MAT540 Topics in Number Theory: L-functions and Modular Forms

Claire Burrin
(Spring 2022, University of Zurich)

Contents

0.1. Asymptotic notation

We write

- $f(x) = g(x) + O(h(x))$ to mean there exists $c > 0$ such that
$$|f(x) - g(x)| \leq c|h(x)| \quad \text{for all } x \in X$$
where the range X should be specified if not clear from context.;
- $f(x) \ll g(x)$ to say $f(x) = O(g(x))$, i.e., f has order of magnitude smaller or equal than g. Once again, the range over which this holds should be clarified. For example, $x \ll x^2$ if $x > 1$ but $x^2 \ll x$ if $x \in (0, 1)$;
- $f(x) \asymp g(x)$ to say $f(x) \ll g(x)$ and $g(x) \ll f(x)$, i.e., f and g have the same order of magnitude
- $f(x) \sim g(x)$ as $x \to x_0$ if
$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1;$$
- $f(x) = o(g(x)$ as $x \to x_0$ if
$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

For a complex number s, we will repeatedly use the variables $\sigma = \mathrm{Re}(s)$, $t = \mathrm{Im}(s)$.

CHAPTER 1

Introduction: Euler, Gauss, Dirichlet

1.1. The Euler product

Riemann's zeta function is

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

with s taken to be a complex variable $s \in \mathbf{C}$ with real part $\mathrm{Re}(s) > 1$.

Proposition 1. *$\zeta(s)$ converges absolutely for $\mathrm{Re}(s) > 1$.*

PROOF. Write $s = \sigma + it$ in rectangular coordinates. Note $|n^s| = |n^\sigma e^{it \log n}| = n^\sigma$. Hence

$$|\zeta(s)| \leq \sum_{n \geq 1} \frac{1}{n^\sigma}.$$

By the integral test, this series converges iff

$$\int_1^\infty x^{-\sigma} dx = \left. \frac{x^{1-\sigma}}{1-\sigma} \right|_1^\infty < +\infty$$

that is, iff $\sigma > 1$. □

Remark 2. *The same argument shows that the harmonic series diverges although this was already known in the mid-14th century. Oresme (circa 1350) gave the following argument:*

$$\begin{aligned} 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \dots & \\ > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \frac{1}{16} + \dots & \\ = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots = \infty & \end{aligned}$$

In other words, the harmonic series diverges by the comparison test.

The series ζ was studied with $s \in \mathbf{N}$ (or even $s \in \mathbf{R}$) well before Riemann considered its complex analytic properties. Around 1650, the **Basel problem** was formulated: find the exact value of the convergent sum

$$\zeta(2) = \sum_{n \geq 1} \frac{1}{n^2}.$$

Similar (...although superficially similar) sums were known to have simple closed form expressions. For example,

$$\sum_{n\geq 1} \frac{1}{n(n+1)} = \sum_{n\geq 1} \left(\frac{1}{n} - \frac{1}{n+1} \right) = 1.$$

The question whether $\zeta(2)$ admitted a closed form formula for $\zeta(2)$ baffled many mathematicians (including the Bernoullis in Basel who worked extensively on the question) for more than half a century until Euler eventually proved that

THEOREM 1 (Euler, 1734).

$$\zeta(2) = \frac{\pi^2}{6}$$

PROOF. The proof consists in playing the following two representations of $\sin(\pi x)$ against each other. On the one side, we have the Maclaurin series

$$\sin(\pi x) = \pi x - \frac{(\pi x)^3}{3!} + \frac{(\pi x)^5}{5!} - \dots$$

and on the other side, Euler proved the infinite product

$$\sin(\pi x) = \pi x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2} \right),$$

which converges by Weierstrass factorization theorem. Regarding $\sin(\pi x)$ as an infinite polynomial, we have the term-by-term identity

$$\begin{aligned} \pi x - \frac{(\pi x)^3}{3!} + \frac{(\pi x)^5}{5!} - \dots &= \pi x - \pi\zeta(2)x^3 + \frac{\pi}{2} \sum_{m\geq 1} \frac{1}{m^2} \sum_{\substack{n\geq 1 \\ n\neq m}} \frac{1}{n^2} x^5 + \dots \\ &= \pi x - \pi\zeta(2)x^3 + \frac{\pi}{2} \sum_{m\geq 1} \frac{1}{m^2} \left(\zeta(2) - \frac{1}{m^2} \right) x^5 + \dots \\ &= \pi x - \pi\zeta(2)x^3 + \frac{\pi}{2} \left(\zeta(2)^2 - \zeta(4) \right) x^5 + \dots \end{aligned}$$

From this we immediately deduce that $\zeta_2 = \pi^2/3!$.

Euler's idea of the product expansion came by analogy to the factorization of finite degree polynomials. By the fundamental theorem of algebra, if $f(x) \in \mathbf{C}[x]$ has degree n and $f(0) = 1$, then

$$f(x) = \left(1 - \frac{x}{a_1} \right) \left(1 - \frac{x}{a_2} \right) \cdots \left(1 - \frac{x}{a_n} \right)$$

where $a_1, \dots a_n$ are the roots of $f(x)$. For the infinite polynomial $\sin(\pi x)$, we have $\sin(\pi x) = 0$ iff $x \in \mathbf{Z}$. This led Euler to the infinite product above, which was eventually formally justified. □

Remark 3. *As a bonus, we may deduce $\zeta(4) = \pi^4/90$ as well. Euler did this for the first few values of $\zeta(2n)$. These are now generally understood in terms of the Bernoulli numbers.*

After this success, Euler turned to the arithmetic properties of $\zeta(s)$, establishing the **Euler product** expansion:

THEOREM 2 (Euler, 1737). *Whenever* $\mathrm{Re}(s) > 1$, *we have*

$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1}$$

PROOF. Starting with $\zeta(s)$ and applying the fundamental theorem of arithmetic[1], we have an expression of the form

$$\prod_p \sum_{j \geq 0} \frac{1}{p^{sj}},$$

which we infer to take the RHS form. Let us make this precise starting from the RHS. Since $p \geq 2$, we have the geometric series

$$\frac{1}{1 - p^{-k}} = 1 + p^{-k} + p^{-2k} + \dots.$$

Restricting to $p = 2, \dots, P$ and multiplying the series together, we see every $n \in \mathbf{N}$ with no prime factor larger than P appearing, i.e.

$$\prod_{p \leq P} \left(1 - p^{-s}\right)^{-1} = \sum_{(P)} \frac{1}{n^s}.$$

With $(P) \supset \{1, \dots, P\}$, we can show

$$0 < \sum_{n \geq 1} \frac{1}{n^s} - \sum_{(P)} \frac{1}{n^s} < \sum_{n > P} \frac{1}{n^s} \to 0$$

as $P \to \infty$. □

The existence of the Euler product yields a new (analytic) proof of the infinitude of primes. Indeed, if there are only finitely many primes, then the RHS is a finite product, and the RHS should always converge, including as $s \to 1$, which is absurd. The existence of infinitely many primes was of course already known since Euclid.

THEOREM 4 (Euclid). *There are infinitely many primes.*

PROOF. Suppose for contradiction that there are only finitely many primes $2, 3, 5, \dots, p$ and set

$$N = 2 \cdot 3 \cdots p + 1.$$

Then by construction $N > p$, so it cannot be prime. Since it is odd, it must have an odd prime factor $q \mid N$. Then q is an element of our finite list. In particular, we have both $q \mid N$ and $q \mid N - 1$, which is absurd. □

1

THEOREM 3 (Euclid, circa 300BC). *Every whole number* $n \geq 2$ *can be written as a product of primes numbers in a unique way* $n = p_1^{k_1} \cdots p_\ell^{k_\ell}$ *(up to reordering of the factors).*

Remark that the uniqueness statement forces us to not consider 1 as a prime.

Actually, Euler proceeds to deduce a stronger statement than just that there are infinitely many primes (although at the time he did so with dubious leaps of faith), namely that

THEOREM 5 (Euler, 1737).

$$\sum_p \frac{1}{p} = \infty.$$

PROOF. Observe that

$$-\sum_p \log(1-p^{-s}) = \sum_p \sum_{r\geq 1} \frac{1}{rp^{rs}},$$

where the last equality follows from the Maclaurin expansion of $\log(1-x)$ with $|x|<1$. Hence for $\mathrm{Re}(s)>1$, we have

$$\zeta(s) = \exp\left(\sum_p \sum_{r\geq 1} \frac{1}{rp^{rs}}\right)$$

which shows that $\zeta(s)\neq 0$ for all $s\in\mathbf{C}$ with $\mathrm{Re}(s)>1$.

Moreover, this also shows that

$$\sum_p \frac{1}{p^s} = \log\zeta(s) - \sum_p \sum_{r\geq 2} \frac{1}{rp^{rs}},$$

where the second sum is bounded by

$$\sum_p \sum_{r\geq 2} \frac{1}{rp^{rs}} \leq \sum_p \sum_{r\geq 2} \frac{1}{2p^{rs}} = \sum_p \frac{1}{2p^s(p^s-1)} < \sum_{n\geq 2} \frac{1}{2n(n-1)} = \frac{1}{2}.$$

The statement now follows by taking $s\to 1$. □

This clearly implies directly that there are infinitely many primes (if the sum was finite, it would converge) but by doing so quantitatively, it provides some more information on the density of primes. (This result often cited as the 'birth of analytic number theory.') For example, we now see that primes are denser than squares (numbers of the form n^2) since $\zeta(2)<\infty$.

1.2. The distribution of prime numbers

Euler's theorem is a remarkable improvement considering how irregular the distribution of primes is. Consider for instance that

Proposition 4. *There exist arbitrary large gaps between successive primes.*

PROOF. Choose $p\in\mathbf{N}$ a large prime. Then

$$p!+2, p!+3, p!+4, \ldots p!+p$$

is a sequence of $p-1$ successive composite numbers. Since there are infinitely many primes, we may take p to be arbitrarily large. □

Although the distribution of the prime numbers is irregular, their average distribution behaves remarkably regularly. Let

$$\pi(x) = \#\{p \leq x,\, p \in \mathcal{P}\}$$

be the number of primes $\leq x$. In the 1790s, Legendre and Gauss both conjectured, based on numerical observations, that the density of prime numbers among the first N whole numbers is approximatively $(\log N)^{-1}$, i.e.,

$$\mathbb{P}_N(n \in \mathcal{P}) := \frac{\pi(N)}{N} \sim \frac{1}{\log N}.$$

This led Gauss to conjecture that a good approximation of $\pi(x)$ is given by the **logarithmic integral**

$$\mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}$$

for $x > 2$. We leave it as an exercise for the reader to check that

$$\mathrm{Li}(x) \sim \frac{x}{\log x}$$

as $x \to \infty$. This conjecture, eventually proved by de la Vallée Poussin and Hadamard independently in 1896 (following a strategy sketched 40 years earlier by Riemann that we will discuss) is called the **prime number theorem** (PNT).

