspond bijectively to the subgroups of the finite group $L/nL \cong (\mathbb{Z}/n\mathbb{Z})^2$, which has n^2 elements, hence only finitely many subgroups. So there are only finitely many L' , and the sum is finite. Homogeneity: the sublattices of index n of λL are exactly the $\lambda L'$, so $(T_n F)(\lambda L) = n^{k-1} \sum F(\lambda L') = \lambda^{-k} (T_n F)(L)$. ■

Correction. In my handwritten notes the conclusion was stated as “ L' is finite”, with question marks. What is finite is the *number* of such sublattices L' (each L' is an infinite group): they are squeezed between nL and L , and L/nL is finite.

Lemma 8.3.3

Let $L = \langle \omega_1, \omega_2 \rangle$ with $\frac{\omega_1}{\omega_2} \in \mathfrak{H}$. A sublattice $L' \subseteq L$ has index n if and only if $L' = \langle \omega'_1, \omega'_2 \rangle$ with $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$ for some $M \in M_{2 \times 2}(\mathbb{Z})$ with $\det M = n$ (and then $\frac{\omega'_1}{\omega'_2} \in \mathfrak{H}$).

Proof. Index of ML . Let $M \in M_{2 \times 2}(\mathbb{Z})$ with $\det M \neq 0$, and let $L_M = \langle \omega'_1, \omega'_2 \rangle$ with $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$. By the elementary divisor theorem (Smith normal form) there are $U, V \in \mathrm{GL}_2(\mathbb{Z})$ with $M = UDV$, $D = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$, $d_1, d_2 \geq 1$. Put $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = V \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$; this is another basis of L (Proposition 2.2.6), and $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = U \begin{pmatrix} d_1 e_1 \\ d_2 e_2 \end{pmatrix}$, so $L_M = \langle d_1 e_1, d_2 e_2 \rangle$ (again by Proposition 2.2.6). Hence $L/L_M \cong \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ and $[L : L_M] = d_1 d_2 = |\det M|$.

($\Leftarrow$) If $\det M = n$, then $[L : L_M] = n$, and $\mathrm{Im} \frac{\omega'_1}{\omega'_2} = \frac{\det M}{|c\tau + d|^2} \mathrm{Im} \tau > 0$ as in Lemma 8.2.1.

($\Rightarrow$) Let $[L : L'] = n$. Then $nL \subseteq L' \subseteq L$ (proof of Proposition 8.3.2), so L' is a subgroup of $L \cong \mathbb{Z}^2$ containing a rank 2 subgroup; hence L' is free of rank 2, and we can choose a basis ω'_1, ω'_2 of L' with $\frac{\omega'_1}{\omega'_2} \in \mathfrak{H}$. Since $L' \subseteq L$, $\begin{pmatrix} \omega'_1 \\ \omega'_2 \end{pmatrix} = M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$ with $M \in M_{2 \times 2}(\mathbb{Z})$, and by the first part $|\det M| = n$; since both quotients lie in $\mathfrak{H}$, $\det M > 0$, so $\det M = n$. ■

Correction. In my handwritten notes Lemma 8.3.3 was stated but its proof was left blank; the proof above uses the elementary divisor theorem.

Let $\mathcal{M}_n = \{M \in M_{2 \times 2}(\mathbb{Z}) : \det M = n\}$. For $M, V \in \mathcal{M}_n$ we have, by Lemma 8.2.1, $L_M = L_V$ iff there is $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $M \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \gamma(V \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}) = (\gamma V) \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$; since ω_1, ω_2 are linearly independent over $\mathbb{R}$ and $M - \gamma V$ has integer entries, this means $M = \gamma V$. Thus $\mathrm{SL}_2(\mathbb{Z})M = \mathrm{SL}_2(\mathbb{Z})V$ (the left cosets generated by M and V coincide), and

$$\{L' \subseteq L \text{ of index } n\} \longleftrightarrow \mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n, \quad L' = L_M \leftrightarrow \mathrm{SL}_2(\mathbb{Z})M.$$

Proposition 8.3.4

A complete set of representatives of $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n$ is

$$\mathcal{R}_n = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : ad = n, a \geq 1, 0 \leq b < d \right\}.$$

Proof. Every coset contains an element of $\mathcal{R}_n$. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_n$ be arbitrary. We would like to show $\gamma M \in \mathcal{R}_n$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. If $c \neq 0$, let $r = \gcd(a, c)$. By Bézout's theorem there are $x, y \in \mathbb{Z}$ with $ax + cy = r$. Thus we can construct

$$\gamma = \begin{pmatrix} x & y \\ -c/r & a/r \end{pmatrix},$$

and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ since $\det \gamma = \frac{a}{r}x + \frac{c}{r}y = \frac{ax+cy}{r} = 1$. So

$$\gamma M = \begin{pmatrix} x & y \\ -c/r & a/r \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ax + cy & bx + dy \\ -\frac{ac}{r} + \frac{ac}{r} & -\frac{bc}{r} + \frac{ad}{r} \end{pmatrix} =: \begin{pmatrix} a^* & b^* \\ 0 & d^* \end{pmatrix}.$$

(If $c = 0$, take $\gamma = I$.) Also $\det(\gamma M) = \det \gamma \det M = 1 \cdot n$, so $a^* d^* = n$; in particular $a^*, d^* \neq 0$. Is $a^* \geq 1$? If $a^* < 0$, left-multiply by $-I \in \mathrm{SL}_2(\mathbb{Z})$: $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a^* & b^* \\ 0 & d^* \end{pmatrix} = \begin{pmatrix} -a^* & -b^* \\ 0 & -d^* \end{pmatrix}$; let $a^{**} = -a^* \geq 1$ and $d^{**} = -d^*$; from $a^* d^* = n \geq 1$, $d^* \leq -1$, so $d^{**} \geq 1$. Is $0 \leq b^* < d^*$? If we left-multiply by $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^* & b^* \\ 0 & d^* \end{pmatrix} = \begin{pmatrix} a^* & b^* + kd^* \\ 0 & d^* \end{pmatrix}.$$

By the Euclidean algorithm there are q, r' with $0 \leq r' < d^*$ and $b^* = qd^* + r'$; choosing $k = -q$ we have $b^* + kd^* = b^* - qd^* = r'$, so $0 \leq r' < d^*$. Therefore the coset contains an element of $\mathcal{R}_n$.

Distinct elements of $\mathcal{R}_n$ lie in distinct cosets. Suppose $\gamma \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix}$ with both in $\mathcal{R}_n$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Then $\gamma = \begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}^{-1}$ is upper triangular with diagonal entries $\frac{a'}{a}, \frac{d'}{d} > 0$ and determinant 1; being integers, $\frac{a'}{a} = \frac{d'}{d} = 1$. So $\gamma = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ and $b' = b + kd$; as $0 \leq b, b' < d$, $k = 0$. Hence the two elements coincide. ■

Correction. In my handwritten proof it was concluded that “ $a^* d^* = 1, a^* = 1$ ”; in fact $a^* d^* = n$ and $a^* = \gcd(a, c)$. Also the irredundancy of $\mathcal{R}_n$ (second part) was not checked there.

The action of T_n on modular forms

Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ correspond to $F \in \mathcal{L}_k$, $f(z) = F(\langle z, 1 \rangle)$. For $L = \langle z, 1 \rangle$ and $L' = \langle az + b, cz + d \rangle = ML$ with $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_n$ (so $[L : L'] = n$ by Lemma 8.3.3),

$$\begin{aligned} F(L') &= F(\langle az + b, cz + d \rangle) = F\left((cz + d) \left\langle \frac{az + b}{cz + d}, 1 \right\rangle\right) = (cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right) \\ &= (\det M)^{-k/2} (f|_k M)(z) = n^{-k/2} (f|_k M)(z). \end{aligned}$$

So, defining $T_n f$ by $(T_n f)(z) = (T_n F)(L_z)$,

$$\begin{aligned} (T_n f)(z) &= n^{k-1} \sum_{[L:L']=n} F(L') = n^{k-1} \sum_{M \in \mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n} n^{-k/2} (f|_k M)(z) \\ &= n^{k/2-1} \sum_{M \in \mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n} (f|_k M)(z). \end{aligned} \quad (8.1)$$

Theorem 8.3.5

T_n is a linear operator $M_k(\mathrm{SL}_2(\mathbb{Z})) \rightarrow M_k(\mathrm{SL}_2(\mathbb{Z}))$, and it maps $S_k(\mathrm{SL}_2(\mathbb{Z}))$ to itself. If $f(z) = \sum_{m \geq 0} a_m q^m$, then

$$(T_n f)(z) = \sum_{m \geq 0} \left(\sum_{\substack{d | \gcd(m, n) \\ d \geq 1}} d^{k-1} a_{mn/d^2} \right) q^m.$$

In particular $a_0(T_n f) = \sigma_{k-1}(n) a_0(f)$ and $a_1(T_n f) = a_n(f)$.

Proof. 1. Well-definedness. Let $\{M_i\}$ and $\{M'_i\}$ be two sets of representatives of $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n$; then (after reordering) there are $\gamma_i \in \mathrm{SL}_2(\mathbb{Z})$ with $M'_i = \gamma_i M_i$. Thus $f|_k M'_i = f|_k \gamma_i M_i = (f|_k \gamma_i)|_k M_i = f|_k M_i$. Since $f|_k M'_i = f|_k M_i$, $T_n(f)$ is well defined.

2. Linearity. By definition of $|_k$, T_n is linear.

3. Invariance under $\mathrm{SL}_2(\mathbb{Z})$. Note that for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,

$$(T_n f)|_k \gamma = n^{k/2-1} \sum_{M \in \mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n} (f|_k M)|_k \gamma = n^{k/2-1} \sum_M f|_k M \gamma.$$

Since $\det(M\gamma) = n \cdot 1 = n$, $M\gamma \in \mathcal{M}_n$; furthermore, right multiplication by γ maps distinct cosets $\mathrm{SL}_2(\mathbb{Z})M$ to distinct cosets and is a bijection of $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_n$ (with inverse right multiplication by γ^{-1}). Let $M' = M\gamma$. Hence $(T_n f)|_k \gamma = n^{k/2-1} \sum_{M'} f|_k M' = T_n(f)$, so $T_n f$ is invariant under γ .

4. Holomorphy on $\mathfrak{H}$. Using the representatives $\mathcal{R}_n$ of Proposition 8.3.4:

$$T_n(f)(z) = n^{k/2-1} \sum_{M \in \mathcal{R}_n} (\det M)^{k/2} (cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right) = n^{k-1} \sum_{\substack{a, d > 0 \\ ad = n}} \sum_{b=0}^{d-1} d^{-k} f\left(\frac{az + b}{d}\right).$$

Since f is holomorphic on $\mathfrak{H}$, so is $T_n(f)$ (a finite sum).

5. Holomorphy at the cusp: the q -expansion. Substituting $f(w) = \sum_m a_m e^{2\pi i m w}$,

$$T_n(f)(z) = n^{k-1} \sum_{\substack{a, d > 0 \\ ad = n}} d^{-k} \sum_{m \geq 0} a_m e^{2\pi i m a z / d} \underbrace{\sum_{b=0}^{d-1} e^{2\pi i m b / d}}_{(1)}.$$

(1) $\sum_{b=0}^{d-1} e^{2\pi i m b / d} = d$ if $d \mid m$, and $= 0$ if $d \nmid m$ (a geometric sum of a d -th root of unity $\neq 1$ vanishes).

So only $m = dr$ survive, and $n^{k-1} d^{-k} \cdot d = n^{k-1} d^{1-k} = a^{k-1}$ (as $n = ad$):

$$T_n(f)(z) = \sum_{\substack{a, d > 0 \\ ad = n}} a^{k-1} \sum_{r \geq 0} a_{dr} e^{2\pi i a r z} = \sum_{\substack{a, d > 0 \\ ad = n}} a^{k-1} \sum_{r \geq 0} a_{dr} q^{ar}.$$

(2) Collect the terms with $ar = m$: then $a \mid n$ and $a \mid m$, and $a_{dr} = a_{\frac{n}{a}, \frac{m}{a}} = a_{nm/a^2}$.