Proposition 5. *Let p_n denote the n-th prime. The PNT is equivalent to $p_n \sim n \log n$.*

Proof. The PNT says that

$$n = \pi(p_n) \sim \frac{p_n}{\log(p_n)}.$$

We show that $\log(p_n) \sim \log(n)$ as $n \to \infty$. Taking the log of the above asymptotic, we have

$$\log n \sim \log p_n - \log\log p_n = \log p_n + o(\log p_n).$$

That is, $\log n \sim \log p_n$. □

Corollary 6. *Let g_n denote the gap between consecutive primes $g_n = p_{n+1} - p_n$. We have $g_n \sim \log n$. In other words, the average size of g_n is $\log n$.* □

Hence, as far as gaps between primes are concerned, while on average $g_n \sim \log n$, we have seen that

$$\limsup_{n\to\infty} g_n = \infty$$

and we believe that

$$\liminf_{n\to\infty} g_n = 2$$

which goes by the name of the **twin prime conjecture** (there exist infinitely many pairs of successive primes that differ by 2). In fact, we currently know that

$$\liminf_{n\to\infty} g_n < N,$$

where the first breakthrough is due to Zhang showing in 2013 that $N = 7 \cdot 10^7$, and was brought down to $N = 600$ (Maynard, 2013) and $N = 246$ (Polymath, 2014).

In conclusion, there is no hope for a nice distribution function to describe the statistical behavior of prime numbers, yet their average behavior is remarkably regular.

1.3. Primes in arithmetic progressions

Clearly, aside from 2, all primes are odd numbers, so of the form $2k+1$. Among those, depending on whether k is even or odd, we have primes p of the form $p \equiv 1 \pmod 4$ and primes p of the form $p \equiv 3 \pmod 4$. (See Aside 15 at the end of this section below for a quick review of Gauss' congruence notation $\equiv$.)

A remarkable pattern arises:

THEOREM 6 (Fermat, 1640). *If $p \equiv 1 \pmod 4$, then p is the sum of two squares, i.e., $p = x^2 + y^2$.*

On the other hand, it is a simple exercise to verify that no whole number of the form $n \equiv 3 \pmod 4$ can be written as a sum of two squares:

Proposition 7. *If $n \equiv 3 \pmod 4$ then n is not the sum of two squares.*

PROOF. Suppose for contradiction that $n = a^2 + b^2$. If a and b are both odd (or even), then this gives $1 \equiv 0 \pmod 2$, which is absurd. We may then assume that a is odd and b is even. We then have $3 \equiv (2k+1)^2 \equiv 1 \pmod 4$, which is again absurd. □

This leads to many natural questions: Are there infinitely many primes of both types? Do we expect primes of the form $p \equiv 3 \pmod 4$ more frequent (in some sense)?

The goal of this section is to discuss the proof strategy of

THEOREM 7 (Dirichlet, 1837). *Every arithmetic progression*

$$a,\ a+n,\ a+2n, \dots$$

with $(a,n) = 1$ contains infinitely many primes.

The base strategy consists in mimicking Euler's proof of Theorem 5, namely to show that

$$\sum_{p \equiv a\ (n)} \frac{1}{p^s} \to \infty$$

as $s \to 1^+$. For this, one needs to find a convenient way to detect only those primes that satisfy $p \equiv a \pmod n$ — or equivalently, $[p] = [a]$ in the group $G = (\mathbf{Z}/n\mathbf{Z})^\times$. Fourier analysis provides the idea. Recall that each continuous function $f : \mathbf{R} \to \mathbf{C}$ for which $f(x+1) = f(x)$ admits a Fourier series

$$f(x) = \sum_{n \in \mathbf{Z}} a_n e^{2\pi i n x},$$

where the Fourier coefficients are given by

$$\int_0^1 f(x) e^{-2\pi i m x} dx = \sum_{n \in \mathbf{Z}} a_n \int_0^1 e^{2\pi i (n-m) x} dx = \sum_{n \in \mathbf{Z}} a_n \delta_n(m) = a_m.$$

To underline here is that the delta-function

$$\delta_n(m) = \begin{cases} 1 & m = n \\ 0 & \text{else} \end{cases}$$

arises from the orthoghonality of the characters $e^{2\pi inx}$. In similar fashion, we will consider characters on the group $G = (\mathbf{Z}/n\mathbf{Z})^\times$. Recall more generally

Definition 8. *Let G be a finite abelian group. A character χ on G is a group homomorphism $\chi : G \to \mathbf{C}^*$ (where $\mathbf{C}^*$ is seen as a group with respect to multiplication).*

Remark 9. *Since G is finite, say of order $|G| = n$, any character on G takes values the n-th roots of unity; indeed, $\chi(g)^n = \chi(g^n) = \chi(e) = 1$. In particular, we have*

$$\chi(g)^{-1} = \frac{1}{\chi(g)} = \frac{\overline{\chi(g)}}{|\chi(g)|^2} = \overline{\chi(g)}.$$

We will use without proof that

THEOREM 8. *Let G be a finite abelian group. The set $\widehat{G}$ of all characters on G is a finite abelian group (called the dual group) of order $|G| = |\widehat{G}|$.*

The group structure is given with respect to the pointwise multiplication of functions, i.e., $(\chi_1 \cdot \chi_2)(g) = \chi_1(g)\chi_2(g)$. The trivial element is the trivial character $\chi_0(g) = 1$ for all $g \in G$.

THEOREM 9. *Let G be a finite abelian group. We have the (dual) orthogonality relations*

$$\frac{1}{|G|}\sum_{g\in G}\chi(g) = \delta_{\chi_0}(\chi) = \begin{cases} 1 & \chi = \chi_0; \\ 0 & \textit{else;} \end{cases}$$

$$\frac{1}{|\widehat{G}|}\sum_{\chi\in\widehat{G}}\chi(g) = \delta_e(g) = \begin{cases} 1 & g = e; \\ 0 & \textit{else.} \end{cases}$$

PROOF. We prove only the second relation. If $g = e$, then the statement is immediate. Suppose that G is not trivial and choose $g \in G$ such that $g \neq e$. Then $\widehat{G}$ is nontrivial and we may choose $\chi_1 \in \widehat{G}$ such that $\chi_1(g) \neq 1$. Consider

$$\chi_1(g)\sum_{\chi\in\widehat{G}}\chi(g) = \sum_{\chi\in\widehat{G}}\chi_1(g)\chi(g) = \sum_{\chi\in\widehat{G}}(\chi_1\cdot\chi)(g) = \sum_{\chi\in\widehat{G}}\chi(g).$$

Since $\chi_1(g) \neq 1$, we conclude that the sum is equal 0. □

A slight modification allows to detect other elements in the group than the trivial ones:

Corollary 10. *Fix $\chi_1 \in \widehat{G}$ and $g_1 \in G$. Then*

$$\frac{1}{|G|}\sum_{g\in G}\overline{\chi_1(g)}\chi(g) = \delta_{\chi_1}(\chi) = \begin{cases}1 & \chi = \chi_1;\\ 0 & \textit{else;}\end{cases}$$

$$\frac{1}{|\widehat{G}|}\sum_{\chi\in\widehat{G}}\overline{\chi(g_1)}\chi(g) = \delta_{g_1}(g) = \begin{cases}1 & g = g_1;\\ 0 & \textit{else.}\end{cases}$$

PROOF. Observe that $\overline{\chi(g_1)}\chi(g) = \chi(g_1^{-1}g)$. By the previous theorem, the sum vanishes whenever $g_1^{-1}g \neq e$. □

Aside 11 (Finite Fourier series)**.** *Let G be a finite abelian group and consider $L(G)$ the set of all functions $f : G \to \mathbf{C}$. This is a vector space over $\mathbf{C}$. Each $f \in L(G)$ can be written as a linear combination*

$$f(x) = \sum_{g\in G} f(g)\delta_g(x).$$

In fact, one can show that $\{\delta_g : g \in G\}$ is a basis for $L(G)$, hence $\dim L(G) = |G|$. Applying the orthogonality of characters as above yields the **finite Fourier series**

$$\begin{aligned} f(x) &= \sum_{g\in G} f(g)\left(\frac{1}{|\widehat{G}|}\sum_{\chi\in\widehat{G}}\overline{\chi(g)}\chi(x)\right)\\ &= \sum_{\chi\in\overline{G}} a_\chi\chi(x) \quad \textit{with} \quad a_\chi = \frac{1}{|G|}\sum_{g\in G} f(g)\overline{\chi(g)}. \end{aligned}$$

We now specialize this discussion to the group $G = (\mathbf{Z}/n\mathbf{Z})^\times$ of order $\varphi(n)$ (where $\varphi(n)$ denotes the Euler totient function).

Example 12. *Let $n = 3$. Then $G = \{1, -1\}$. Since $\chi(-1)^2 = 1$, the two possible characters on G are given by*

a	1	-1
χ_0	1	1
χ_1	1	-1

Example 13. *Exercise: Write down the character table for $n = 5$.*

The orthogonality relation now spells

$$\frac{1}{\varphi(n)}\sum_{\chi\,(n)}\overline{\chi(a)}\chi(m) = \begin{cases}1 & m \equiv a\,(n)\\ 0 & \text{else,}\end{cases}$$

where the summation index is taken to mean that we are summing over all characters on the group G as above. Note that we trivially extend a character on G to $\mathbf{Z}$ by considering it as

$$\chi(m) = \begin{cases}\chi([m]) & (m,n) = 1\\ 0 & (m,n) \neq 1.\end{cases}$$

Recall that we want to prove

$$\sum_{p\equiv a\,(n)} \frac{1}{p^s} = \frac{1}{\varphi(n)} \sum_{\chi\,(n)} \overline{\chi(a)} \sum_p \frac{\chi(p)}{p^s} \to \infty$$

as $s \to 1^+$.

Definition 14. *Let $n \in \mathbf{N}$. A* **Dirichlet character (mod n)** *is a function $\chi : \mathbf{Z} \to \mathbf{C}$ with the following properties*

(1) *χ is completely multiplicative, i.e., $\chi(ab) = \chi(a)\chi(b)$ for all $a, b \in \mathbf{Z}$;*
(2) *χ is periodic mod n, i.e., $\chi(a) = \chi(b)$ if $a \equiv b$ (mod n);*
(3) *χ is induced by a character on $(\mathbf{Z}/n\mathbf{Z})^\times$, i.e., $\chi(a) = 0$ if $(a, n) \neq 1$.*

Moreover, we call a Dirichlet character induce by the trivial character a **principal character**.

Fix a Dirichlet character $\chi \pmod q$. Consider the **Dirichlet L-function**

$$L(s, \chi) = \sum_{n\geq 1} \frac{\chi(n)}{n^s}$$

Then by comparison with the Riemann ζ-function, we see that $L(s, \chi)$ converges absolutely whenever $\mathrm{Re}(s) > 1$ and by the multiplicativity of Dirichlet characters, it admits the Euler product

$$\sum_{n\geq 1} \frac{\chi(n)}{n^s} = \prod_p (1 - \chi(p)p^{-s})^{-1}$$

whenever $\mathrm{Re}(s) > 1$. Mimicking Euler's proof, we find that

$$\log L(s, \chi) = \sum_p \frac{\chi(p)}{p^s} + O(1).$$

Using that (see exercise sheet)

$$L(s, \chi_0) = \sum_{\substack{n\geq 1 \\ (n,q)=1}} \frac{1}{n^s} = \prod_{p|n} (1 - p^{-s})\zeta(s),$$

we have

$$\begin{aligned} \sum_{p\equiv a\,(n)} \frac{1}{p^s} &= \frac{1}{\varphi(n)} \sum_{\chi\,(n)} \overline{\chi(a)} \log L(s, \chi) + O(1) \\ &= \frac{1}{\varphi(n)} \left(\log \zeta(s) + \sum_{\chi\neq\chi_0} \overline{\chi(a)} \log L(s, \chi) \right) + O(1). \end{aligned}$$

Clearly $\log \zeta(s) \to \infty$ as $s \to 1^+$. It remains to see that we don't have $L(s, \chi) \to 0$ as $s \to 1^+$. We will soon show that in fact $L(1, \chi) \neq 0$ whenever $\chi \neq \chi_0$.

Aside 15 (Gauss' *Disquisitiones arithmeticae*, 1801)**.** *Gauss' famous manuscript was the first modern systematic treaty of number theory, offering a synthesis (with complete proofs) of the earlier results of Fermat, Euler, Lagrange, and Legendre, and Gauss' own original works, which set the basis of what is now algebraic number theory. Formally, it also set what is still today the golden standard of mathematical writing: the statement of a theorem, followed by its proof and corollaries, as well as the recourse to abundant (numerical) examples.*

Although the notions of groups, rings, and fields appeared only later, much of the structure of the rings $\mathbf{Z}/n\mathbf{Z}$*, their unit groups* $G = (\mathbf{Z}/n\mathbf{Z})^\times$ *and the finite fields* $\mathbf{F}_p$ *was dealt with here. Among others, we owe Gauss the* **congruence notation***; we say that* $a \equiv b \pmod n$ *('*a *is congruent to* b *modulo* n*') iff* $a = b + kn$ *for some* $k \in \mathbf{Z}$*, or equivalently,* $n \mid (a-b)$*. The congruence relation defines an equivalence relation, so that we may assign all* $a \equiv b\ (n)$ *to the same equivalence class* $[a]$*, where we choose* a *to be the unique representative* $a \in \{0, \dots, n-1\}$*. Hence*

$$\mathbf{Z}/n\mathbf{Z} = \{[a] : a = 0, \dots, n-1\},\ (\mathbf{Z}/n\mathbf{Z})^\times = \{[a] : a = 0, \dots, n-1,\ (a,n) = 1\}.$$

Even if '$\equiv$*' behaves mostly like '*$=$*', beware for example that the cancellation law doesn't completely hold. In fact,*

$$ac \equiv bc \quad (n) \implies a \equiv b \quad \left(\frac{n}{(n,c)}\right).$$

A good simple illustration of the efficiency of the congruence notation is given by divisibility criteria. Take $n \in \mathbf{N}$ *and consider its digit expansion*

$$n = \sum_{i=1}^{k} a_i 10^i = a_k 10^k + \cdots + a_1 \cdot 10 + a_0.$$

Then $n \equiv a_0 \pmod 2$ *so that we need only check the last digit to know if* n *is even or odd, or, less obviously,* $n \equiv a_0 + a_1 + \dots a_k \pmod 3$ *so that* n *is divisible by 3 iff the sum of its digits is, and so on.*

1.4. Dirichlet series

Let $(a_n)_{n\geq 1} \subset \mathbf{C}$. Its **Dirichlet series** is the generating function

$$F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where s is understood to be a complex variable. We have already seen a few examples: $\zeta(s)$, $L(s,\chi)$, $1/\zeta(s)$ (see exercise set 1 for the last one).

1.4.1. The arithmetic of formal Dirichlet series. . As formal series, we may add and multiply Dirichlet series as follows. Let

$$F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}, \qquad G(s) = \sum_{n=1}^{\infty} \frac{b_n}{n^s}.$$

Then

$$F(s)+G(s)=\sum_{n=1}^{\infty}\frac{a_n+b_n}{n^s} \qquad \text{and} \qquad F(s)\cdot G(s)=\sum_{n=1}^{\infty}n^{-s}\sum_{d|n}a_d b_{n/d},$$

where the last inner sum is called the **Dirichlet convolution** $(a*b)(n)$. The formal Dirichlet series forms a ring. The most interesting Dirichlet series for us arise when $a: n\mapsto a_n$ is multiplicative:

Definition 16. *We say that $f:\mathbf{N}\to\mathbf{C}$ is* **multiplicative** *if $f(mn)=f(m)f(n)$ whenever $(m,n)=1$. We say that f is* **completely multiplicative** *if $f(mn)=f(m)f(n)$ for all $m,n\geq 1$.*

1.4.2. Examples. For ζ, $1/\zeta$, we know moreover that we have absolute convergence when $\mathrm{Re}(s)>1$. Using only these two functions, we can generate new examples of Dirichlet series, e.g.,

$$\begin{aligned}\zeta(s)^2&=\sum_{n\geq 1}\frac{d(n)}{n^s}\\ \frac{\zeta(s)^2}{\zeta(2s)}&=\sum_{n\geq 1}\frac{2^{\omega(n)}}{n^s}\\ \frac{\zeta(s)}{\zeta(2s)}&=\sum_{n\geq 1}\frac{|\mu(n)|}{n^s}\\ \frac{\zeta(s-1)}{\zeta(s)}&=\sum_{n\geq 1}\frac{\varphi(n)}{n^s}\end{aligned}$$

where $d(n)$ is the number of positive divisors of n, $\omega(n)$ is the number of distinct prime factors of n, $\mu(n)$ is the Möbius function, and $\varphi(n)$ is the totient function. There are many more examples of course (see exercise sheet 2).

We add to the list the Dirichlet representation of the functions $\log\zeta(s)$ and $\log L(s,\chi)$, which are key in the arguments of Euler and Dirichlet.

Proposition 17. *Let $\mathrm{Re}(s)>1$ and let χ be a Dirichlet character mod q. Then*

$$\log\zeta(s)=\sum_{n=2}^{\infty}\frac{\Lambda(n)}{(\log n)n^s} \qquad \textit{and} \qquad \log L(s,\chi)=\sum_{n=2}^{\infty}\frac{\chi(n)\Lambda(n)}{(\log n)n^s}$$

where

$$\Lambda(n)=\begin{cases}\log p & n=p^r \text{ for some } r\geq 1,\\ 0 & \text{otherwise}\end{cases}$$

is the **von Mangoldt function**.

Proof. We have

$$\log\zeta(s)=\sum_{p}\sum_{r\geq 1}\frac{1}{rp^{rs}}=\sum_{n=p^r}\frac{1}{rn^s}=\sum_{n=p^r}\frac{\log p}{(\log n)n^s}=\sum_{n\geq 2}\frac{\Lambda(n)}{(\log n)n^s}.$$

The second identity is derived in the exact same way. □

1.4.3. Analytic properties of Dirichlet series. We now address the question of convergence. In this section, $F(s)$ will always denote the Dirichlet series

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

The following lemma generalizes the argument of Proposition 1.

Lemma 18. *If $a_n n^{-\sigma_0} = O(1)$ then $F(s)$ converges absolutely for $\mathrm{Re}(s) > 1 + \sigma_0$.*

PROOF. We have

$$|F(s)| \leq \sum_{n\geq 1} \frac{|a_n|}{n^\sigma} \ll \sum_{n\geq 1} n^{\sigma_0 - \sigma}$$

and by the integral test this converges iff $\sigma - \sigma_0 > 1$. □

THEOREM 10. *If $F(s)$ converges at $s = s_0$. Then $F(s)$ converges uniformly in every compact set of the halfplane $\sigma > \sigma_0$ and defines a holomorphic function in that halfplane. Moreover, its derivatives can be computed termwise, i.e.,*

$$F^{(k)}(s) = \sum_{n\geq 1} \frac{a_n(-\log n)^k}{n^s}$$

and converge uniformly on compact subsets of $\sigma > \sigma_0$ as well.

To prove this, we will need the following important trick, which plays the role of integration by parts in our context.

Lemma 19 (Abel summation)**.** *Let $f \in C^1(\mathbf{R}_{>0})$, $(a_n)_{n\geq 1} \subset \mathbf{C}$ and set*

$$A(x) = \sum_{n\leq x} a_n$$

for $x > 1$. Then

$$\sum_{n\leq N} a_n f(n) = A(N)f(N) - \int_1^N A(x)f'(x)dx.$$

PROOF. Using that $a_n = A(n) - A(n-1)$, we have that

$$\begin{aligned}\sum_{n\leq N} a_n f(n) &= \sum_{n\leq N} (A(n) - A(n-1))f(n) = \sum_{n=1}^{N} A(n)f(n) - \sum_{n=1}^{N-1} A(n)f(n+1)\\ &= A(N)f(N) - \sum_{n=1}^{N-1} A(n)\left(f(n+1) - f(n)\right)\\ &= A(N)f(N) - \sum_{n=1}^{N-1} A(n)\int_n^{n+1} f'(x)dx.\end{aligned}$$

Since $A(n) = A(\lfloor x \rfloor) = A(x)$ for $x \in [n, n+1)$, we can move $A(n)$ inside the integral and we arrive at the desired formula. □

PROOF OF THEOREM 10. We may assume that $s_0 = 0$ by replacing $F(s)$ by

$$G(s) = \sum_{n\geq 1} \frac{a_n n^{-s_0}}{n^s} = F(s + s_0).$$

Hence wlog let $s_0 = 0$.