(3) Rewrite a as d :

$$T_n(f)(z) = \sum_{m \geq 0} \left(\sum_{\substack{d|n \\ d|m}} d^{k-1} a_{nm/d^2} \right) q^m.$$

There are no negative powers of q , so $T_n(f)$ is holomorphic at ∞ , and $T_n(f) \in M_k(\mathrm{SL}_2(\mathbb{Z}))$. For $m = 0$ every $d \mid n$ divides 0, so $a_0(T_n f) = \sum_{d|n} d^{k-1} a_0 = \sigma_{k-1}(n) a_0$; hence cusp forms go to cusp forms. For $m = 1$ only $d = 1$ occurs, so $a_1(T_n f) = a_n$. ■

The operator T_n on $M_k(\mathrm{SL}_2(\mathbb{Z}))$ is called the n -th Hecke operator.

Proposition 8.3.6

For a prime p , $T_p f = U_p f + p^{k-1} V_p f$, where

$$(U_p f)(z) = \frac{1}{p} \sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right) = \sum_{n \geq 0} a_{np} q^n, \quad (V_p f)(z) = f(pz) = \sum_{m \geq 0} a_m q^{pm}.$$

Proof. If $n = p$, by Proposition 8.3.4 the set of coset representatives of $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathcal{M}_p$ is $\left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : ad = p, a \geq 1, 0 \leq b < d \right\}$. Since p is prime, the divisors d of p are $d \in \{1, p\}$:

- (1) if $d = p$, then $a = 1$, so $0 \leq b < p$, and there are p matrices of the form $\begin{pmatrix} 1 & b \\ 0 & p \end{pmatrix}$;
- (2) if $d = 1$, then $a = p$, and $0 \leq b < 1$ forces $b = 0$: there is only one matrix, $\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$.

Therefore there are $p + 1$ representatives, and by the formula in step 4 above

$$\begin{aligned} T_p(f) &= p^{k-1} \sum_{b=0}^{p-1} p^{-k} f\left(\frac{z+b}{p}\right) + p^{k-1} \cdot 1^{-k} f(pz) \\ &= \frac{1}{p} \sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right) + p^{k-1} f(pz) = U_p(f) + p^{k-1} V_p(f). \end{aligned}$$

By $f(z) = \sum_m a_m q^m = \sum_m a_m e^{2\pi i m z}$,

$$U_p(f) = \frac{1}{p} \sum_{b=0}^{p-1} \sum_{m \geq 0} a_m e^{2\pi i m \frac{z+b}{p}} = \frac{1}{p} \sum_{m \geq 0} a_m e^{2\pi i m z/p} \sum_{b=0}^{p-1} e^{2\pi i m b/p} = \frac{1}{p} \sum_{\substack{m \geq 0 \\ p|m}} a_m q^{m/p} \cdot p = \sum_{n \geq 0} a_{np} q^n,$$

writing $\frac{m}{p} = n$, i.e. $m = np$; and $V_p(f) = f(pz) = \sum_m a_m q^{pm}$. ■

Correction. In my handwritten notes the formula for $T_p(f)$ was first written with $(cz + d)^k$ instead of $(cz + d)^{-k}$, and the evaluation of $U_p(f)$ cancelled the sum over b without first using the vanishing of $\sum_b e^{2\pi i m b/p}$ for $p \nmid m$.

8.4 Multiplicativity of Hecke operators

Theorem 8.4.1

For any $n, m \geq 1$,

$$T_n T_m = \sum_{\substack{d \mid \gcd(n, m) \\ d \geq 1}} d^{k-1} T_{nm/d^2}$$

as operators on $M_k(\mathrm{SL}_2(\mathbb{Z}))$.

Proof. Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and write $a_r(g)$ for the r -th coefficient of g . By Theorem 8.3.5, $a_r(T_n f) = \sum_{d \mid \gcd(n, r)} d^{k-1} a_{nr/d^2}(f)$. We observe the r -th coefficient of $T_n(T_m f)$:

$$\begin{aligned} a_r(T_n(T_m f)) &= \sum_{d \mid \gcd(n, r)} d^{k-1} a_{nr/d^2}(T_m f) = \sum_{d \mid \gcd(n, r)} d^{k-1} \left(\sum_{c \mid \gcd(m, \frac{nr}{d^2})} c^{k-1} a_{\frac{mnr}{c^2 d^2}}(f) \right) \\ &= \sum_{d \mid \gcd(n, r)} \sum_{c \mid \gcd(m, \frac{nr}{d^2})} (cd)^{k-1} a_{\frac{nmr}{(cd)^2}}(f). \end{aligned} \quad (\star)$$

To show $T_n T_m = \sum_{e \mid (n, m)} e^{k-1} T_{nm/e^2}$, we need to show that the r -th coefficient of the right-hand side equals $(\star)$. Note that it is

$$\begin{aligned} \sum_{e \mid \gcd(n, m)} e^{k-1} a_r(T_{nm/e^2} f) &= \sum_{e \mid \gcd(n, m)} e^{k-1} \sum_{g \mid \gcd(\frac{nm}{e^2}, r)} g^{k-1} a_{\frac{nmr}{e^2 g^2}}(f) \\ &= \sum_{e \mid \gcd(n, m)} \sum_{g \mid \gcd(\frac{nm}{e^2}, r)} (eg)^{k-1} a_{\frac{nmr}{(eg)^2}}(f). \end{aligned} \quad (\star\star)$$

In both double sums the summand only depends on the product $x = cd$, resp. $x = eg$. So it suffices to show that for every fixed $x \geq 1$,

$$\begin{aligned} A(x) &:= \#\left\{ (d, c) : d \mid n, d \mid r, c \mid m, c \mid \frac{nr}{d^2}, cd = x \right\} \\ &\stackrel{!}{=} B(x) := \#\left\{ (e, g) : e \mid n, e \mid m, g \mid r, g \mid \frac{nm}{e^2}, eg = x \right\}. \end{aligned}$$

Observe that $B(x)$ is obtained from $A(x)$ by exchanging the roles of m and r (and renaming $d \rightarrow e, c \rightarrow g$). So it suffices to show that $A(x)$ is symmetric in m and r . All conditions are divisibility conditions, so they can be checked prime by prime: a pair (d, c) with $cd = x$ is the same as a choice, for every prime p , of $\delta = v_p(d)$ and $\gamma = v_p(c) = v_p(x) - \delta$, and the conditions hold iff they hold for the exponents at every p . Hence $A(x) = \prod_p A_p$, where, with $\alpha = v_p(n)$, $\beta = v_p(m)$, $\varrho = v_p(r)$, $\xi = v_p(x)$,

$$A_p = \#\{\delta \in \mathbb{Z} : 0 \leq \delta \leq \alpha, \delta \leq \varrho, 0 \leq \xi - \delta \leq \beta, \xi - \delta \leq \alpha + \varrho - 2\delta\}.$$

The conditions say $\max(0, \xi - \beta) \leq \delta \leq \min(\alpha, \varrho, \xi, \alpha + \varrho - \xi)$, so

$$\begin{aligned} A_p &= \max(0, \min(\alpha, \varrho, \xi, \alpha + \varrho - \xi) - \max(0, \xi - \beta) + 1) \\ &= \max(0, 1 + \min(\alpha, \varrho, \beta, \xi, \alpha + \varrho - \xi, \alpha + \beta - \xi, \varrho + \beta - \xi, \alpha + \varrho + \beta - 2\xi)), \end{aligned}$$

using $\min(u_i) - \max(0, v) = \min(u_i, u_i - v)$ and expanding. The set $\{\alpha, \varrho, \beta, \xi, \alpha + \varrho - \xi, \alpha + \beta - \xi, \varrho + \beta - \xi, \alpha + \varrho + \beta - 2\xi\}$ is invariant under $\beta \leftrightarrow \varrho$. Hence A_p , and therefore $A(x)$, is symmetric in m and r , so $A(x) = B(x)$ and $(\star) = (\star\star)$. $\blacksquare$

Correction. In my handwritten proof the equality of $(\star)$ and $(\star\star)$ was reduced to the equality of the numbers of ways to split $x = c \cdot d$ and $x = e \cdot g$, and I observed that “these two statements are equivalent, just m and r swapped” (with a question mark at the bounds

for d). The prime-by-prime count above makes this symmetry precise.

Corollary 8.4.2

For all m, n : $T_n T_m = T_m T_n$; and if $\gcd(n, m) = 1$, then $T_n T_m = T_{nm}$.

Proof. The right-hand side of Theorem 8.4.1 is symmetric in n, m , because $\gcd(n, m) = \gcd(m, n)$ and $nm = mn$. If $\gcd(n, m) = 1$, then $d \mid \gcd(n, m)$ forces $d = 1$, so $T_n T_m = \sum_{d \mid 1} d^{k-1} T_{nm/d^2} = 1^{k-1} T_{nm/1^2} = T_{nm}$. ■

Proposition 8.4.3

The operators T_p (p prime) generate all T_n : T_{p^α} is a polynomial in T_p with integer coefficients, determined by $T_{p^{\alpha+1}} = T_{p^\alpha} T_p - p^{k-1} T_{p^{\alpha-1}}$, and $T_n = \prod_i T_{p_i^{\alpha_i}}$ for $n = \prod_i p_i^{\alpha_i}$.

Proof. Note that we already have $T_n T_m = \sum_{d \mid \gcd(n, m)} d^{k-1} T_{nm/d^2}$, and $T_n T_m = T_{nm}$ if $\gcd(n, m) = 1$. Let $p \mid n$.

(1) If $p \nmid \frac{n}{p}$ (i.e. $\gcd(\frac{n}{p}, p) = 1$), then $T_{n/p} T_p = T_{\frac{n}{p} \cdot p} = T_n$.

(2) If $p \mid \frac{n}{p}$, then $\gcd(\frac{n}{p}, p) = p$, and

$$T_{n/p} T_p = \sum_{d \mid p} d^{k-1} T_{\frac{(n/p)p}{d^2}} = 1^{k-1} T_n + p^{k-1} T_{n/p^2} \implies T_n = T_{n/p} T_p - p^{k-1} T_{n/p^2}.$$

Let $n = p^\alpha$: $T_{p^\alpha} = T_{p^{\alpha-1}} T_p - p^{k-1} T_{p^{\alpha-2}}$ (polynomial in T_p), starting from $T_{p^0} = T_1 = I$ and $T_{p^1} = T_p$; e.g. $T_{p^2} = T_p T_p - p^{k-1} T_1 = T_p^2 - p^{k-1} I$. So if $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$, then, since the $p_i^{\alpha_i}$ are pairwise coprime, $T_n = T_{p_1^{\alpha_1}} T_{p_2^{\alpha_2}} \cdots T_{p_r^{\alpha_r}} = \prod_i (\text{polynomial in } T_{p_i})$. This proves the proposition. ■

Correction. In my handwritten notes the two cases were written as “ $p \mid n$ and $p \nmid n^2$ ” and “ $p \mid n$ and $p \mid n^2$ ”; the correct case distinction is $p \nmid \frac{n}{p}$ versus $p \mid \frac{n}{p}$.

8.5 Hecke eigenforms

Definition 8.5.1 (Hecke eigenform)

$f \in M_k(\text{SL}_2(\mathbb{Z}))$, $f \neq 0$, is called a *Hecke eigenform* if for every $n \geq 1$ there is $\lambda_n \in \mathbb{C}$ (eigenvalue) with $T_n(f) = \lambda_n f$.

Suppose $f(z) = \sum_{m \geq 0} a_m q^m$ is a Hecke eigenform. Then the first coefficient of $T_n(f)$ is a_n (Theorem 8.3.5), so by the definition, $a_n = \lambda_n a_1$ for all $n \geq 1$. If $a_1 = 0$, then $a_n = 0$ for all $n \geq 1$ and f is constant, which is impossible for $k > 0$; so $a_1 \neq 0$, and we normalise f such that $a_1 = 1$. Then $\lambda_n = a_n$, i.e. $T_n f = a_n f$ for all $n \geq 1$. This normalised f is called a *normalised Hecke eigenform*.