By assumption, $\sum_{n\geq 1} a_n$ converges. Hence for any $\varepsilon > 0$, there exists $M_0 \geq 1$ such that for all $M \geq M_0$, we have

$$\left| \sum_{n=M+1}^{N} a_n \right| < \varepsilon.$$

Let $\mathrm{Re}(s) = \sigma > 0$. By Abel summation, we have

$$\sum_{n=M+1}^{N} \frac{a_n}{n^s} = \sum_{n=M+1}^{N} a_n N^{-s} + s \int_{M+1}^{N} \left(\sum_{n=M+1}^{x} a_n \right) \frac{dx}{x^{s+1}}$$

for all $1 \leq M < N$, and hence

$$\left| \sum_{n=M+1}^{N} \frac{a_n}{n^s} \right| < \varepsilon N^{-\sigma} + |s| \left. \frac{t^{-\sigma}}{(-\sigma)} \right|_{M+1}^{N} = \varepsilon \left(N^{-\sigma} + \frac{|s|}{\sigma}((M+1)^{-\sigma} - N^{-\sigma}) \right) < \varepsilon \left(1 + \frac{|s|}{\sigma} \right).$$

Fix a compact set K in $\mathrm{Re}(s) > 0$. Then there is $\delta > 0$ and $R \geq 1$ such that for all $s \in K$, we have $|s| < R$ and $\sigma > \delta$. The uniform convergence follows.

The rest of the statement follows by standard complex analysis. Fix a compact set K. Given we have the uniform convergence $F_n \to F$, we have

$$\lim_{n\to\infty} \int_\gamma F_n(s) ds = \int_\gamma F(s) ds$$

for any simple closed curve γ in K. By Cauchy's theorem $\int_\gamma F_n(s) ds = 0$ and hence $\int_\gamma F(s) ds = 0$. By Morera's theorem, it follows that $F(s)$ is holomorphic. The statement for the derivatives is derived similarly using the formula

$$F_n'(s') = \frac{1}{2\pi i} \int \frac{F_n(s)}{(s - s')^2} ds.$$

□

From these arguments, we gain the following criterion for convergence of a Dirichlet series (in the spirit of Lemma 18).

Lemma 20. *Suppose that $A(N) = \sum_{n\leq N} a_n n^{-s_0} = O(1)$. Then $F(s)$ is convergent and holomorphic for $\sigma > \sigma_0$.*

PROOF. We may assume that $s_0 = 0$ (as in the previous proof). Let $\mathrm{Re}(s) > 0$. By Abel summation, we have the identity

$$\sum_{n=1}^{N} \frac{a_n}{n^s} = A(N) N^{-s} + s \int_1^N \frac{A(x)}{x^{s+1}} dx.$$

By assumption, $\sum_{n\leq N} a_n = O(1)$. In particular, $|A(N)N^{-s}| \ll N^{-\sigma} \to 0$ as $N \to \infty$ while $\int_1^N |A(x)|x^{-\sigma-1}dx = O(1)$ as $N \to \infty$. Hence taking $N \to \infty$, we have the identity

$$F(s) = \sum_{n\geq 1} \frac{a_n}{n^s} = s\int_1^\infty \frac{A(x)}{x^{s+1}}dx$$

for which we have just seen that the RHS (right handside) converges absolutely. This concludes. □

1.4.4. Cancellation and analytic continuation: Example 1. We will compare the convergence of the Dirichlet series $\zeta(s)$ and

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

Clearly, both converge absolutely for $\sigma > 1$ and clearly $\zeta(s)$ is not holomorphic at $s = 1$. On the other hand, the alternating signs in $\eta(n)$ provide a larger half-plane of convergence. Indeed, $\sum_{n\leq N}(-1)^n \in \{0, 1\}$ and so by Lemma 20, $\eta(n)$ converges for $\mathrm{Re}(s) > 0$. This is interesting since clearly the two series are related. Indeed, observe that

$$\begin{aligned}\eta(s) &= \sum_{n \text{ even}} \frac{(-1)^n}{n^s} + \sum_{n \text{ odd}} \frac{(-1)^n}{n^s} \\ &= \sum_{n \text{ even}} \frac{1}{n^s} - \sum_{n \text{ odd}} \frac{1}{n^s} = 2\sum_{n\geq 1} \frac{1}{(2n)^s} - \zeta(s) = (2^{1-s} - 1)\zeta(s).\end{aligned}$$

This allows to define

$$\zeta(s) = \frac{\eta(s)}{2^{1-s} - 1}$$

as the analytic continuation of the ζ-function to $\mathrm{Re}(s) > 0$. It follows immediately that this analytic continuation is holomorphic with a simple pole at each zero of $2^{1-s} - 1$, that is, at

$$s_k = 1 - \frac{2\pi i}{\log 2}k$$

for each $k \in \mathbf{Z}$. Actually, Widder proved that $\eta(s_k) = 0$ for $k \neq 0$, so that $\zeta(s)$ has only one simple pole, at $s = 1$, and you can check that its residue is

$$\lim_{s\to 1}(s - 1)\zeta(s) = 1.$$

1.4.5. Cancellation and analytic continuation: Example 2. We now consider the Dirichlet L-functions $L(s, \chi)$, where χ is a Dirichlet character mod q.

When $\chi = \chi_0$ is the principal character, then we recall (exercise sheet 1) that

$$L(s, \chi_0) = \prod_{p|q}(1 - p^{-s})\zeta(s).$$

By the previous discussion of the ζ-function, $L(s, \chi_0)$ has abscissa of convergence $\sigma_c = 1$ and admits an analytic continuation to $\sigma > 0$ that is holomorphic except for the simple pole at $s = 1$.

For Dirichlet L-functions attached to nonprincipal characters, we will now see that we have a larger halfplane of convergence thanks to cancellation in the sums $\sum_1^N \chi(n)$ as a consequence of the symmetries expressed by the orthogonality relation. More explicitly, we show

Proposition 21. *Let χ be a nonprincipal Dirichlet character mod q. Then*

$$\left|\sum_{n=1}^{N} \chi(n)\right| \leq q.$$

PROOF. Let k be the largest integer such that $kq \leq N$. Observe that for each $j = 1, \ldots, k-1$, we have

$$\sum_{n=(j-1)q}^{jq} \chi(n) = \sum_{n=1}^{q} \chi(n) = \sum_{n\,(q)} \chi(n) = 0.$$

(Recall properties of Dirichlet characters from last section.) Then

$$\left|\sum_{n=1}^{N} \chi(n)\right| = \left|\sum_{n=jq+1}^{N} \chi(n)\right| \leq q.$$

□

Since $\sum_{n\leq N} \chi(n) = O(1)$, we conclude by Lemma 20 that $L(s, \chi)$ converges for all $\sigma > 0$ and is holomorphic in that halfplane.

Remark 22. *The Pólya–Vinogradov inequality (1918) is the much stronger bound*

$$\left|\sum_{n\leq N} \chi(n)\right| \leq 2\sqrt{q} \ln q$$

for any nonprincipal Dirichlet character (mod q). This bound plays an important role in the study of the distribution of quadratic residues.

1.4.6. Cancellation and analytic continuation: Example 3. For the series

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s},$$

the **Riemann hypothesis** conjectures that $\sigma_c = 1/2$.

For comparison, recall that $\sigma_c(\zeta) = 1$ and $\sigma_c(\eta) = 0$. Consider now the Dirichlet series for which $\mu(n)$ is replaced by a sequence of random variables $X_n \in \{-1, 1\}$ such that all X_n are iid (independent and identically distributed) with the uniformly distribution. By standard probabilistic arguments (Kolmogorov's law of iterated logarithm), we expect that $\sigma_c = 1/2$. That is, we may understand the Riemann hypothesis to say that the statistical behavior of the Möbius function $\mu(n)$ is 'pseudorandom', i.e., it

behaves similarly to what one would expect from a series with random ± 1 in the place of $\mu(n)$, yet the $\mu(n)$ themselves from a deterministic sequence.

1.5. Landau's proof of Dirichlet's theorem

The goal of this section is to prove that $L(1,\chi) \neq 0$ for all $\chi \neq \chi_0$. Following the discussion in Section 2, this finishes the proof of Dirichlet's theorem on the existence of infinitely many primes in an arithmetic progression.

The following result says that if a Dirichlet series with nonnegative coefficients can be extended to a neighborhood $s = \sigma_c$, then it has a singularity at $s = \sigma_c$ (which generalizes our observations for $\zeta(s)$.)

Lemma 23 (Landau's lemma)**.** *Let $F(s) = \sum a_n n^{-s}$ be a Dirichlet series with abscissa of convergence σ_c and such that $a_n \in \mathbf{R}$, $a_n \geq 0$. Then $s = \sigma_c$ is a singularity of $F(s)$.*

PROOF. We once again assume that $\sigma_c = 0$. Suppose for contradiction that $F(s)$ is analytic at $s = 0$, and hence in a small δ-neighborhood of $s = 0$. Then the power series of $F(s)$ at $s = 1$, given by

$$\sum_{k\geq 0} \frac{F^{(k)}(1)}{k!}(s-1)^k = \sum_{k\geq 0} \frac{1}{k!} \sum_{n\geq 1} \frac{a_n (\log n)^k}{n} (1-s)^k$$

has radius of convergence $1+\varepsilon$ for some small $0 < \varepsilon < \delta$. Observe that for $s \in \mathbf{R}_{<1}$, all summands are nonnegative. By Cauchy's double summation theorem, we can thus exchange the summations to get

$$\sum_{n\geq 1} \frac{a_n}{n} \sum_{k\geq 0} \frac{(\log n(1-s))^k}{k!} = \sum_{n\geq 1} \frac{a_n}{n} e^{(1-s)\log n} = \sum_{n\geq 1} \frac{a_n}{n^s}.$$

Hence we conclude that $F(s)$ converges at $s = -\varepsilon/2$, which is impossible if $\sigma_c = 0$. □

THEOREM 11. *Let $\chi \neq \chi_0$ be a Dirichlet character mod q. Then $L(1,\chi) \neq 0$.*

PROOF. First note that since $\chi \neq \chi_0$, $L(1,\chi)$ is well defined.

We will treat characters separately, depending on whether they are complex(-valued) or quadratic, the idea being that complex characters are easy to handle. Note that if χ is real(-valued), it can only take values ± 1, and hence $\chi^2 = \chi_0$.

Suppose first that χ is complex and that $L(1,\chi) = 0$. Then $\overline{L(1,\chi)} = L(1,\overline{\chi}) = 0$ as well. Looking at the product $\prod L(s,\chi)$ over all Dirichlet characters χ mod q, we see that the factor $L(s,\chi)L(s,\overline{\chi})$ contributes a double zero at $s = 1$ while the factor $L(s,\chi_0)$ contributes a simple pole at $s = 1$. Since all factors beside $L(s,\chi_0)$ in the product are analytic at $s = 1$, we conclude that

$$\prod_{\chi\,(q)} L(1,\chi) = 0.$$

This is absurd since for all $\sigma > 1$

$$\prod_{\chi\,(q)} L(s,\chi) = \exp\left(\sum_{\chi\,(q)} \log L(s,\chi)\right) = \exp\left(\sum_{\chi\,(q)}\sum_{n\geq 2} \frac{\chi(n)\Lambda(n)}{n^s \log n}\right)$$
$$= \exp\left(\varphi(q) \sum_{\substack{n\geq 2 \\ n\equiv 1\,(q)}} \frac{\Lambda(n)}{n^s \log n}\right)$$

as follows from the orthogonality of characters. Indeed taking $s \in \mathbf{R}_{>0}$ and letting $s \to 1^+$, we see $\prod L(s,\chi) \geq 1$.