Proposition 8.5.2 (Behaviour of the Fourier coefficients)

If $f = \sum a_m q^m$ is a normalised Hecke eigenform, then

$$a_m a_n = \sum_{d|\gcd(m,n)} d^{k-1} a_{mn/d^2} \quad (m, n \geq 1).$$

In particular $a_m a_n = a_{mn}$ if $\gcd(m, n) = 1$ (multiplicativity), and $a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}$ for primes p (prime power recurrence). Hence f is uniquely determined by its prime Fourier coefficients a_p .

Proof. Since the normalised Hecke eigenform f satisfies $T_n f = a_n f$, we get $T_n(T_m f) = T_n(a_m f) = a_m(T_n f) = a_m a_n f$, so by Theorem 8.4.1

$$\sum_{d|\gcd(n,m)} d^{k-1} (T_{nm/d^2} f) = a_m a_n f \implies \left(\sum_{d|\gcd(n,m)} d^{k-1} a_{nm/d^2} \right) f = a_m a_n f,$$

and comparing (as $f \neq 0$) gives $a_m a_n = \sum_{d|\gcd(n,m)} d^{k-1} a_{nm/d^2}$. Multiplicativity: if $\gcd(n, m) = 1$, then $d = 1$, so $a_m a_n = a_{mn}$. Prime power recurrence: if $m = p$, $n = p^r$ ($r \geq 1$), then $\gcd(n, m) = p$, so $d \in \{1, p\}$, and $a_p a_{p^r} = a_{p^{r+1}} + p^{k-1} a_{p^{r-1}}$, i.e. $a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}$. As before, a_{p^r} is a polynomial in a_p , and with multiplicativity all a_n ($n \geq 1$) are determined by the a_p ; a_0 is then determined as well (for $k > 0$, by Theorem 8.3.5: $\sigma_{k-1}(n) a_0 = a_n a_0$ for all n , and if $a_0 \neq 0$ this forces $a_n = \sigma_{k-1}(n)$). ■

Example 8.5.3 (The discriminant). Let $\Delta_1 = (2\pi)^{-12} \Delta = \sum_{n \geq 1} \tau(n) q^n = q - 24q^2 + 252q^3 + \dots$ (Theorem 4.8.4); τ is *Ramanujan's tau function*. Since $\Delta_1 \in S_{12}(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C} \Delta_1$ is one-dimensional and $T_n : S_{12} \rightarrow S_{12}$ (Theorem 8.3.5), $T_n \Delta_1 \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$, so there exists $\lambda_n \in \mathbb{C}$ with $T_n \Delta_1 = \lambda_n \Delta_1$. Since $a_1 = \tau(1) = 1$, Δ_1 is a normalised Hecke eigenform: the first coefficient of the left-hand side is $\tau(n)$ (for $m = 1$, $\gcd(n, m) = 1$, so $d = 1$), and the first coefficient of the right-hand side is $\lambda_n \tau(1) = \lambda_n$. Hence $\lambda_n = \tau(n)$. Explicitly, $T_n \Delta_1 = \lambda_n \Delta_1$ reads (weight 12, so $d^{k-1} = d^{11}$)

$$\sum_{m \geq 1} \left(\sum_{d|\gcd(n,m)} d^{11} \tau\left(\frac{nm}{d^2}\right) \right) q^m = \sum_{m \geq 1} \lambda_n \tau(m) q^m \implies \tau(n) \tau(m) = \sum_{d|\gcd(n,m)} d^{11} \tau\left(\frac{nm}{d^2}\right).$$

In particular $\tau(mn) = \tau(m) \tau(n)$ for $\gcd(m, n) = 1$ and $\tau(p^{r+1}) = \tau(p) \tau(p^r) - p^{11} \tau(p^{r-1})$: the first two of Ramanujan's conjectures, proven by Mordell (1917).

Example 8.5.4 (Eisenstein series). For even $k \geq 4$ let

$$\tilde{G}_k(z) = -\frac{B_k}{2k} + \sum_{m \geq 1} \sigma_{k-1}(m) q^m = -\frac{B_k}{2k} E_k(z) \in M_k(\mathrm{SL}_2(\mathbb{Z})).$$

Note that its coefficients are $a_m = \sigma_{k-1}(m)$ ($m \geq 1$), and specifically $a_1 = \sigma_{k-1}(1) = 1$. We show $T_n(\tilde{G}_k) = \sigma_{k-1}(n) \tilde{G}_k$. The constant term of $T_n \tilde{G}_k$ is $\sigma_{k-1}(n) a_0$ (Theorem 8.3.5); for $m \geq 1$ we need

$$\sigma_{k-1}(n) \sigma_{k-1}(m) = \sum_{d|\gcd(n,m)} d^{k-1} \sigma_{k-1}\left(\frac{mn}{d^2}\right). \quad (\dagger)$$

It is known that σ_{k-1} is multiplicative, and both sides of $(\dagger)$ factor over the primes (a divisor d of $\gcd(n, m)$ is a choice of $v_p(d) \leq \min(v_p(n), v_p(m))$ for each p , and $\sigma_{k-1}(mn/d^2) = \prod_p \sigma_{k-1}(p^{v_p(mn/d^2)})$). So it suffices to treat prime powers $n = p^r$, $m = p^s$ (WLOG $r \leq s$, so $\gcd(p^r, p^s) = p^r$ and $d = p^j$, $j = 0, \dots, r$). Note that, with $X = p^{k-1}$,

$$\sigma_{k-1}(p^r) = \sum_{d|p^r} d^{k-1} = \sum_{i=0}^r (p^i)^{k-1} = \sum_{i=0}^r (p^{k-1})^i = \frac{X^{r+1} - 1}{X - 1}.$$

Thus

$$\begin{aligned}
 \sigma_{k-1}(p^r)\sigma_{k-1}(p^s) &= \left(\frac{X^{r+1}-1}{X-1}\right)\left(\frac{X^{s+1}-1}{X-1}\right) = \frac{X^{r+s+2}-X^{r+1}-X^{s+1}+1}{(X-1)^2} \\
 &= \frac{1}{X-1}\left(\frac{X^{r+s+2}-X^{s+1}}{X-1} - \frac{X^{r+1}-1}{X-1}\right) \\
 &= \frac{1}{X-1}\left(X^{s+1}\frac{X^{r+1}-1}{X-1} - \frac{X^{r+1}-1}{X-1}\right) \\
 &= \frac{1}{X-1}\left(X^{s+1}\sum_{j=0}^r X^j - \sum_{j=0}^r X^j\right) \\
 &= \frac{1}{X-1}\left(\sum_{j=0}^r X^{r+s+1-j} - \sum_{j=0}^r X^j\right) \\
 &= \frac{1}{X-1}\sum_{j=0}^r (X^{r+s+1-j} - X^j) = \sum_{j=0}^r X^j \frac{X^{(r+s-2j)+1}-1}{X-1} \\
 &= \sum_{j=0}^r (p^j)^{k-1} \sigma_{k-1}(p^{r+s-2j}) \\
 &= \sum_{d|\gcd(p^r, p^s)} d^{k-1} \sigma_{k-1}\left(\frac{p^{r+s}}{d^2}\right),
 \end{aligned}$$

where in the fifth line we reindexed $j \mapsto r-j$ in the first sum. So $(\dagger)$ holds for all $m, n \geq 1$, and $T_n \tilde{G}_k = \sigma_{k-1}(n) \tilde{G}_k$: the Eisenstein series is a normalised Hecke eigenform.

Correction. In my handwritten notes $\sigma_{k-1}(p^r)$ was written as $\sum_{i=1}^r (p^i)^{k-1}$; the sum starts at $i = 0$. Also the conclusion was stated “for any m, n where $\gcd(m, n) \neq 1$ ”; identity $(\dagger)$ holds for all m, n (for coprime m, n it is the multiplicativity of σ_{k-1}). Finally, the series called G_k in my notes is $\tilde{G}_k = \frac{(k-1)!}{2(2\pi i)^k} G_k$ in the notation of Chapter 4.

Chapter 9

The Hecke Algebra

To define Hecke operators for congruence subgroups, it is convenient to replace sums over sublattices by *double cosets* $\Gamma\alpha\Gamma$. In this chapter we develop the abstract Hecke algebra of a group Γ with respect to a semigroup Δ of its commensurator, following [5, Ch. 3] and [3, §2.7].

9.1 Commensurable subgroups and the commensurator

Let G be a group and Γ_1, Γ_2 subgroups of G .

Definition 9.1.1 (Commensurability)

Γ_1 and Γ_2 are *commensurable*, written $\Gamma_1 \approx \Gamma_2$, if

$$[\Gamma_1 : \Gamma_1 \cap \Gamma_2] < \infty \quad \text{and} \quad [\Gamma_2 : \Gamma_1 \cap \Gamma_2] < \infty.$$

Lemma 9.1.2

Commensurability is an equivalence relation.

Proof. Reflexivity and symmetry are clear. For transitivity we use: if $H, K \leq G$ and $A \leq H$, then $[A : A \cap K] \leq [H : H \cap K]$ (the map $A/(A \cap K) \rightarrow H/(H \cap K)$, $a(A \cap K) \mapsto a(H \cap K)$, is well defined and injective). Let $\Gamma_1 \approx \Gamma_2 \approx \Gamma_3$. Then

$$[\Gamma_1 : \Gamma_1 \cap \Gamma_2 \cap \Gamma_3] = [\Gamma_1 : \Gamma_1 \cap \Gamma_2] [\Gamma_1 \cap \Gamma_2 : (\Gamma_1 \cap \Gamma_2) \cap \Gamma_3] \leq [\Gamma_1 : \Gamma_1 \cap \Gamma_2] [\Gamma_2 : \Gamma_2 \cap \Gamma_3] < \infty,$$

and since $\Gamma_1 \cap \Gamma_2 \cap \Gamma_3 \subseteq \Gamma_1 \cap \Gamma_3 \subseteq \Gamma_1$, also $[\Gamma_1 : \Gamma_1 \cap \Gamma_3] < \infty$. By symmetry $[\Gamma_3 : \Gamma_1 \cap \Gamma_3] < \infty$. ■

Example 9.1.3. Any two congruence subgroups $\Gamma_1, \Gamma_2 \subseteq \mathrm{SL}_2(\mathbb{Z})$ are commensurable: $\Gamma(N_1) \subseteq \Gamma_1$ and $\Gamma(N_2) \subseteq \Gamma_2$, so

$$\mathrm{SL}_2(\mathbb{Z}) \supseteq \Gamma_1, \Gamma_2 \supseteq \Gamma_1 \cap \Gamma_2 \supseteq \Gamma(N_1) \cap \Gamma(N_2) \supseteq \Gamma(\mathrm{lcm}(N_1, N_2)),$$

and $\Gamma(\mathrm{lcm}(N_1, N_2))$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$ (Proposition 3.2.7); so is $\Gamma_1 \cap \Gamma_2$, in both Γ_1 and Γ_2 .

Definition 9.1.4 (Commensurator)

For $\Gamma \leq G$, the set $\tilde{\Gamma} = \{\alpha \in G : \Gamma \approx \alpha^{-1}\Gamma\alpha\}$ is called the *commensurator* of Γ in G . It is a subgroup of G (if $\Gamma \approx \alpha^{-1}\Gamma\alpha$ and $\Gamma \approx \beta^{-1}\Gamma\beta$, then conjugating the first by β gives $\beta^{-1}\Gamma\beta \approx (\alpha\beta)^{-1}\Gamma(\alpha\beta)$, and transitivity gives $\Gamma \approx (\alpha\beta)^{-1}\Gamma\alpha\beta$; conjugating by α^{-1} gives $\alpha\Gamma\alpha^{-1} \approx \Gamma$). Commensurable subgroups have the same commensurator.

Proposition 9.1.5

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and $G = \mathrm{GL}_2^+(\mathbb{R})$. Then $\tilde{\Gamma} = \mathbb{R}^\times \cdot \mathrm{GL}_2^+(\mathbb{Q})$ (the matrices $\lambda\beta$ with $\lambda \in \mathbb{R}^\times$, $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$).

$\supseteq$. If Γ is a congruence subgroup, then $\Gamma(n) \subseteq \Gamma$ for some n , where $\Gamma(n) = \mathrm{SL}_2(\mathbb{Z}) \cap (I + nM_{2 \times 2}(\mathbb{Z}))$. Now let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Multiplying α by a positive integer does not change $\alpha^{-1}\Gamma\alpha$, so WLOG $\alpha \in M_{2 \times 2}(\mathbb{Z})$. Let $d = \det(\alpha)$; then $d\alpha^{-1} = \mathrm{adj}(\alpha) \in M_{2 \times 2}(\mathbb{Z})$. Thus

$$\alpha\Gamma(nd)\alpha^{-1} \subseteq \alpha(I + ndM_{2 \times 2}(\mathbb{Z}))\alpha^{-1} = I + n\alpha M_{2 \times 2}(\mathbb{Z})(d\alpha^{-1}) \subseteq I + nM_{2 \times 2}(\mathbb{Z}),$$

and its elements have determinant 1 and integer entries, so $\alpha\Gamma(nd)\alpha^{-1} \subseteq \Gamma(n) \subseteq \Gamma$. So $\Gamma(nd) \subseteq \alpha^{-1}\Gamma\alpha$. On the other side, $\Gamma(nd) \subseteq \Gamma(n) \subseteq \Gamma$. Thus

$$\Gamma(nd) \subseteq \Gamma \cap \alpha^{-1}\Gamma\alpha \subseteq \Gamma \subseteq \mathrm{SL}_2(\mathbb{Z}),$$

and since $\Gamma(nd)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$, $\Gamma \cap \alpha^{-1}\Gamma\alpha$ has finite index in Γ . Replacing α by α^{-1} , there is d' with $\Gamma(nd') \subseteq \Gamma \cap \alpha\Gamma\alpha^{-1} \subseteq \Gamma$; thus

$$[\alpha^{-1}\Gamma\alpha : \Gamma \cap \alpha^{-1}\Gamma\alpha] = [\Gamma : \alpha\Gamma\alpha^{-1} \cap \Gamma] < \infty$$

(conjugating by α). So $\alpha^{-1}\Gamma\alpha \approx \Gamma$, and $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ commensurates Γ . Since scalar matrices commute with every element of $\mathrm{GL}_2(\mathbb{R})$, i.e. $(cI)^{-1}\gamma(cI) = \gamma$ for all $\gamma \in \Gamma$ and $c \in \mathbb{R}^\times$, we have $(cI)^{-1}\Gamma(cI) = \Gamma$: cI commensurates Γ as well, and so does $cI \cdot \alpha$. Hence $\mathbb{R}^\times \cdot \mathrm{GL}_2^+(\mathbb{Q}) \subseteq \tilde{\Gamma}$.

$[\subseteq]$ Let $\alpha \in \tilde{\Gamma}$, so $\alpha^{-1}\Gamma\alpha \approx \Gamma$.

Step 1: α maps cusps to cusps. Let $s \in \mathbb{Q} \cup \{\infty\}$, and choose a parabolic $\gamma \in \Gamma$, $\gamma \neq \pm I$, fixing s (e.g. $\sigma^{-1}T^N\sigma$ with $\sigma s = \infty$, which lies in $\Gamma(N) \subseteq \Gamma$). Since $\Gamma \cap \alpha\Gamma\alpha^{-1}$ has finite index in Γ , by the pigeonhole principle two of the powers $\gamma, \gamma^2, \dots$ lie in the same coset, so $\gamma^m \in \alpha\Gamma\alpha^{-1}$ for some $m \geq 1$. Then $\alpha^{-1}\gamma^m\alpha \in \Gamma$ is parabolic (same trace ± 2 , and $\neq \pm I$) and fixes $\alpha^{-1}s$. A parabolic $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ fixes ∞ if $c = 0$, and otherwise the root of $cz^2 + (d-a)z - b = 0$, whose discriminant is $(d-a)^2 + 4bc = (a+d)^2 - 4 = 0$; so the fixed point is $\frac{a-d}{2c} \in \mathbb{Q}$. Hence $\alpha^{-1}s \in \mathbb{Q} \cup \{\infty\}$. Since $\tilde{\Gamma}$ is a group, the same applies to α^{-1} , so $\alpha(\mathbb{Q} \cup \{\infty\}) = \mathbb{Q} \cup \{\infty\}$.

Step 2. Let $x_1 = \alpha\infty$, $x_2 = \alpha 0$, $x_3 = \alpha 1$; these are three distinct points of $\mathbb{Q} \cup \{\infty\}$. There is $\beta \in \mathrm{GL}_2(\mathbb{Q})$ with $\beta\infty = x_1$, $\beta 0 = x_2$, $\beta 1 = x_3$ (if all x_i are finite, $\beta = \begin{pmatrix} x_1u & x_2v \\ u & v \end{pmatrix}$ with rational $u, v \neq 0$ solving $u(x_1 - x_3) = v(x_3 - x_2)$; similarly if one x_i is ∞). Then $\beta^{-1}\alpha$ fixes $\infty, 0, 1$, so it is a scalar matrix λI (a Möbius transformation fixing three points is the identity), $\lambda \in \mathbb{R}^\times$ since α is real. Thus $\alpha = \lambda\beta$, and $\det \alpha = \lambda^2 \det \beta > 0$ gives $\det \beta > 0$, i.e. $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$. So $\alpha \in \mathbb{R}^\times \cdot \mathrm{GL}_2^+(\mathbb{Q})$. ■

Correction. In my handwritten notes the labels $[\subseteq]$ and $[\supseteq]$ were interchanged, and the second inclusion concluded from $\alpha\infty = \frac{a}{c}$, $\alpha 0 = \frac{b}{d} \in \mathbb{Q} \cup \{\infty\}$ directly that α is a real multiple of a rational matrix; two values do not suffice (they only fix the ratios $a : c$ and $b : d$), which is why the third point $\alpha 1$ is used above. That α maps cusps to cusps was justified by “ $\alpha^{-1}\Gamma\alpha$ has the same cusps as Γ ”; Step 1 proves this.

9.2 Double cosets

Proposition 9.2.1

Let $\alpha \in \tilde{\Gamma}$. Then $\Gamma\alpha\Gamma$ is a finite disjoint union of right cosets and of left cosets:

$$\Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \Gamma\alpha\gamma_i = \bigsqcup_{j=1}^e \delta_j\alpha\Gamma,$$

where $\Gamma = \bigsqcup_{i=1}^d (\Gamma \cap \alpha^{-1}\Gamma\alpha)\gamma_i$ and $\Gamma = \bigsqcup_{j=1}^e \delta_j(\Gamma \cap \alpha\Gamma\alpha^{-1})$, $d = [\Gamma : \Gamma \cap \alpha^{-1}\Gamma\alpha]$, $e = [\Gamma : \Gamma \cap \alpha\Gamma\alpha^{-1}]$.

Proof. Covering. Let $\gamma \in \Gamma$; write $\gamma = \delta\gamma_i$ with $\delta \in \Gamma \cap \alpha^{-1}\Gamma\alpha$. Then $\alpha\delta\alpha^{-1} \in \Gamma$, and

$$\Gamma\alpha\gamma = \Gamma\alpha\delta\gamma_i = \Gamma(\alpha\delta\alpha^{-1})\alpha\gamma_i = \Gamma\alpha\gamma_i.$$

Hence $\Gamma\alpha\Gamma = \bigcup_{\gamma} \Gamma\alpha\gamma = \bigcup_{i=1}^d \Gamma\alpha\gamma_i$.

Disjointness. Let $\Gamma\alpha\gamma_i = \Gamma\alpha\gamma_j$. Then $\alpha\gamma_i\gamma_j^{-1}\alpha^{-1} \in \Gamma$ (right coset property), so $\gamma_i\gamma_j^{-1} \in \alpha^{-1}\Gamma\alpha$, hence $\gamma_i\gamma_j^{-1} \in \Gamma \cap \alpha^{-1}\Gamma\alpha$, i.e. γ_i and γ_j lie in the same coset of $\Gamma \cap \alpha^{-1}\Gamma\alpha$, so $i = j$.

The statement for left cosets is proved in the same way. ■

Notes. My handwritten proof started from $\alpha^{-1}\Gamma\alpha\Gamma$ and used $\Gamma = \bigcup_i (\Gamma \cap \alpha^{-1}\Gamma\alpha)\gamma_i$ inside it; the direct computation above is the same idea. Commensurability is exactly what makes the double coset a *finite* union of single cosets.

9.3 The Hecke algebra

Definition 9.3.1 (Hecke module)

Let $\Gamma_1 \approx \Gamma_2$ and let Δ be a subsemigroup of G with $\Gamma_1, \Gamma_2 \subseteq \Delta \subseteq \tilde{\Gamma}_1 = \tilde{\Gamma}_2$. Let R be a commutative ring. We denote by

$$\mathcal{H}_R(\Gamma_1, \Gamma_2; \Delta) = \left\{ \sum_{\alpha} c_{\alpha} \Gamma_1 \alpha \Gamma_2 : c_{\alpha} \in R, c_{\alpha} = 0 \text{ for all but finitely many double cosets} \right\}$$

the free R -module with basis the double cosets $\Gamma_1 \alpha \Gamma_2$, $\alpha \in \Delta$, and call it the *Hecke module*. If $\Gamma_1 = \Gamma_2 = \Gamma$ we simply write $\mathcal{H}_R(\Gamma, \Delta)$.

Definition 9.3.2 (Multiplication)

Let $\Gamma_1 \approx \Gamma_2 \approx \Gamma_3$ and $\Gamma_1, \Gamma_2, \Gamma_3 \subseteq \Delta \subseteq \tilde{\Gamma}_i$. For $\alpha, \beta \in \Delta$ write $\Gamma_1 \alpha \Gamma_2 = \bigsqcup_i \Gamma_1 \alpha_i$ and $\Gamma_2 \beta \Gamma_3 = \bigsqcup_j \Gamma_2 \beta_j$ (finite, Proposition 9.2.1), and define

$$(\Gamma_1 \alpha \Gamma_2)(\Gamma_2 \beta \Gamma_3) = \sum_{\Gamma_1 \gamma \Gamma_3 \subseteq \Gamma_1 \alpha \Gamma_2 \beta \Gamma_3} c_{\gamma} \Gamma_1 \gamma \Gamma_3, \quad c_{\gamma} = \#\{(i, j) : \Gamma_1 \alpha_i \beta_j = \Gamma_1 \gamma\},$$

extended bilinearly to $\mathcal{H}_R(\Gamma_1, \Gamma_2; \Delta) \times \mathcal{H}_R(\Gamma_2, \Gamma_3; \Delta) \rightarrow \mathcal{H}_R(\Gamma_1, \Gamma_3; \Delta)$. For given α, β , there are only finitely many double cosets $\Gamma_1 \gamma \Gamma_3$ with $c_{\gamma} \neq 0$ (the products $\alpha_i \beta_j$ are finitely many).

Lemma 9.3.3

The multiplication is well defined: c_{γ} does not depend on the choice of the representatives α_i, β_j , nor on the choice of γ in $\Gamma_1 \gamma \Gamma_3$. Moreover

$$c_{\gamma} = \#\{\Gamma_2 y \subseteq \Gamma_2 \beta \Gamma_3 : \gamma y^{-1} \in \Gamma_1 \alpha \Gamma_2\} = \#\{x \Gamma_2 \subseteq \Gamma_1 \alpha \Gamma_2 : x^{-1} \gamma \in \Gamma_2 \beta \Gamma_3\}.$$

Proof. For fixed j , the condition $\Gamma_1 \alpha_i \beta_j = \Gamma_1 \gamma$ means $\gamma \beta_j^{-1} \in \Gamma_1 \alpha_i$; since the $\Gamma_1 \alpha_i$ partition $\Gamma_1 \alpha \Gamma_2$, for each j there is at most one such i , and there is one iff $\gamma \beta_j^{-1} \in \Gamma_1 \alpha \Gamma_2$. Thus $c_{\gamma} = \#\{j : \gamma \beta_j^{-1} \in \Gamma_1 \alpha \Gamma_2\}$, which is the first formula (replacing β_j by $\delta \beta_j$, $\delta \in \Gamma_2$, does not change

the condition, since $\gamma\beta_j^{-1}\delta^{-1} \in \Gamma_1\alpha\Gamma_2$ iff $\gamma\beta_j^{-1} \in \Gamma_1\alpha\Gamma_2$). This expression no longer involves the α_i , and does not depend on the representatives β_j . If $\gamma' = \delta_1\gamma\delta_3$ ($\delta_1 \in \Gamma_1$, $\delta_3 \in \Gamma_3$), then $\Gamma_2y \mapsto \Gamma_2y\delta_3$ is a bijection between the two sets, so $c_{\gamma'} = c_\gamma$.