Suppose now that χ is quadratic and that $L(1,\chi) = 0$. Then $L(s,\chi)\zeta(s)$ has an analytic continuation to the halfplane $\sigma > 0$ where it is holomorphic (the zero 'cancels' the pole). We can write this function down as the Dirichlet series

$$L(s,\chi)\zeta(s) = \sum_{n\geq 1} n^{-s} \sum_{d|n} \chi(d).$$

Let

$$r(n) = \sum_{d|n} \chi(d).$$

First observe that $r(n)$ is multiplicative (exercise). Hence to understand $r(n)$, we need to understand its values on prime powers

$$r(p^\alpha) = \sum_{k=0}^{\alpha} \chi(p)^k.$$

Clearly $r(p^\alpha)$ is completely determined by whether $\chi(p) = 1$ or -1. In fact, we have

$$r(p^\alpha) = \begin{cases} \alpha + 1 & \chi(p) = 1, \\ 1 & \chi(p) = -1,\ \alpha \text{ even}, \\ 0 & \chi(p) = -1,\ \alpha \text{ odd}. \end{cases}$$

By Landau's lemma, $L(s,\chi)\zeta(s)$ converges for $\sigma > 0$. We find a contradiction by showing that

$$\sum_{n\geq 1} \frac{r(n)}{n^{1/2}} \geq \sum_{n\geq 1} \frac{r(n^2)}{n} \geq \sum_{n\geq 1} \frac{1}{n}. \tag{1.1}$$

The second inequality follows from $r(n^2) = r(p_1^{2\alpha_1}) \cdots r(p_k^{2\alpha_k}) \geq 1$ for all $n \in \mathbf{N}$ and this proves (1.1). □

1.6. Exercises

Exercise 24. *Show directly that there exist infinitely many primes $p \equiv 1$ (mod 4) and infinitely many primes $p \equiv 3$ (mod 4).*

Exercise 25. *Let $m, n \in \mathbf{N}$. Show that there exists a sequence of n consecutive numbers each of which is divisible by at least m distinct prime numbers.*

Exercise 26. *Let φ denote the Euler totient function.*

(1) *Show that φ is multiplicative, i.e., $\varphi(mn) = \varphi(m)\varphi(n)$ if $(m, n) = 1$.*
(2) *Let p be a prime, $k \in \mathbf{N}$. Show that $\varphi(p^k) = p^k - p^{k-1}$.*
(3) *Show that*

$$\varphi(n) = n \prod_{p|n} (1 - p^{-1}).$$

Exercise 27. *Since ζ is nonvanishing when $\mathrm{Re}(s) > 1$, we have*

$$\frac{1}{\zeta(s)} = \prod_p (1 - p^{-s}) = (1 - 2^{-s})(1 - 3^{-s})(1 - 5^{-s}) \cdots$$

$$= 1 - 2^{-s} - 3^{-s} - 5^{-s} + 6^{-s} - 7^{-s} + 10^{-s} - \cdots = \sum_{n \geq 1} \frac{\mu(n)}{n^s}$$

where the **Möbius function** *$\mu(n)$ is defined to be $\mu(1) = 1$, $\mu(n) = 0$ if n is divisible by a square and $\mu(n) = (-1)^{\omega}(n)$ if n is squarefree where $\omega(n)$ is the number of distinct prime factors of n.*

(1) *Show that*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & n = 1, \\ 0 & n > 1 \end{cases}$$

(2) *Prove the* **Möbius inversion formula***: Let $f, g : \mathbf{N} \to \mathbf{C}$ be two arithmetic functions, then*

$$g(n) = \sum_{d|n} f(d) \iff f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d).$$

Exercise 28. *Let $q \in \mathbf{N}$ and let χ_0 be the principal Dirichlet character mod q.*

(1) *Show that*

$$\sum_{d|q} \frac{\mu(d)}{d^s} = \prod_{p|q} \left(1 - p^{-s}\right).$$

(2) *Deduce that*

$$L(s, \chi_0) = \prod_{p|q} \left(1 - p^{-s}\right) \zeta(s).$$

Exercise 29. *Let a_n be the number of finite abelian groups of order n up to isomorphism.*

(1) *Show that a_n is multiplicative.*
(2) *Show that a_{p^k} is equal to the partition number of k*
(3) *Deduce that*

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} = \zeta(s)\zeta(2s)\zeta(3s) \cdots$$

CHAPTER 2

Riemann's memoir

2.1. Euler's Γ-function

The factorial $n! = n(n-1)\cdots 2\cdot 1$ has the integral representation

$$n! = \int_0^\infty e^{-t}t^n dt.$$

(This is easily checked by integration by parts.) Euler's **Γ-function** is defined similarly; let

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

for all $s \in \mathbf{C}$ for which the integral converges.

Proposition 30. *$\Gamma(s)$ converges absolutely for $\sigma > 0$.*

Proof. We consider separately the two integrals arising from

$$\Gamma(s) = \int_1^\infty e^{-t}t^s \frac{dt}{t} + \int_0^1 e^{-t}t^s \frac{dt}{t}.$$

In the first case, we use that e^{-t} tends to 0 faster than any monomial in t. In fact, for any $M \geq 1$, we see that

$$e^{-t} \leq \frac{M!}{t^M}$$

by taking the Maclaurin expansion of e^t. We can always choose M sufficiently large such that

$$\left|\int_1^\infty e^{-t}t^s \frac{dt}{t}\right| \leq \int_1^\infty e^{-t}t^{\sigma-1}dt \ll \int_1^\infty t^{\sigma-M-1}dt < +\infty.$$

In the second case, we have

$$\int_0^1 e^{-t}t^{\sigma-1}dt \leq t^\sigma|_0^1 < +\infty$$

iff $\sigma > 0$. □

By comparison to the factorial, we have $n! = \Gamma(n+1)$ so that $n! = n(n-1)!$ implies the recursion formula $\Gamma(n+1) = n\Gamma(n)$. In fact, it is an exercise in integration by parts to show that

$$\Gamma(s+1) = s\Gamma(s)$$

whenever $\Gamma(s)$ is defined. In fact, this identity allows to analytically continue $\Gamma(s)$ to all $s \in \mathbf{C}$; we use for that

$$\Gamma(s) = \frac{\Gamma(s+1)}{s} = \frac{\Gamma(s+2)}{s(s+1)} = \cdots = \frac{\Gamma(s+n+1)}{s(s+1)\cdots(s+n)}.$$

Hence

Proposition 31. *$\Gamma(s)$ admits an analytic continuation to all $s \in \mathbf{C}$ that is analytic except for countably many simple poles at $s = 0, -1, -2, -3, \ldots$ with residue*

$$\mathrm{Res}_{s=-n}\Gamma(s) = \lim_{s\to -n}(s+n)\Gamma(s) = \frac{(-1)^n}{n!}.$$

□

We can almost as easily show that $\Gamma(s)$ has no zeroes; i.e., $\Gamma(s) \neq 0$ for all $s \in \mathbf{C}$. One proof of this fact can be deduced from the **Euler refection formula**:

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

as a meromorphic function on $\mathbf{C}$ with simple poles at $s \in \mathbf{Z}$. (See exercise sheet 3.)

2.2. Mellin transform

In more modern language, one introduces $\Gamma(s)$ as the Mellin transform of the function e^{-t}.

Definition 32. *The* **Mellin transform** *of $f : \mathbf{R}_{>0} \to \mathbf{C}$ is given by*

$$\mathcal{M}(f)(s) = \int_0^\infty f(t)t^s \frac{dt}{t}$$

for all $s \in \mathbf{C}$ such that this integral converges. If the set of all such s is nonempty, it is given by a vertical strip $\sigma_1 < \mathrm{Re}(s) < \sigma_2$.

Remark 33. *We think of $\frac{dt}{t}$ as Haar measure with respect to multiplication by positive scalars. Let $\alpha > 0$ and consider the change of variable $u = \alpha t$. Then $\frac{du}{u} = \frac{dt}{t}$.*

The Mellin transform is closely related to the Fourier transform. Let $t = e^u$ and write $s = \sigma + i\tau$. Then for $\sigma_1 < \sigma < \sigma_2$, the Mellin transform can be expressed as

$$\int_0^\infty f(t)t^s \frac{dt}{t} = \int_{-\infty}^\infty \left(f(e^u)e^{\sigma u}\right) e^{i\tau u} du,$$

where the RHS is the usual Fourier transform of $F(u) = f(e^u)e^{\sigma u}$. Correspondingly we have a Mellin inversion theorem

Theorem 12 (Mellin inversion). *Let $f \in C(\mathbf{R}_{>0})$. Let $\sigma_1 < \sigma_2$ delimit the vertical strip of convergence of its Mellin transform $\mathcal{M}(f)$. Then for any $\sigma_1 < \sigma < \sigma_2$, we have*

$$f(y) = \frac{1}{2\pi i}\int_{(\sigma)} \mathcal{M}(f)(s)y^{-s}ds := \frac{y^{-\sigma}}{2\pi i}\lim_{T\to\infty}\int_{\sigma-iT}^{\sigma+iT} \mathcal{M}(f)(\sigma+it)y^{-it}dt.$$

Proof. Follows from the Fourier inversion theorem via change of variables. □

Remark 34. *Building on Fourier theory, the inversion theorem also holds under the condition that f be piecewise C^1. Then*

$$\frac{f(y^+)+f(y^-)}{2} = \frac{1}{2\pi i}\int_{(\sigma)} \mathcal{M}(f)(s)y^{-s}ds$$

for all $\sigma \in (\sigma_1, \sigma_2)$.

As an application to the theory of Dirichlet series, we prove Perron's inversion formula.