For the second formula, the maps $\Gamma_2y \mapsto \gamma y^{-1}\Gamma_2$ and $x\Gamma_2 \mapsto \Gamma_2x^{-1}\gamma$ are well defined (replacing y by δy changes γy^{-1} to $\gamma y^{-1}\delta^{-1}$, in the same left coset) and inverse to each other; and $\gamma y^{-1} = x$ satisfies $x \in \Gamma_1\alpha\Gamma_2$ iff $y = x^{-1}\gamma \in \Gamma_2\beta\Gamma_3$ (given $\gamma = xy$). So the two sets are in bijection. ■

Theorem 9.3.4 (Hecke algebra)

$\mathcal{H}_R(\Gamma, \Delta)$ with the above multiplication is an associative R -algebra with unit element $e_{\mathcal{H}} = \Gamma e \Gamma = \Gamma$. It is called the *Hecke algebra* of Γ with respect to Δ .

Proof sketch. It is an R -module by definition, and the multiplication is R -bilinear by definition. The unit: $\Gamma = \Gamma e$ is a single coset, and $(\Gamma)(\Gamma\beta\Gamma)$ has $c_\gamma = \#\{j : \Gamma\beta_j = \Gamma\gamma\} = 1$ for $\gamma \in \Gamma\beta\Gamma$, so $(\Gamma)(\Gamma\beta\Gamma) = \Gamma\beta\Gamma$; similarly on the other side. Associativity: identify $\mathcal{H}$ with the space of finite formal sums of right cosets $\Gamma\xi$ which are invariant under right multiplication by Γ (via $\Gamma\alpha\Gamma \mapsto \sum_i \Gamma\alpha_i$); the multiplication is then $(\sum_i \Gamma\alpha_i)(\sum_j \Gamma\beta_j) = \sum_{i,j} \Gamma\alpha_i\beta_j$, which is visibly associative. See [5, Prop. 3.4] for details. ■

Proposition 9.3.5 (1) If $|\Gamma \backslash \Gamma\alpha\Gamma| = |\Gamma\alpha\Gamma/\Gamma|$, then $\Gamma\alpha\Gamma$ has a common set of right and left coset representatives, i.e. there are α_i with

$$\Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \Gamma\alpha_i = \bigsqcup_{i=1}^d \alpha_i\Gamma.$$

(2) Suppose there is an involution $\iota : \Delta \rightarrow \Delta$, $\alpha \mapsto \alpha^\iota$, such that for all $\alpha, \beta \in \Delta$:

$$\begin{aligned} (\alpha\beta)^\iota &= \beta^\iota\alpha^\iota \text{ (anti-homomorphism),} & (\alpha^\iota)^\iota &= \alpha \text{ (involution),} \\ \Gamma^\iota &= \Gamma \text{ (preserves the group),} & \Gamma\alpha^\iota\Gamma &= \Gamma\alpha\Gamma \text{ (preserves double cosets).} \end{aligned}$$

Then (i) for all $\alpha \in \Delta$, $\Gamma \backslash \Gamma\alpha\Gamma$ and $\Gamma\alpha\Gamma/\Gamma$ have a common set of representatives; and (ii) $\mathcal{H}_R(\Gamma, \Delta)$ is commutative.

Proof. (1) Let $\Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \Gamma\alpha_i = \bigsqcup_{j=1}^d \beta_j\Gamma$. We first show that every right coset meets every left coset. Suppose, by way of contradiction, that there are i_0, j_0 with $\Gamma\alpha_{i_0} \cap \beta_{j_0}\Gamma = \emptyset$. Since $\Gamma\alpha_{i_0} \subseteq \bigcup_{j=1}^d \beta_j\Gamma$ and $\Gamma\alpha_{i_0} \cap \beta_{j_0}\Gamma = \emptyset$, we get $\Gamma\alpha_{i_0} \subseteq \bigcup_{j \neq j_0} \beta_j\Gamma$. Multiplying both sides by Γ on the right,

$$\Gamma\alpha\Gamma = \Gamma\alpha_{i_0}\Gamma \subseteq \bigcup_{j \neq j_0} \beta_j\Gamma\Gamma = \bigcup_{j \neq j_0} \beta_j\Gamma,$$

since $\Gamma\alpha_{i_0}\Gamma$ generates all of $\Gamma\alpha\Gamma$ ($\alpha_{i_0} \in \Gamma\alpha\Gamma$). This contradicts $\Gamma\alpha\Gamma = \bigsqcup_{j=1}^d \beta_j\Gamma$ (the coset $\beta_{j_0}\Gamma$ is missing). Hence for every i we may choose $\alpha'_i \in \Gamma\alpha_i \cap \beta_i\Gamma$. Then $\Gamma\alpha'_i = \Gamma\alpha_i$ and $\alpha'_i\Gamma = \beta_i\Gamma$, so $\Gamma\alpha\Gamma = \bigsqcup_i \Gamma\alpha'_i = \bigsqcup_i \alpha'_i\Gamma$.

(2)(i) Let $\Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \Gamma\alpha_i$. Applying the involution ι to both sides (it is a bijection of Δ and reverses products),

$$(\Gamma\alpha\Gamma)^\iota = \bigsqcup_{i=1}^d \alpha_i^\iota\Gamma^\iota \implies \Gamma^\iota\alpha^\iota\Gamma^\iota = \bigsqcup_{i=1}^d \alpha_i^\iota\Gamma \implies \Gamma\alpha^\iota\Gamma = \bigsqcup_{i=1}^d \alpha_i^\iota\Gamma \implies \Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \alpha_i^\iota\Gamma,$$

using the assumptions $\Gamma^\iota = \Gamma$ and $\Gamma\alpha^\iota\Gamma = \Gamma\alpha\Gamma$. So $|\Gamma\alpha\Gamma/\Gamma| = |\Gamma \backslash \Gamma\alpha\Gamma| = d$, and by (1) a common set of representatives exists.

(2)(ii) Let $X = \Gamma\alpha\Gamma$ and $Y = \Gamma\beta\Gamma$; it suffices to show $XY = YX$. Since $\Gamma\gamma^t\Gamma = \Gamma\gamma\Gamma$, the coefficient of $\Gamma\gamma\Gamma$ in YX may be computed at the representative γ^t (Lemma 9.3.3). By the first formula of Lemma 9.3.3,

$$c_{\gamma^t}(YX) = \#\{\Gamma x \subseteq X : \gamma^t x^{-1} \in Y\}.$$

Applying ι : since $(\gamma^t x^{-1})x = \gamma^t$, the anti-homomorphism property gives $x^t(\gamma^t x^{-1})^t = \gamma$, i.e. $(\gamma^t x^{-1})^t = (x^t)^{-1}\gamma$ (inverses taken in G); and $Y^t = \Gamma\beta^t\Gamma = Y$; also $\Gamma x \subseteq X$ iff $x^t\Gamma \subseteq X^t = X$, and $\Gamma x \mapsto x^t\Gamma$ is a bijection between right and left cosets in X . Hence

$$c_{\gamma^t}(YX) = \#\{x^t\Gamma \subseteq X : x'^{-1}\gamma \in Y\} = c_{\gamma}(XY)$$

by the second formula of Lemma 9.3.3. So XY and YX have the same coefficients, i.e. $XY = YX$. By bilinearity $\mathcal{H}_R(\Gamma, \Delta)$ is commutative. ■

Correction. In my handwritten notes Proposition 9.3.5(1) was annotated “not generally” at the hypothesis: indeed, the numbers of right and left cosets in $\Gamma\alpha\Gamma$ need not agree in general, which is why it is assumed. The proof in my notes showed that every right coset meets every left coset; the final choice $\alpha'_i \in \Gamma\alpha_i \cap \beta_i\Gamma$ completes it. Part (2)(ii) was left as “Open Notes!”; the proof above is the standard argument of [5, Prop. 3.8].

Chapter 10

Hecke Operators for $\Gamma_0(N)$

In this final chapter the Hecke algebra of Chapter 9 is made to act on modular forms with character, and specialised to $\Gamma_0(N)$: we define the Hecke operators T_n and T_n^* , show that the Hecke algebra of $\Gamma_0(N)$ is commutative, decompose the basic double cosets into single cosets, and compute the action of T_p on q -expansions, which generalises Proposition 8.3.6. We follow [3, §4.5] and [5, Ch. 3].

10.1 The Hecke algebra acts on $M_k(\Gamma, \chi)$

Definition 10.1.1

Let Γ be a congruence subgroup and Δ a subsemigroup with $\Gamma \subseteq \Delta \subseteq \tilde{\Gamma} = \mathbb{R}^\times \cdot \mathrm{GL}_2^+(\mathbb{Q})$ (Proposition 9.1.5). Let $\chi : \Delta \rightarrow \mathbb{C}^\times$ be a multiplicative map ($\chi(\alpha\beta) = \chi(\alpha)\chi(\beta)$) whose restriction to Γ is a group character of finite order (so its values on Γ are roots of unity, and $\chi(\gamma)^{-1} = \overline{\chi(\gamma)}$). Let

$$M_k(\Gamma, \chi) = \{f \in M_k(\Gamma \cap \ker \chi) : f|_k \gamma = \chi(\gamma)f \text{ for all } \gamma \in \Gamma\},$$

i.e. f holomorphic on $\mathfrak{H}$ and at the cusps with $f|_k \gamma = \chi(\gamma)f$; similarly $S_k(\Gamma, \chi)$ (vanishing at the cusps). For $\alpha \in \Delta$ with $\Gamma\alpha\Gamma = \bigsqcup_{i=1}^d \Gamma\alpha_i$ ($\alpha_i \in \Delta$) we define the *slash operator of the double coset*

$$f|_k[\Gamma\alpha\Gamma] = (\det \alpha)^{k/2-1} \sum_{i=1}^d \overline{\chi(\alpha_i)} f|_k \alpha_i,$$

and extend it linearly to $\mathcal{H}_{\mathbb{Z}}(\Gamma, \Delta)$.

Theorem 10.1.2 (1) The definition is independent of the choice of the representatives α_i .

(2) If $f \in M_k(\Gamma, \chi)$, then $f|_k[\Gamma\alpha\Gamma] \in M_k(\Gamma, \chi)$; similarly for $S_k(\Gamma, \chi)$.

Proof. (1) Note that $\det \alpha_i = \det \alpha$ for all i (elements of Γ have determinant 1). For any $\beta \in \Gamma$ and $f \in M_k(\Gamma, \chi)$, let α_i be a representative and $\beta\alpha_i$ another representative of the same coset $\Gamma\alpha_i$. Then

$$\overline{\chi(\beta\alpha_i)} f|_k \beta\alpha_i = \overline{\chi(\beta)\chi(\alpha_i)} (f|_k \beta)|_k \alpha_i = \overline{\chi(\beta)} \overline{\chi(\alpha_i)} \chi(\beta) f|_k \alpha_i = \overline{\chi(\alpha_i)} f|_k \alpha_i,$$

since $\overline{\chi(\beta)}\chi(\beta) = |\chi(\beta)|^2 = 1$. So each summand is independent of the representative.

(2) *Transformation law.* Apply $|_k \gamma$, $\gamma \in \Gamma$, to both sides of the definition:

$$\begin{aligned} (f|_k[\Gamma\alpha\Gamma])|_k \gamma &= (\det \alpha)^{k/2-1} \sum_{i=1}^d \overline{\chi(\alpha_i)} f|_k \alpha_i \gamma = (\det \alpha)^{k/2-1} \chi(\gamma) \overline{\chi(\gamma)} \sum_{i=1}^d \overline{\chi(\alpha_i)} f|_k \alpha_i \gamma \\ &= \chi(\gamma) (\det \alpha)^{k/2-1} \sum_{i=1}^d \overline{\chi(\alpha_i \gamma)} f|_k \alpha_i \gamma. \end{aligned}$$

Notice that $\{\alpha_i \gamma\}$ is again a set of representatives of $\Gamma \backslash \Gamma\alpha\Gamma$: right multiplication by γ maps $\Gamma\alpha_i$ into $\Gamma\alpha\Gamma\gamma = \Gamma\alpha\Gamma$, and $\Gamma\alpha_i\gamma = \Gamma\alpha_j\gamma$ implies $\Gamma\alpha_i = \Gamma\alpha_j$; so it permutes the d cosets. By (1), the last sum is again $f|_k[\Gamma\alpha\Gamma]$, so $(f|_k[\Gamma\alpha\Gamma])|_k \gamma = \chi(\gamma) f|_k[\Gamma\alpha\Gamma]$.

Holomorphy on $\mathfrak{H}$ and at the cusps follows since each $f|_k \alpha_i$ is holomorphic on $\mathfrak{H}$ and at all cusps (Proposition 7.1.2; a positive real scalar acts trivially by Remark 6.1.3, and $\alpha_i \in \mathbb{R}^\times \mathrm{GL}_2^+(\mathbb{Q})$), and the same holds for vanishing at the cusps. ■

Correction. In my handwritten definition the coefficient was written as $\chi(\alpha_i)$ and in the proof as $\chi(\alpha_i)$; the conjugate is the correct one (it is what makes (1) work). The question mark at “ $\{\alpha\gamma_i\gamma\}$ is also a set of representatives” is answered by the argument above (right multiplication permutes the right cosets in $\Gamma\alpha\Gamma$). The condition “ $\chi(\alpha^{-1}\gamma\alpha) = \chi(\gamma)$ ” in my notes is replaced here by the multiplicativity of χ on Δ , which is what is used.

10.2 The Hecke algebras of $\Gamma_0(N)$

Definition 10.2.1

We define the semigroups

$$\Delta_0(N) = \left\{ \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2 \times 2}(\mathbb{Z}) : \gcd(a, N) = 1, c \equiv 0 \pmod{N}, \det \alpha > 0 \right\},$$

$$\Delta_0^*(N) = \left\{ \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2 \times 2}(\mathbb{Z}) : \gcd(d, N) = 1, c \equiv 0 \pmod{N}, \det \alpha > 0 \right\}.$$

They are closed under multiplication: for α_1, α_2 in $\Delta_0(N)$ the product has lower left entry $c_1 a_2 + d_1 c_2 \equiv 0$ and upper left entry $a_1 a_2 + b_1 c_2 \equiv a_1 a_2 \pmod{N}$, coprime to N ; similarly the lower right entry $c_1 b_2 + d_1 d_2 \equiv d_1 d_2$ for $\Delta_0^*(N)$. Clearly $\Gamma_0(N) \subseteq \Delta_0(N)$ and $\Gamma_0(N) \subseteq \Delta_0^*(N)$ (for $\gamma \in \Gamma_0(N)$, $ad \equiv 1 \pmod{N}$, so a and d are coprime to N , and $\det \gamma = 1$), and $\Delta_0(N), \Delta_0^*(N) \subseteq \mathrm{GL}_2^+(\mathbb{Q}) \subseteq \widetilde{\Gamma_0(N)}$. Now define the Hecke algebras

$$\mathcal{H}_N = \mathcal{H}_{\mathbb{Z}}(\Gamma_0(N), \Delta_0(N)) = \left\{ \sum_{\alpha \in \Delta_0(N)} c_\alpha \Gamma_0(N) \alpha \Gamma_0(N) : c_\alpha \in \mathbb{Z}, c_\alpha = 0 \text{ for almost all } \alpha \right\},$$

$$\mathcal{H}_N^* = \mathcal{H}_{\mathbb{Z}}(\Gamma_0(N), \Delta_0^*(N)),$$

generated by the double cosets $\Gamma_0(N) \alpha \Gamma_0(N)$ for $\alpha \in \Delta_0(N)$, resp. $\alpha \in \Delta_0^*(N)$.

Characters. Let χ be a Dirichlet character modulo N . We extend $\gamma \mapsto \chi(d)$ (the Nebentypus of Chapter 7) to the semigroups by

$$\chi(\alpha) = \overline{\chi(a)} \quad (\alpha \in \Delta_0(N)), \quad \chi^*(\alpha) = \chi(d) \quad (\alpha \in \Delta_0^*(N)).$$

Both are multiplicative (by the congruences above), and for $\gamma \in \Gamma_0(N)$, $\overline{\chi(a)} = \chi(a)^{-1} = \chi(d)$ since $ad \equiv 1 \pmod{N}$; so both restrict to the Nebentypus on $\Gamma_0(N)$, and $M_k(\Gamma_0(N), \chi) = M_k(N, \chi)$ in the notation of Definition 10.1.1.

Lemma 10.2.2 (☺) (1) If $\alpha \in \Delta_0(N)$, there are unique positive integers l, m with $l \mid m$, $\gcd(l, N) = 1$ and

$$\Gamma_0(N) \alpha \Gamma_0(N) = \Gamma_0(N) \begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix} \Gamma_0(N);$$

here l is the gcd of the entries of α and $lm = \det \alpha$.

(2) If $l \mid m$ and $\gcd(lm, N) = 1$, then $\Gamma_0(N) \begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix} \Gamma_0(N) = \Gamma_0(N) \begin{pmatrix} m & 0 \\ 0 & l \end{pmatrix} \Gamma_0(N)$.

(3) If $\alpha \in \Delta_0^*(N)$, there are unique $l \mid m$, $\gcd(l, N) = 1$, with $\Gamma_0(N)\alpha\Gamma_0(N) = \Gamma_0(N) \begin{pmatrix} m & 0 \\ 0 & l \end{pmatrix} \Gamma_0(N)$.

Proof. Write $\Gamma_0 = \Gamma_0(N)$. *Uniqueness.* For $\gamma_1, \gamma_2 \in \mathrm{SL}_2(\mathbb{Z})$, the matrix $\gamma_1\alpha\gamma_2$ has the same determinant as α and the same gcd of entries (the entries of $\gamma_1\alpha\gamma_2$ are integer combinations of those of α , and conversely as γ_i^{-1} are integral). Since $\mathrm{diag}(l, m)$ with $l \mid m$ has entry-gcd l and determinant lm , l and m are determined by the double coset.

(1) *Reduction to entry-gcd 1.* Let l be the gcd of the entries of $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Since $l \mid a$ and $\gcd(a, N) = 1$, $\gcd(l, N) = 1$; and $\alpha' = \alpha/l$ has integer entries, upper left entry coprime to N , and lower left entry $c/l \equiv 0 \pmod{N}$ (as $N \mid c$ and $\gcd(l, N) = 1$). So $\alpha' \in \Delta_0(N)$ has entry-gcd 1, and $\det \alpha' = n/l^2$ where $n = \det \alpha$. If we show $\Gamma_0\alpha'\Gamma_0 = \Gamma_0 \mathrm{diag}(1, n/l^2)\Gamma_0$, then multiplying by the central matrix lI gives $\Gamma_0\alpha\Gamma_0 = \Gamma_0 \mathrm{diag}(l, n/l)\Gamma_0$, with $m = n/l$ and $l \mid m$ (as $l^2 \mid n$).

The case of entry-gcd 1. Let $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Delta_0(N)$ with $\gcd(a, b, c, d) = 1$ and $\det \alpha = n$.

Step 1: upper triangular form. Write $c = Nc'$ and $g = \gcd(a, c')$, and put $z = c'/g$, $w = -a/g$. Then $\gcd(z, w) = 1$ and $\gcd(w, N) = 1$ (as $\gcd(a, N) = 1$), so $\gcd(Nz, w) = 1$, and by Bézout there are x, y with $xw - yNz = 1$. Then $\gamma_1 = \begin{pmatrix} x & y \\ Nz & w \end{pmatrix} \in \Gamma_0$, and the lower left entry of $\gamma_1\alpha$ is $Nza + wc = N(\frac{c'}{g}a - \frac{a}{g}c') = 0$. So $\gamma_1\alpha = \begin{pmatrix} a_1 & b_1 \\ 0 & d_1 \end{pmatrix}$ with $a_1d_1 = n$; it still lies in $\Delta_0(N)$ and has entry-gcd 1. Replacing α by $\gamma_1\alpha$ we may assume $c = 0$, so $\alpha = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$, $ad = n$, $\gcd(a, N) = 1$, $\gcd(a, b, d) = 1$.

Step 2: a good first column. We choose an integer x with $\gcd(x, N) = 1$ and $\gcd(ax + bN, dN) = 1$, by the Chinese remainder theorem, imposing for each prime $p \mid dN$:

- if $p \mid N$: $x \not\equiv 0 \pmod{p}$; then $ax + bN \equiv ax \not\equiv 0 \pmod{p}$, since $p \nmid a$;
- if $p \nmid N$, $p \mid d$ and $p \nmid a$: $x \not\equiv -bNa^{-1} \pmod{p}$; then $ax + bN \not\equiv 0 \pmod{p}$;
- if $p \nmid N$, $p \mid d$ and $p \mid a$: then $p \nmid b$ (as $\gcd(a, b, d) = 1$), so $ax + bN \equiv bN \not\equiv 0 \pmod{p}$ for every x ; no condition.

Each condition excludes one residue class modulo p , so a solution x exists. Since $\gcd(x, N) = 1$, there is $\gamma_2 = \begin{pmatrix} x & * \\ N & * \end{pmatrix} \in \Gamma_0$ (complete (x, N) to a matrix of determinant 1 by Bézout). The first column of $\alpha\gamma_2$ is $\begin{pmatrix} ax+bN \\ dN \end{pmatrix} =: \begin{pmatrix} u \\ v \end{pmatrix}$, with $\gcd(u, v) = 1$ and $N \mid v$.

Step 3. Since $\gcd(u, v) = 1$, there are r, s with $us - rv = 1$, and $\gamma_3^{-1} := \begin{pmatrix} u & r \\ v & s \end{pmatrix} \in \Gamma_0$ (as $N \mid v$). Then $\gamma_3\alpha\gamma_2$ has first column $\gamma_3 \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, so $\gamma_3\alpha\gamma_2 = \begin{pmatrix} 1 & e \\ 0 & n \end{pmatrix}$ (its determinant is n), and $\begin{pmatrix} 1 & e \\ 0 & n \end{pmatrix} T^{-e} = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}$ with $T^{-e} \in \Gamma_0$. Hence $\Gamma_0\alpha\Gamma_0 = \Gamma_0 \mathrm{diag}(1, n)\Gamma_0$.

(2) The matrix $\mathrm{diag}(m, l)$ lies in $\Delta_0(N)$ (as $\gcd(m, N) = 1$), has determinant lm and entry-gcd $\gcd(l, m) = l$. By (1), its double coset is $\Gamma_0 \mathrm{diag}(l, m)\Gamma_0$.