THEOREM 13 (Perron's formula). *Let $F(s) = \sum a_n n^{-s}$ be a Dirichlet series with abscissa of convergence σ_c. Then for $\sigma > \max\{0, \sigma_c\}$, $x > 0$, we have*

$$\sum_{n\leq x}{}' a_n = \frac{1}{2\pi i}\int_{(\sigma)} F(s)\frac{x^s}{s}ds,$$

where the 'prime' superscript on the sum means that

$$\sum_{n\leq x}{}' a_n = \begin{cases} a_1 + \cdots + a_N & \text{if } x \in (N, N+1), \\ a_1 + \cdots + a_{N-1} + \dfrac{a_N}{2} & \text{if } x = N. \end{cases}$$

PROOF. Set $A(x) = \sum_{n\leq x} a_n$ and $A^*(x) = \sum_{n\leq x}{}' a_n$. Recall that by Abel summation, we have

$$F(s) = s\int_0^\infty A(x)x^{-s}\frac{dx}{x}.$$

We may view the RHS as a Mellin transform. Clearly, $A(y)$ is not continuous, but piecewise continuous. By observing that $A^*(x) = \frac{1}{2}(A(x^+) + A(x^-))$ and applying Mellin inversion in the form given in Remark 34 concludes. □

2.3. Riemann's memoir

Recall that in the 1790s, Gauss conjectured that

$$\pi(x) \sim \mathrm{Li}(x)$$

as $x \to \infty$. The proof strategy was eventually outlined in a short article of Riemann in 1859. (His only published work in number theory.) He first considered $\zeta(s)$ as a function of s as a complex variable and replaced it by the 'nicer' function

$$\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s),$$

which we call the **completed Riemann ζ-function**. We will introduce and discuss the main properties of the Euler gamma-function $\Gamma(s)$ soon, but for the moment we will ignore it to give a short informal description of the work of Riemann.

Riemann showed that

(I) $\xi(s)$ has an analytic continuation to all $s \in \mathbf{C}$ that is holomorphic except for simple poles at $s = 0$ and $s = 1$. Moreover, it satisfies the **functional equation**

$$\xi(s) = \xi(1-s);$$

(II) $\xi(s)$ has infinitely many complex zeroes $\rho+i\tau$ in the **critical strip** $0 \le \rho \le 1$. More precisely, if $N(T)$ is the number of zeroes in the critical strip of the form $\rho + i\tau$ with $0 \le \rho \le 1$ and $0 < \tau < T$, then

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + O(\log T).$$

Viewing $\xi(s)$ has a function on the whole complex plane, Riemann stated (a preliminary form of) the **explicit formula**:

(III) For any noninteger $x > 1$, we have

$$\psi(x) = \sum_{n\le x}\Lambda(n) = x - \lim_{T\to\infty}\sum_{\substack{\xi(\rho+i\tau)=0\\ |\tau|\le T}}\frac{x^{\rho+i\tau}}{\rho+i\tau} - \log(2\pi) - \frac{1}{2}\log(1-x^{-2}),$$

where the middle sum is determined by the zeroes of ξ.

Recall (see exercise sheet 2) that the PNT (i.e., $\pi(x) \sim \mathrm{Li}(x)$) is equivalent to $\psi(x) \sim x$ as $x \to \infty$. Hence the explicit formula implies the PNT if we can show that $\rho < 1$. In fact, it not only confirms Gauss' conjecture, but shows that the distribution of primes is controlled by the location of the zeroes of the complex analytic function $\xi(s)$, and as such provides a strategy towards proving the PNT. It took until 1895 for a complete proof of the explicit formula, due to von Mangoldt. The PNT was proved one year later, independently by Hadamard and by de la Vallée Poussin, by showing that $\xi(1+it) \neq 0$ for all $t \in \mathbf{R}$. Regarding the location of the zeroes of $\xi(s)$, Riemann additionally conjectured that

(IV) **Riemann hypothesis** (RH): Every zero of $\xi(s)$ is on the line $\mathrm{Re}(s) = 1/2$.

If we take for granted the validity of Riemann's assertions (I)–(III) above, what RH tells us in terms of the distribution of prime numbers is

Proposition 35.

$$\pi(x) = \mathrm{Li}(x) + O(\sqrt{x}\log x) \iff \psi(x) = x + O(\sqrt{x}\log^2 x) \iff \textit{RH}.$$

Proof. The first equivalence is a consequence of Abel summation (see exercise sheet 2). Set $E(x) = \psi(x) - x$. For $\sigma > 1$,

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= \sum_{n\ge1}\frac{\Lambda(n)}{n^s} = s\int_1^\infty \frac{\psi(x)}{x^{s+1}}dx \\ &= s\int_1^\infty \frac{dx}{x^s} + s\int_1^\infty E(x)\frac{dx}{x^{s+1}} \\ &= \frac{s}{s-1} + s\int_1^\infty E(x)\frac{dx}{x^{s+1}}. \end{aligned}$$

If $E(x) = O(\sqrt{x}\log^2 x)$, then we see that the integral converges absolutely whenever $\sigma > 1/2$. This shows that $-\zeta'(s)/\zeta(s)$ admits a meromorphic continuation to the halfplane $\sigma > 1/2$. In particular, this implies that $\zeta(s)$ has no zeroes when $\sigma > 1/2$. Together with the functional equation for ζ, i.e.,

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s),$$

it also has no zeroes when $0 \leq \sigma < 1/2$. Hence all the zeroes in the critical strip are on the line passing at $1/2$.

For the converse, we will also assume that the following weaker form of the explicit formula holds: for all $x > 0$, $T \geq 2$,

$$\psi(x) = x - \sum_{|\tau| \leq T} \frac{x^s}{s} + O\left(\frac{x \log^2(xT)}{T} + \log x\right).$$

If RH holds, the middle sum becomes

$$\begin{aligned}\sqrt{x} \sum_{|\tau|<T} \frac{1}{|1/2 + i\tau|} &\leq \sqrt{x} \sum_{n \leq T} \sum_{\tau \in [n, n+1)} \frac{1}{|1/2 + i\tau|} \\ &\leq \sqrt{x} \sum_{n \leq T} \frac{1}{n} \sum_{\tau \in [n, n+1)} 1 \ll \sqrt{x} \sum_{n \leq T} \frac{\log(n+1)}{n} \ll \sqrt{x} \log^2(T).\end{aligned}$$

We conclude by choosing $T = x$. □

We can also state Riemann's assertions In terms of the original ζ-function:

(I') $\zeta(s)$ has an analytic continuation to all $s \in \mathbf{C}$ that is holomorphic except for a simple pole at $s = 1$ (with residue 1) and satisfies the functional equation

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

(II') $\zeta(s)$ has infinitely many complex **"nontrivial zeroes"** $\rho + i\tau$ in the critical strip $0 \leq \rho \leq 1$ and simple **"trivial zeroes"** at $s = -2, -4, -6, \ldots$. (These come from the poles of $\Gamma(2s)$.)

(IV') **Riemann hypothesis**: Every nontrivial zero of $\zeta(s)$ is on the line $\mathrm{Re}(s) = 1/2$.

2.4. RH and the 'random behavior of primes'

Unlike the other statements of Riemann, the Riemann hypothesis is famously still open; it is one of the millenium problems[1].

In this short section, we discuss a probabilistic heuristic (the Cramér model) through which we may view the Riemann hypothesis as saying that we expect prime numbers to "behave randomly."

Consider a sequence of independent Bernoulli random variables (X_n). This means that for each X_n, we have $X_n = 1$ with probability p_n, and $X_n = 0$ with probability $1 - p_n$. To be consistent with the prime number theorem, we choose p_n to be

$$p_n = \frac{1}{\log n}$$

for $n \geq 3$, $p_2 = 1$ and $p_1 = 0$. This yields a sequence $\mathcal{P} := \{n : X_n = 1\}$ that will play the rôle of "random primes". For instance, the 'random prime counting function'

$$\Pi(x) = \sum_{\substack{n \leq x \\ n \in \mathcal{P}}} 1 = \sum_{n \leq x} X_n,$$

[1] https://www.claymath.org/millennium-problems

amounts to be a sum of independent random variables. We find that

$$\mathbb{E}[\pi(x)] = \sum_{n\leq x} \frac{1}{\log n} = 1 + \sum_{n\geq 3} \frac{1}{\log n} = \mathrm{Li}(x) + O(1)$$

and

$$\mathbb{V}[\pi(x)] = \sum_{n\leq x} p_n(1-p_n) \ll \mathrm{Li}(x).$$

By Kolmogorov's law of iterated logarithm (1929), we have

$$\Pi(x) = \mathbb{E}[\Pi(x)] + O\left(\sqrt{\mathbb{V}[\Pi(x)]\log\log\mathbb{V}[\Pi(x)]}\right) = \mathrm{Li}(x) + O\left(\sqrt{x}\left(\frac{\log\log x}{\log x}\right)^{1/2}\right)$$

with probability 1 as $x \to \infty$. Considering Proposition 35, this is seen as saying that the $\Pi(x)$ satisfies RH with probability 1.

The (rough) idea behind Cramér's model (1936) is that if a statement holds true with probability 1 for $\mathcal{P}$, then something similar should be true for the set of prime numbers. Of course, there are limitations to Cramér's model — after all the sequence of primes is deterministic and not truly random — yet it has shown to be a solid heuristic model to examine the statistical behavior of primes. We will not delve further in probabilistic number theory; the interested reader is encouraged to consult the lecture notes of Kowalski's course at ETH.

2.5. Truncated Perron and Cauchy's theorem

The goal of this section is to sketch out some of the principal elements of proof behind Riemann's explicit formula. Recall Perron's inversion formula (Theorem 13): starting from

$$-\frac{\zeta'(s)}{\zeta(s)} = s\int_0^\infty \frac{\psi(x)}{x^{s+1}}dx,$$

Mellon inversion yields the dual formula

$$\psi(x) = \frac{1}{2\pi i}\int_{(\sigma)} \left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^s}{s}ds,$$

for $\sigma > 1$ and where we choose $x > 0$, $x \notin \mathbf{N}$. Complex analysis provides us with a powerful tool to compute such line integrals:

THEOREM 14 (Cauchy's residue theorem). *Let $\Omega \subset \mathbf{C}$ be a simply connected open subset, $f : \Omega \to \mathbf{C}$ a holomorphic function except for at most finitely many points $\{a_1, \ldots, a_n\} \subset \Omega$., and let $\gamma \subset \Omega$ be an oriented simple closed curve in $\Omega \setminus \{a_1, \ldots, a_n\}$. Then*

$$\oint_\gamma f(z)dz = 2\pi i \sum_{k=1}^n \mathrm{Res}_{z=a_i} f(z).$$

Lemma 36. *As $T \to \infty$, we have*

$$\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\frac{y^s}{s}ds = \begin{cases} O\left(\dfrac{x^\sigma}{T\log x}\right) & \text{if } x<1, \\ \dfrac{1}{2}+O\left(\dfrac{1}{T}\right) & \text{if } x=1, \\ 1+O\left(\dfrac{x^\sigma}{T\log x}\right) & \text{if } x>1. \end{cases}$$