(3) Let $w_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$, which normalises Γ_0 (Report Problem 7.4.3). For $\alpha \in \Delta_0^*(N)$, $w_N \alpha w_N^{-1} = \begin{pmatrix} d & -c/N \\ -Nb & a \end{pmatrix} \in \Delta_0(N)$ (same computation as in that problem), so by (1) $\Gamma_0 w_N \alpha w_N^{-1} \Gamma_0 = \Gamma_0 \text{diag}(l, m) \Gamma_0$. Conjugating back with w_N and using $w_N^{-1} \text{diag}(l, m) w_N = \text{diag}(m, l)$ (direct computation) gives $\Gamma_0 \alpha \Gamma_0 = \Gamma_0 \text{diag}(m, l) \Gamma_0$. ■

Notes. My handwritten proof of (1) used lattices: with $L = \mathbb{Z} \oplus \mathbb{Z}$ and $L_0 = \mathbb{Z} \oplus N\mathbb{Z}$ (columns $\begin{pmatrix} x \\ Ny \end{pmatrix}$), one has $\alpha L_0 \subseteq L_0$ for $\alpha \in \Delta_0(N)$, and $[L : \alpha L_0] = nN$; then the elementary divisor theorem should give bases adapted to $L \supseteq \alpha L$ and $L_0 \supseteq \alpha L_0$ simultaneously. The existence of such simultaneous bases needs an extra argument; the matrix reduction above avoids it. The underlying fact is that $\Gamma_0(N)$ is exactly the stabiliser of L_0 in $\text{SL}_2(\mathbb{Z})$.

Theorem 10.2.3

$\mathcal{H}_N$ and $\mathcal{H}_N^*$ are commutative.

Proof. Consider $\iota : \Delta_0(N) \rightarrow \Delta_0(N)$,

$$\alpha^\iota = \iota \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & c/N \\ bN & d \end{pmatrix} = W \alpha^\top W^{-1}, \quad W = \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix}.$$

It maps $\Delta_0(N)$ to itself ($c/N \in \mathbb{Z}$, upper left entry a , lower left entry $bN \equiv 0$, same determinant). We need to show that ι satisfies the four properties of Proposition 9.3.5(2). It is clear that $(\alpha\beta)^\iota = W\beta^\top \alpha^\top W^{-1} = (W\beta^\top W^{-1})(W\alpha^\top W^{-1}) = \beta^\iota \alpha^\iota$ and $(\alpha^\iota)^\iota = W(W\alpha^\top W^{-1})^\top W^{-1} = WW^{-1} \alpha W W^{-1} = \alpha$ (as W is diagonal), by matrix multiplication. For $\Gamma_0(N)^\iota = \Gamma_0(N)$: if $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, then $c \equiv 0 \pmod{N}$ so $c/N \in \mathbb{Z}$; also $bN \equiv 0 \pmod{N}$ and $\det \gamma^\iota = 1$; thus $\gamma^\iota \in \Gamma_0(N)$, and equality follows since ι is an involution. For the last property: by Lemma 10.2.2, any $\Gamma_0(N) \alpha \Gamma_0(N)$ can be represented by a diagonal matrix $\delta = \text{diag}(l, m)$, where $\delta^\iota = \delta$. So $\Gamma_0(N) \alpha^\iota \Gamma_0(N) = (\Gamma_0(N) \alpha \Gamma_0(N))^\iota = (\Gamma_0(N) \delta \Gamma_0(N))^\iota = \Gamma_0(N) \delta^\iota \Gamma_0(N) = \Gamma_0(N) \alpha \Gamma_0(N)$. By Proposition 9.3.5(2)(ii), $\mathcal{H}_N$ is commutative. The same ι preserves $\Delta_0^*(N)$ (it fixes d), and Lemma 10.2.2(3) gives diagonal representatives, so $\mathcal{H}_N^*$ is commutative too. ■

10.3 The Hecke operators T_n and T_n^*

Definition 10.3.1

For $l \mid m$ with $\gcd(l, N) = 1$ let

$$T_{l,m} = \Gamma_0(N) \begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix} \Gamma_0(N) \in \mathcal{H}_N, \quad T_{m,l}^* = \Gamma_0(N) \begin{pmatrix} m & 0 \\ 0 & l \end{pmatrix} \Gamma_0(N) \in \mathcal{H}_N^*.$$

For $n \geq 1$ define the Hecke operators as the sums of all *distinct* double cosets of determinant n :

$$T_n = \sum_{\substack{\Gamma_0(N) \alpha \Gamma_0(N) \\ \alpha \in \Delta_0(N), \det \alpha = n}} \Gamma_0(N) \alpha \Gamma_0(N) = \sum_{\substack{lm=n, l \mid m \\ \gcd(l, N)=1}} T_{l,m} \in \mathcal{H}_N, \quad T_n^* = \sum_{\substack{lm=n, l \mid m \\ \gcd(l, N)=1}} T_{m,l}^* \in \mathcal{H}_N^*,$$

the equalities by Lemma 10.2.2. By Lemma 10.2.2, the $T_{l,m}$ form a $\mathbb{Z}$ -basis of $\mathcal{H}_N$.

Correction. In my handwritten notes T_n was written as $\sum_{\alpha \in \Delta_0(N), \det \alpha = n} \Gamma_0(N) \alpha \Gamma_0(N)$, which literally sums over infinitely many α ; the sum is over the distinct double cosets.

Proposition 10.3.2 (1) For every prime p : $T_p = T_{1,p}$ (and $T_p^* = T_{p,1}^*$).

(2) If p is prime and $p \mid N$, then $T_{p^e} = T_{1,p^e}$.

(3) If p is prime and $p \nmid N$, then $T_{p^2} = T_{1,p^2} + T_{p,p}$.

(4) If $\gcd(n, N) = 1$, then $T_{n,n} = \Gamma_0(N) \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix}$ (a single coset), and $T_{n,n} T_{l,m} = T_{nl,nm}$.

Proof. Recall $T_n = \sum_{\gcd(l,N)=1, l \mid m, lm=n} T_{l,m}$.

(1) For $n = p$, $lm = p$, so the only possible pairs (l, m) are $(1, p)$ and $(p, 1)$. If $(l, m) = (1, p)$, then $1 \mid p$ and $\gcd(1, N) = 1$, so $T_{1,p}$ is counted. If $(l, m) = (p, 1)$, it is not true that $p \mid 1$, so $T_{p,1}$ is not counted. Thus $T_p = T_{1,p}$ (analogously $T_p^* = T_{p,1}^*$).

(2) Let $n = p^e$; the pairs with $lm = p^e$ are $(l, m) = (p^r, p^s)$ with $r + s = e$, and $r \leq s$ (to satisfy $l \mid m$). Note that we need $\gcd(l, N) = \gcd(p^r, N) = 1$. Since $p \mid N$ (by assumption), for $r \geq 1$ we have $\gcd(p^r, N) \geq p \neq 1$, which cannot be; so only $r = 0$, implying $s = e$, and $(l, m) = (p^0, p^e) = (1, p^e)$. Thus $T_{p^e} = T_{1,p^e}$.

(3) For $n = p^2$: all possible pairs (l, m) with $lm = p^2$ are $(1, p^2)$, (p, p) , $(p^2, 1)$; since $\gcd(p, N) = 1$ by assumption and l is a power of p , $\gcd(l, N) = 1$ for all pairs. If $(l, m) = (1, p^2)$, then $1 \mid p^2$ is true, so this case is counted; if $(l, m) = (p, p)$, then $p \mid p$ is true, so this case is also counted; if $(l, m) = (p^2, 1)$, then $p^2 \mid 1$ is false, so this case is not counted. Thus $T_{p^2} = T_{1,p^2} + T_{p,p}$.

(4) Since $\begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} = nI$ and for any $\gamma \in \Gamma_0(N)$, $(nI)\gamma = \gamma(nI)$ (they commute), we have $\Gamma_0(N)(nI)\Gamma_0(N) = \Gamma_0(N)(nI)$; and $T_{n,n}$ is a legitimate basis element since $n \mid n$ and $\gcd(n, N) = 1$ (from the assumption). Furthermore, writing $T_{l,m} = \bigsqcup_j \Gamma_0(N)\beta_j$, the product $T_{n,n}T_{l,m}$ is computed from the products $(nI)\beta_j$, which represent pairwise distinct cosets whose union is $\Gamma_0(N)(nI)\begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix}\Gamma_0(N)$; hence every coefficient c_γ equals 1 and

$$T_{n,n}T_{l,m} = \Gamma_0(N) \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix} \Gamma_0(N) = \Gamma_0(N) \begin{pmatrix} nl & 0 \\ 0 & nm \end{pmatrix} \Gamma_0(N) = T_{nl,nm}.$$

Analogously for $T_{n,n}^*$. ■

Theorem 10.3.3

Let $f \in M_k(N, \chi)$, where $T_{l,m}$ acts via χ on $\Delta_0(N)$ and $T_{m,l}^*$ via χ^* on $\Delta_0^*(N)$. If $\gcd(lm, N) = 1$, then

$$f|_k T_{m,l}^* = \overline{\chi(lm)} f|_k T_{l,m}.$$

Furthermore, if $\gcd(n, N) = 1$, then $f|_k T_n^* = \overline{\chi(n)} f|_k T_n$.

Proof. Since $\gcd(lm, N) = 1$, from Lemma 10.2.2(2) we have

$$\Gamma_0(N) \begin{pmatrix} l & 0 \\ 0 & m \end{pmatrix} \Gamma_0(N) = \Gamma_0(N) \begin{pmatrix} m & 0 \\ 0 & l \end{pmatrix} \Gamma_0(N) = \bigsqcup_{i=1}^d \Gamma_0(N) \alpha_i, \quad (*)$$

for α_i in this double coset. Each $\alpha_i = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has $c \equiv 0 \pmod{N}$ and $\det \alpha_i = lm$, so $ad \equiv lm \pmod{N}$ is a unit; hence $\gcd(a, N) = \gcd(d, N) = 1$ and $\alpha_i \in \Delta_0(N) \cap \Delta_0^*(N)$. Thus

$$\chi(lm) = \chi(\det \alpha_i) = \chi(ad - bc) = \chi(ad) = \chi(a)\chi(d).$$

By definition $\chi(\alpha_i) = \overline{\chi(a)}$, so $\chi(a) = \overline{\chi(\alpha_i)}$, and $\chi^*(\alpha_i) = \chi(d)$. Therefore $\chi(lm) = \overline{\chi(\alpha_i)}\chi^*(\alpha_i)$, i.e. $\chi^*(\alpha_i) = \chi(lm)\chi(\alpha_i)$ (multiplying by $\chi(\alpha_i)$, $|\chi(\alpha_i)| = 1$), and so $\overline{\chi^*(\alpha_i)} = \overline{\chi(lm)}\overline{\chi(\alpha_i)}$. Since $\det \alpha_i = lm$,

$$\begin{aligned} f|_k T_{m,l}^* &= (lm)^{k/2-1} \sum_{i=1}^d \overline{\chi^*(\alpha_i)} f|_k \alpha_i = (lm)^{k/2-1} \sum_{i=1}^d \overline{\chi(lm)} \overline{\chi(\alpha_i)} f|_k \alpha_i \\ &= \overline{\chi(lm)} \left[(lm)^{k/2-1} \sum_{i=1}^d \overline{\chi(\alpha_i)} f|_k \alpha_i \right] = \overline{\chi(lm)} f|_k T_{l,m}. \end{aligned}$$

If $\gcd(n, N) = 1$, every pair (l, m) with $lm = n$ has $\gcd(lm, N) = 1$, and summing over the pairs gives $f|_k T_n^* = \overline{\chi(n)} f|_k T_n$. ■

10.4 Decomposition into single cosets

Lemma 10.4.1

Let $\gcd(n, N) = 1$. Then the map $\Gamma_0(N)\alpha \mapsto \mathrm{SL}_2(\mathbb{Z})\alpha$ is a bijection from $\Gamma_0(N) \setminus \{\alpha \in \Delta_0(N) : \det \alpha = n\}$ onto $\mathrm{SL}_2(\mathbb{Z}) \setminus \mathcal{M}_n$. Consequently, the set $\mathcal{R}_n$ of Proposition 8.3.4 is a complete set of representatives of $\Gamma_0(N) \setminus \{\alpha \in \Delta_0(N) : \det \alpha = n\}$.

Proof. Injective. If $\alpha, \alpha' \in \Delta_0(N)$, $\det = n$, and $\alpha' = \gamma\alpha$ with $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, then $n\gamma = \alpha' \mathrm{adj}(\alpha)$. Modulo N , both α' and $\mathrm{adj}(\alpha) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ are upper triangular, so $n\gamma$ is upper triangular modulo N ; as $\gcd(n, N) = 1$, the lower left entry of γ is $\equiv 0 \pmod{N}$, i.e. $\gamma \in \Gamma_0(N)$.

Surjective. Let $\beta \in \mathcal{M}_n$. Modulo N , β is invertible (its determinant n is a unit). Let $u = n \bmod N$ and $g = \begin{pmatrix} 1 & 0 \\ 0 & u \end{pmatrix} \bar{\beta}^{-1} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ (determinant $u \cdot u^{-1} = 1$). By Proposition 3.2.6 there

is $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma \equiv g \pmod{N}$. Then $\gamma\beta \equiv \begin{pmatrix} 1 & 0 \\ 0 & u \end{pmatrix} \pmod{N}$, so $\gamma\beta \in \Delta_0(N)$ and $\mathrm{SL}_2(\mathbb{Z})\beta = \mathrm{SL}_2(\mathbb{Z})\gamma\beta$.

Finally, every element of $\mathcal{R}_n$ is upper triangular with upper left entry $a \mid n$, hence coprime to N , so $\mathcal{R}_n \subseteq \Delta_0(N)$. ■

Proposition 10.4.2

Let p be a prime and $e \geq 1$. Then $T_{1,p^e} = \Gamma_0(N) \begin{pmatrix} 1 & 0 \\ 0 & p^e \end{pmatrix} \Gamma_0(N) = \bigsqcup_{\nu} \Gamma_0(N)\alpha_{\nu}$, where

$$\alpha_{\nu} \text{ runs over } \begin{cases} \begin{pmatrix} 1 & m \\ 0 & p^e \end{pmatrix}, & 0 \leq m < p^e, & \text{if } p \mid N, \\ \begin{pmatrix} p^{e-f} & m \\ 0 & p^f \end{pmatrix}, & 0 \leq f \leq e, \ 0 \leq m < p^f, \ \gcd(p^{e-f}, m, p^f) = 1, & \text{if } p \nmid N. \end{cases}$$

In particular $T_{1,p} = \bigsqcup_{m=0}^{p-1} \Gamma_0(N) \begin{pmatrix} 1 & m \\ 0 & p \end{pmatrix} \sqcup \Gamma_0(N) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$ if $p \nmid N$ ($p+1$ cosets), and $T_{1,p} = \bigsqcup_{m=0}^{p-1} \Gamma_0(N) \begin{pmatrix} 1 & m \\ 0 & p \end{pmatrix}$ if $p \mid N$ (p cosets).

Proof. By Lemma 10.2.2, T_{1,p^e} consists exactly of the $A \in \Delta_0(N)$ with $\det A = p^e$ and entry-gcd 1; and the entry-gcd is constant on cosets $\Gamma_0(N)A$.

Case $p \nmid N$. By Lemma 10.4.1 the cosets $\Gamma_0(N)A$, $\det A = p^e$, are represented by $\mathcal{R}_{p^e} = \left\{ \begin{pmatrix} p^{e-f} & m \\ 0 & p^f \end{pmatrix} : 0 \leq f \leq e, 0 \leq m < p^f \right\}$; those in T_{1,p^e} are those with entry-gcd 1, i.e. $\gcd(p^{e-f}, m, p^f) = 1$.

Case $p \mid N$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in T_{1,p^e}$; then $\gcd(a, N) = 1$, so $p \nmid a$. By Step 1 of the proof of Lemma 10.2.2 there is $\gamma_1 \in \Gamma_0(N)$ with $\gamma_1 A = \begin{pmatrix} a_1 & b_1 \\ 0 & d_1 \end{pmatrix}$, $\gcd(a_1, N) = 1$, $a_1 d_1 = p^e$. Hence $a_1 \mid p^e$ and $p \nmid a_1$, so $a_1 = \pm 1$; multiplying by $-I$ if necessary, $a_1 = 1$ and $d_1 = p^e$. Then $T^k \gamma_1 A = \begin{pmatrix} 1 & b_1 + kp^e \\ 0 & p^e \end{pmatrix}$, and choosing k we get $0 \leq b_1 + kp^e < p^e$. So every coset has a representative $\begin{pmatrix} 1 & m \\ 0 & p^e \end{pmatrix}$, $0 \leq m < p^e$. These are distinct: if $\begin{pmatrix} 1 & m' \\ 0 & p^e \end{pmatrix} = \gamma \begin{pmatrix} 1 & m \\ 0 & p^e \end{pmatrix}$, then $\gamma = \begin{pmatrix} 1 & m' \\ 0 & p^e \end{pmatrix} \begin{pmatrix} 1 & m \\ 0 & p^e \end{pmatrix}^{-1} = \begin{pmatrix} 1 & (m' - m)/p^e \\ 0 & 1 \end{pmatrix}$, which is integral only if $m' \equiv m \pmod{p^e}$, i.e. $m' = m$. Conversely each $\begin{pmatrix} 1 & m \\ 0 & p^e \end{pmatrix}$ lies in $\Delta_0(N)$ with entry-gcd 1, hence in T_{1,p^e} . ■

Correction. My handwritten proof of this proposition stopped after observing that $\gcd(a, cN) \mid p^e$ (with the question “ $\gcd(a, cN) = 1$?”). The argument above instead brings A to upper triangular form inside $\Gamma_0(N)A$.

10.5 Generators of the Hecke algebra and the action on q -expansions

Proposition 10.5.1

If $p \mid N$, then $T_{1,p}T_{1,p^e} = T_{1,p^{e+1}}$; hence $T_{p^e} = T_p^e$.

Proof. By Proposition 10.4.2, $T_{1,p} = \bigsqcup_{m < p} \Gamma_0(N) \begin{pmatrix} 1 & m \\ 0 & p \end{pmatrix}$ and $T_{1,p^e} = \bigsqcup_{m' < p^e} \Gamma_0(N) \begin{pmatrix} 1 & m' \\ 0 & p^e \end{pmatrix}$. The products are

$$\begin{pmatrix} 1 & m \\ 0 & p \end{pmatrix} \begin{pmatrix} 1 & m' \\ 0 & p^e \end{pmatrix} = \begin{pmatrix} 1 & m' + mp^e \\ 0 & p^{e+1} \end{pmatrix},$$

and as (m, m') runs over $[0, p) \times [0, p^e)$, $m' + mp^e$ runs over $[0, p^{e+1})$ exactly once. So the $p \cdot p^e$ products are exactly the representatives of $T_{1,p^{e+1}}$, each once: all coefficients are 1 and $T_{1,p}T_{1,p^e} = T_{1,p^{e+1}}$. With Proposition 10.3.2(1),(2), $T_{p^e} = T_{1,p^e} = T_{1,p}^e = T_p^e$. ■

Theorem 10.5.2 ($\circledast$, generators of the Hecke algebra)

$\mathcal{H}_N$ is generated as a ring by T_p and $T_{p,p}$ for primes $p \nmid N$, and T_p for primes $p \mid N$:

$$\mathcal{H}_N = \mathbb{Z}[T_p, T_{p,p} \ (p \nmid N); T_p \ (p \mid N)].$$

Proof sketch. The $T_{l,m}$ ($l \mid m$, $\gcd(l, N) = 1$) form a $\mathbb{Z}$ -basis. By Proposition 10.3.2(4), $T_{l,m} = T_{l,l}T_{1,m/l}$, and $T_{l,l} = \prod_p T_{p,p}^{v_p(l)}$. One shows, using the coset decompositions of Proposition 10.4.2 as in Proposition 10.5.1, that

- $T_{1,m}T_{1,m'} = T_{1,mm'}$ if $\gcd(m, m') = 1$;
- for $p \nmid N$: $T_{1,p}T_{1,p^e} = T_{1,p^{e+1}} + (p+1)T_{p,p}$ if $e = 1$, and $= T_{1,p^{e+1}} + pT_{p,p}T_{1,p^{e-1}}$ if $e \geq 2$;
- for $p \mid N$: $T_{1,p}T_{1,p^e} = T_{1,p^{e+1}}$ (Proposition 10.5.1).

By induction on e , every T_{1,p^e} is a polynomial in $T_{1,p} = T_p$ and $T_{p,p}$, and the claim follows. See [5, §3.3] or [3, §4.5]; in fact $\mathcal{H}_N$ is the polynomial ring in these generators. ■

Correction. My handwritten notes stated this theorem (marked $\circledast$) and began the coset decomposition needed for its proof, but did not finish; the coset decompositions are now Proposition 10.4.2, the case $p \mid N$ is proved completely (Proposition 10.5.1), and for $p \nmid N$ we quote the standard relation.

Proposition 10.5.3 (Action of T_p on q -expansions)

Let $f = \sum_{n \geq 0} a_n q^n \in M_k(N, \chi)$ and p a prime. Then

$$f|_k T_p = \frac{1}{p} \sum_{m=0}^{p-1} f\left(\frac{z+m}{p}\right) + \chi(p) p^{k-1} f(pz) = \sum_{n \geq 0} \left(a_{np} + \chi(p) p^{k-1} a_{n/p}\right) q^n,$$

where $a_{n/p} = 0$ if $p \nmid n$, and $\chi(p) = 0$ if $p \mid N$. For $N = 1$ this is Proposition 8.3.6.

Proof. By Proposition 10.3.2(1) and Proposition 10.4.2, $T_p = T_{1,p}$ has the representatives $\alpha_m = \begin{pmatrix} 1 & m \\ 0 & p \end{pmatrix}$, $0 \leq m < p$, and, if $p \nmid N$, also $\alpha_\infty = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$. With $\chi(\alpha) = \overline{\chi(a)}$ we have $\overline{\chi(\alpha_m)} = \chi(1) = 1$ and $\overline{\chi(\alpha_\infty)} = \chi(p)$. Moreover $f|_k \alpha_m(z) = p^{k/2} p^{-k} f\left(\frac{z+m}{p}\right)$ and $f|_k \alpha_\infty(z) = p^{k/2} f(pz)$. Hence

$$f|_k T_p = p^{k/2-1} \left(\sum_{m=0}^{p-1} p^{-k/2} f\left(\frac{z+m}{p}\right) + \chi(p) p^{k/2} f(pz) \right) = \frac{1}{p} \sum_{m=0}^{p-1} f\left(\frac{z+m}{p}\right) + \chi(p) p^{k-1} f(pz),$$

the second term being absent if $p \mid N$ (consistent with $\chi(p) = 0$). The q -expansion follows exactly as in Proposition 8.3.6: $\frac{1}{p} \sum_m f\left(\frac{z+m}{p}\right) = \sum_n a_{np} q^n$ and $f(pz) = \sum_n a_n q^{pn}$. ■

Remark 10.5.4. Since $\mathcal{H}_N$ is commutative and (by the Petersson inner product, not treated here) the operators T_n with $\gcd(n, N) = 1$ are normal on $S_k(N, \chi)$, the space $S_k(N, \chi)$ has a basis of simultaneous eigenforms for these T_n ; for newforms one obtains, as in Proposition 8.5.2, multiplicative Fourier coefficients with $a_{p^{r+1}} = a_p a_{p^r} - \chi(p) p^{k-1} a_{p^{r-1}}$. This is the beginning of Atkin–Lehner–Li theory [1, §5.6–5.8] and the end of these notes.

Bibliography

- [1] F. Diamond and J. Shurman, *A First Course in Modular Forms*, Graduate Texts in Mathematics 228, Springer, 2005.
- [2] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, 2nd ed., Graduate Texts in Mathematics 97, Springer, 1993.
- [3] T. Miyake, *Modular Forms*, Springer Monographs in Mathematics, Springer, 1989 (reprint 2006).
- [4] J.-P. Serre, *A Course in Arithmetic*, Graduate Texts in Mathematics 7, Springer, 1973.
- [5] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Publications of the Mathematical Society of Japan 11, Iwanami Shoten and Princeton University Press, 1971.
- [6] E. M. Stein and R. Shakarchi, *Complex Analysis*, Princeton Lectures in Analysis II, Princeton University Press, 2003.
- [7] E. M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton Lectures in Analysis I, Princeton University Press, 2003.