PROOF. Let $f(s) = s^{-1}y^s$; it is holomorphic in the whole complex plane except for the simple pole at $s = 1$. Suppose first that $y > 1$. Let $A > 0$ and let γ be the rectangle with sides

$$[\sigma - iT, \sigma + iT] \cup [\sigma + iT, -A + iT] \cup [-A + iT, -A - iT] \cup [-A - iT, \sigma - iT]$$

oriented counterclockwise. Then by Cauchy's theorem, $\oint_\gamma f(s)ds = 1$. Hence

$$\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\frac{y^s}{s}ds = 1 + \int_{-A+iT}^{\sigma+iT}\frac{y^s}{s}ds + \int_{-A-iT}^{-A+iT}\frac{y^s}{s}ds - \int_{-A-iT}^{\sigma-iT}\frac{y^s}{s}ds.$$

For the first integral, one has the estimate

$$\ll \int_{-A}^{\sigma}\frac{x^\rho}{T+|\rho|} \ll \frac{x^\sigma}{T\log x}$$

uniformly in A. The third integral is bounded similarly. For the second integral, we have

$$\ll x^{-A}\int_{-T}^{T}\frac{dt}{A+|t|} \to 0$$

as $A \to \infty$, for all $T > 0$. This proves the estimate for when $x > 1$. The situation when $x < 1$ is similar. Suppose now that $x = 1$. Then we may compute the integral directly:

$$\begin{aligned}\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\frac{1}{s}ds &= \frac{1}{2\pi}\int_{-T}^{T}\frac{dt}{\sigma+it} = \frac{1}{2\pi}\int_0^T\frac{dt}{\sigma+it} + \frac{1}{2\pi}\int_0^T\frac{dt}{\sigma-it} \\ &= \frac{\sigma}{\pi}\int_0^T\frac{dt}{\sigma^2+t^2} = \frac{\sigma}{\pi}\frac{\operatorname{atan}(\frac{T}{\sigma})}{\sigma} = \frac{1}{\pi}\left(\frac{\pi}{2}+O\left(\frac{1}{T}\right)\right)\end{aligned}$$

as $T \to \infty$. □

We can now apply this lemma to "truncate" Perron's formula. Let $x \notin \mathbf{N}$ and consider

$$\begin{aligned}\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^s}{s}ds &= \sum_{n<x}\Lambda(n)\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\left(\frac{x}{n}\right)^s\frac{ds}{s} + \sum_{n>x}\Lambda(n)\frac{1}{2\pi i}\int_{\sigma-iT}^{\sigma+iT}\left(\frac{x}{n}\right)^s\frac{ds}{s} \\ &= \psi(x) + O\left(\frac{x^\sigma}{T}\sum_{n\geq}\frac{|\Lambda(n)|}{n^\sigma(\log x - \log n)}\right).\end{aligned}$$

The explicit formula follows from applying contour integration again to the LHS and estimating the error term. To carry this program out, one needs information

specific to the Riemann ζ-function: how its logarithmic derivative behaves outside of its halfplane of absolute convergence (especially for T large) and an estimate on the number of zeroes in the critical strip.

2.6. A Mellin–transform proof of the functional equation

We will prove the analytic continuation of $\xi(s)$ and its functional equation together.

THEOREM 15. *Let $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ be the completed ζ-function. Then*

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

It follows that $\xi(s)$ admits a meromorphic continuation to $s \in \mathbf{C}$ with simple poles at $s = 0$ and $s = 1$ and that $\xi(s) = \xi(1-s)$.

PROOF. For $\sigma > 1$, we have

$$\begin{aligned}\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s) &= \sum_{n\geq 1}\int_0^\infty e^{-t}\left(\frac{t}{\pi n^2}\right)^{s/2}\frac{dt}{t} \\ &= \sum_{n\geq 1}\int_0^\infty e^{-t\pi n^2}t^{s/2}\frac{dt}{t} \\ &= \int_0^\infty \left(\sum_{n\geq 1} e^{-t\pi n^2}\right)t^{s/2}\frac{dt}{t}.\end{aligned}$$

The inner sum is $O(e^{-t\pi})$ and so the integral converges for $t \geq 1$ uniformly in s. On the other hand, as Riemann recognized, the inner sum is closely related to the Jacobi theta series

$$\tilde{\theta}(t) = \sum_{n\in\mathbf{Z}} e^{-\pi t n^2},$$

$t > 0$, for which

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

for all $t > 0$. (We will give the proof later in this section; for now we admit it as black box.) In fact, we have

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

Furthermore,

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

Hence

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

and this also shows that

$$\xi(1-s) = \int_1^\infty \frac{\tilde\theta(t)-1}{t}\left(t^{s/2}+t^{(1-s)/2}\right)\frac{dt}{t} - \frac{1}{s} + \frac{1}{s-1} = \xi(s).$$

□

The key to this proof is of course Jacobi's inversion formula for the $\tilde\theta$-series. First we want to explain where this series originates. Consider the related series

$$\sum_{n\in\mathbf{Z}} q^{n^2}.$$

Under the assumption that $|q|<1$, this infinite series converges. Let's view q as a complex number via the parametrization

$$q = e^{2\pi i z}e^{2\pi i(x+iy)} = e^{-2\pi y}e^{2\pi i x}.$$

The convergence of the series is guaranteed for all such parametrization with $y>0$. This leads us to the **theta-function**

$$\theta(z) = \sum_{n\in\mathbf{Z}} q^{n^2}$$

defined on the upper halfplane $\mathbf{H} = \{x+iy \in \mathbf{C} : y>0\}$. The relation to the theta-series above is given by

$$\tilde\theta(t) = \theta(it/2).$$

As an aside, let us mention that θ yields the generating series of some interesting number theoretic functions. Indeed, observe that

$$\theta^2(z) = \sum_{a\in\mathbf{Z}} q^{a^2}\sum_{b\in\mathbf{Z}} q^{b^2} = \sum_{a,b\in\mathbf{Z}} q^{a^2+b^2} = \sum_{n\geq 0} r_2(n)q^n,$$

where $r_2(n)$ is the number of ways we can write n as a sum of two squares $n = a^2 + b^2$, and more generally

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

We also want to make some preliminary remarks on the proof of the inversion formula for future references. The proof is a direct application of **Poisson summation** and more generally reflects the fact that $\theta(z)$ is what is called a modular form, the subject of the second half of this course. We will later see that in fact the Poisson summation formula is a powerful device to establish that a given function defines a modular form.

Before stating the Poisson summation formula, we will review a few elements of Fourier theory. Recall that a function $f \in L^1(\mathbf{R})$ has Fourier transform

$$\widehat{f}(t) = \int_{-\infty}^{\infty} f(x) e^{2\pi i t x} dx$$

and that if $\widehat{f} \in L^1(\mathbf{R})$, then we recover by Fourier inversion

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(t) e^{2\pi i t x} dt.$$

Example 37. *One important attribute of the Fourier transform is that when applied to a function f with large support, $\widehat{f}$ is very localized. For example, take $f(x) = 1_{[-T,T]}(x)$, then $\widehat{f}(t) = \frac{\sin(tT)}{t}$.*

Example 38. *A modification of this first example consists in taking*

$$f(x) = \begin{cases} \dfrac{T - |x|}{T} & |x| < T, \\ 0 & |x| \geq T. \end{cases}$$

Then

$$\widehat{f}(t) = \begin{cases} T & t = 0, \\ \dfrac{1 - \cos(2\pi t T)}{2\pi^2 t^2 T} & t \neq 0. \end{cases}$$

There is a (much smaller) class of functions that is preserved by the Fourier transform: a function f is Schwartz if $f \in C^\infty(\mathbf{R})$ and

$$\sup_{x \in \mathbf{R}} \left| x^m \frac{d^n}{dx^n} f(x) \right| < \infty$$

for all $m, n \in \mathbf{N}$.

Example 39. *The Gaussian function $f(x) = e^{-\pi x^2}$ is a special example of a Schwartz function. In fact, we will soon see that $\widehat{f}(x) = f(x)$.*

THEOREM 16 (Poisson summation formula). *Let $f \in C^\infty(\mathbf{R})$ be a Schwartz function. Then*

$$\sum_{n \in \mathbf{Z}} f(n) = \sum_{n \in \mathbf{Z}} \widehat{f}(n).$$

PROOF. Consider

$$F(x) = \sum_{n \in \mathbf{Z}} f(x+n).$$

Then F is Schwartz as well, and since $F(x) = F(x+1)$, it admits the Fourier series expansion

$$F(x) = \sum_{n \in \mathbf{Z}} \widehat{F}(n) e^{2\pi i n x}.$$

Hence

$$F(0) = \sum_{n \in \mathbf{Z}} f(n) = \sum_{n \in \mathbf{Z}} \widehat{F}(n),$$

where

$$\begin{aligned} \widehat{F}(n) &= \int_0^1 F(x) e^{-2\pi i n t x} dx = \sum_{m \in \mathbf{Z}} \int_0^1 f(x+m) e^{-2\pi i n t x} dx \\ &= \sum_{m \in \mathbf{Z}} \int_m^{m+1} f(x) e^{2\pi i n (x+m)} dx = \int_{-\infty}^{\infty} f(x) e^{2\pi i x} dx = \widehat{f}(n). \end{aligned}$$

□

Remark 40. *Observe that Poisson summation does not apply to Example 37 and trivially applies to Example 39. Moreover its conclusion also holds for Example 38; it should be noted that the Poisson summation formula may hold under a weaker condition on f than being Schwartz.*

We can now finally complete the proof of Riemann with Jacobi's inversion formula:

THEOREM 17 (Jacobi's inversion formula). *For each $t > 0$, we have $\tilde{\theta}(t) = t^{-1/2} \tilde{\theta}(1/t)$.*

PROOF. Let $t > 0$. We apply Poisson summation to $f(x) = e^{-\pi x^2 t}$. For this we need to compute $\widehat{f}(x)$. We have

$$\begin{aligned} \frac{d}{dx} \widehat{f}(x) &= \frac{d}{dx} \int_{\mathbf{R}} f(u) e^{-2\pi i u x} du = -2\pi i \int_{\mathbf{R}} f(u) u e^{-2\pi i u x} du \\ &= \frac{i}{t} \int_{\mathbf{R}} f'(u) e^{-2\pi i u x} du \\ &= -\frac{i}{t} \int_{\mathbf{R}} f(u) (-2\pi i x e^{-2\pi u x}) du = -\frac{2\pi x}{t} \widehat{f}(x). \end{aligned}$$

This ODE has solution

$$\widehat{f}(x) = C e^{-\pi x^2 / t}$$

for a constant C that can be computed via the usual computation of the Gaussian integral:

$$C = \widehat{f}(0) = \int_{\mathbf{R}} e^{-\pi x^2 t} dx = \frac{1}{\sqrt{t}}.$$

□

2.7. Exercises

Exercise 41. *Prove the Euler recurrence formula* $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$ *and deduce from it that* $\Gamma(s)$ *(seen as a meromorphic function on* $\mathbf{C}$ *with simple poles at* $s \in \mathbf{Z}$*) has no zeroes.*

Exercise 42.
(1) *Let* $f \in C_c^\infty(\mathbf{R}_{>0})$*. Show that its Mellin transform* $\mathcal{M}(f)(s)$ *is analytic for* $\mathrm{Re}(s) > 0$*.*
(2) *Let* $F(s) = \sum a_n n^{-s}$ *be a Dirichlet series that converges absolutely for* $\mathrm{Re}(s) = \sigma > 0$*. Show that*

$$\sum a_n f(n/x) = \frac{1}{2\pi i} \int_{(\sigma)} F(s)\mathcal{M}(f)(s)x^s ds$$

for all $x \geq 1$*.*

Exercise 43. *Let* $d(n)$ *be the number of positive divisors of* $n \in \mathbf{N}$*.*
(1) *Prove that* $d(n) \ll n^\varepsilon$ *for any* $\varepsilon > 0$*.*
(2) *We will now think of* n *as a random variable: Let* X *be a random variable that takes value* n *in* $\{1, \ldots, N\}$ *with respect to the discrete uniform distribution, and let* X_m *be random variables depending on* X *given by* $X_m = 1$ *if* $m \mid n$ *and 0 otherwise. Set* $d(n) = \sum X_m$*. Show that* $\mathbb{E}[d(n)] = \log N + O(1)$*.*

CHAPTER 3

Dirichlet L-functions

In this chapter, we will establish the analytic properties of Dirichlet L-functions. Let χ be a Dirichlet character mod q; recall Definition 14. Based on Riemann's proof of the analytic continuation of the zeta-function, we have

$$\pi^{-s/2}\Gamma(s/2)L(s,\chi) = \int_0^\infty \left(\sum_{n\geq 1} \chi(n)e^{-\pi n^2 t}\right) t^{s/2}\frac{dt}{t}$$

and consider the twisted theta series

$$\tilde{\theta}(t) = \sum_{n\in\mathbf{Z}} \chi(n)e^{-\pi n^2 t} = \sum_{n\geq 1} \left(\chi(n) + \chi(-1)\chi(n)\right) e^{-\pi n^2 t}.$$

Definition 44. *We say that a Dirichlet character χ is odd if $\chi(-1) = -1$ and even if $\chi(-1) = 1$.*

The above twisted theta series is identically zero if χ is odd. This leads to the following 'correct' definition of the twisted theta series: for a Dirichlet character χ (mod q), we define $\epsilon \in \{0,1\}$ via $\chi(-1) = (-1)^\epsilon$ and set

$$\tilde{\theta}_\chi(t) = \sum_{n\in\mathbf{Z}} n^\epsilon \chi(n)e^{-\pi n^2 t}.$$

We leave it as an exercise to check the identity

$$\pi^{-(s+\epsilon)/2}\Gamma(\tfrac{s+\epsilon}{2})L(s,\chi) = \int_0^\infty \tilde{\theta}_\chi(t)t^{(s+\epsilon)/2}\frac{dt}{t}$$

with $\epsilon \in \{0,1\}$ as above.

Our goal for the rest of this chapter is to prove the 'modularity' of these twisted theta series. In the setting of the Riemann zeta function, the modularity is deduced from Poisson summation. Here we have the problem that the summands in

$$\sum_{n\in\mathbf{Z}} \chi(n)f(n) = \sum_{n\in\mathbf{Z}} \chi(n)e^{-\pi n^2 t}$$

are a priori not functions on $\mathbf{R}$. The first thing we will see is that we can interpolate (primitive) characters defined on $\mathbf{Z}$ to smooth functions defined on $\mathbf{R}$ via the device of Gauss sums.

3.1. Primitive characters

From working out character tables for $(\mathbf{Z}/q\mathbf{Z})^\times$, one might observe that

Proposition 45. *If $d \mid q$ and given a Dirichlet character χ^* (mod d),*

$$\chi(n) = \begin{cases} \chi^*(n) & \text{if } (n,q) = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

defines a Dirichlet character mod q.

PROOF. Note that χ is a Dirichlet character mod q if it is completely multiplicative and periodic mod q.

Complete multiplicativity: If $(mn, q) = 1$ then $\chi(mn) = \chi(m)\chi(n)$ by definition. Otherwise there is a prime p such that $p \mid q$ and $p \mid mn$, which, since p is prime, implies that $p \mid m$ or $p \mid n$. Hence $(m, q) = 1$ or $(n, q) = 1$ and $\chi(mn) = 0 = \chi(m)\chi(n)$.

Periodicity: If $(m, q) = 1$, then $\chi(m) = \chi^*(m)$. Then for all $n \equiv m \pmod{q}$, we automatically have $n \equiv m \pmod{d}$ (since $d \mid q$) and $(n, q) = 1$. Hence $\chi(m) = \chi(n)$. If $(m, q) \neq 1$ then for all $n \equiv m \pmod{q}$, we also have $(n, q) \neq 1$ and hence $\chi(m) = \chi(n)$. □

Definition 46. *Let χ be a Dirichlet character (mod q). We call d a* **quasiperiod** *of χ if $\chi(m) = \chi(n)$ when $m \equiv n$ (mod d) and $(mn, q) = 1$. The smallest quasiperiod of χ is called the* **conductor** *of χ. If χ has conductor q, we say that χ is* **primitive**.

The terminology 'primitive' comes from the observation that if χ is primitive then it can not be induced by any character of smaller conductor. Indeed, observe for the induced character given by (3.1) that if $(mn, q) = 1$, we have

$$\chi(m) = \chi(n) \iff \chi^*(m) = \chi^*(n) \iff m \equiv n \,(\text{mod } d).$$

Hence if $d < q$, χ is not primitive. A character that is not primitive is called imprimitive. We leave it as an exercise to show that any Dirichlet character is induced by a primitive character. More precisely,

Proposition 47. *Let χ be a Dirichlet character (mod q) with conductor d. Then $d \mid q$ and there exists a unique primitive character χ^* (mod d) that induces χ.*

We have the following criteria for primitivity.

Proposition 48. *Let χ be a Dirichlet character (mod q). The following statements are equivalent:*

1. *χ is primitive;*
2. *If $d \mid q$, $d < q$, there is $c \in \mathbf{Z}$ such that $c \equiv 1$ (mod d), $(c, q) = 1$, $\chi(c) \neq 1$;*
3. *If $d \mid q$, $d < q$, then for every $h \in \mathbf{Z}$, we have*

$$\sum_{\substack{a=1 \\ a \equiv h\,(d)}}^{q} \chi(a) = 0.$$

PROOF. If χ is primitive, then for each $d \mid q$ with $d < q$, we have a pair of integers m, n such that $(mn, q) = 1$, $m \equiv n \pmod{d}$ and $\chi(m) \neq \chi(n)$. We may choose an integer c such that $(c, q) = 1$ and $cm \equiv n \pmod{q}$. Then

$$\chi(n) = \chi(cm) = \chi(c)\chi(m)$$

implies that $\chi(c) \neq 1$.

Let $d \mid q$ with $d < q$. For any c such that $(c, q) = 1$ and $c \equiv 1 \ (d)$, we have

$$\sum_{\substack{a \ (q) \\ a \equiv h \ (d)}} \chi(a) = \sum_{\substack{a \ (q) \\ a \equiv h \ (d)}} \chi(ca) = \chi(c) \sum_{\substack{a \ (q) \\ a \equiv h \ (d)}} \chi(a).$$

The existence of such a c for which $\chi(c) \neq 1$ yields (3).

Suppose that χ has conductor d and assume that (3) holds. Since $\chi(1) = 1$ and

$$\sum_{\substack{a=1 \\ a \equiv 1 \ (d)}}^{q} \chi(a) = 0$$

there is some $a \equiv 1 \pmod{d}$ such that $\chi(a) \neq 1$. But since χ has conductor d, we have $\chi(a) = \chi(1) = 1$, a contradiction. □

Example 49. *The two Dirichlet characters mod 4 are: the principal character*

$$\chi_0(n) = \begin{cases} 1 & (n, q) = 1, \\ 0 & (n, 1) \neq 1 \end{cases}$$

and the odd character

$$\chi_4(n) = \begin{cases} 1 & n \equiv 1 \ (4), \\ -1 & n \equiv 3 \ (4), \\ 0 & n \ \textit{even} \end{cases}$$

$$= \begin{cases} (-1)^{\frac{n-1}{2}} & n \ \textit{odd}, \\ 0 & n \ \textit{even}. \end{cases}$$

Observe that χ_4 is primitive, and χ_0 is not.

Example 50. *The principle character*

$$\chi_0(n) = \begin{cases} 1 & (n, q) = 1, \\ 0 & (n, 1) \neq 1 \end{cases}$$

is primitive iff $q = 1$. In this case, $\chi_0(n) = 1$ for all $n \in \mathbf{Z}$ and $L(s, \chi_0) = \zeta(s)$.

3.2. Gauss sums

We will need Gauss sums

$$\tau(\chi) = \sum_{a=1}^{q} \chi(a) e(a/q),$$

where χ is a Dirichlet character (mod q) and we use the notation $e(x) = e^{2\pi i x}$. The reader can check that $\overline{\tau(\chi)} = \chi(-1)\tau(\overline{\chi})$.

We will use the Gauss sum to interpolate a (primitive) character defined on $\mathbf{Z}$ to a smooth function on $\mathbf{R}$. The idea is that the Gauss sum can be thought of as an inner product of a multiplicative character $\chi(a)$ with an additive character $e(a/q)$, which

should be reminiscent of the construction of the Γ-function and its rôle in regularizing the ζ-function.

THEOREM 18. *Let χ be a primitive Dirichlet character (mod q). Then*

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) e(an/q)$$

and $|\tau(\chi)| = \sqrt{q}$.

PROOF. Suppose first that $(n, q) = 1$. Then

$$\tau(\chi) = \sum_{a=1}^{q} \chi(an) e(an/q) = \chi(n) \sum_{a=1}^{q} \chi(a) e(an/q),$$

which is seen to be equivalent to

$$\chi(n)\tau(\overline{\chi}) = \sum_{a\,(q)} \overline{\chi}(a) e(an/q) \tag{3.2}$$

if we use that $\chi(n)^{-1} = \overline{\chi}(n)$ and $\overline{\tau(\chi)} = \chi(-1)\tau(\overline{\chi})$.

Now we consider $(n, q) > 1$. The LHS of (3.2) is 0 — since by definition $\chi(n) = 0$ if $(n, q) \neq 1$ — and we want to show that the RHS of (3.2) vanishes as well. Since $(n, q) \neq 1$, we have $\frac{n}{q} = \frac{m}{d}$ for some integers $(m, d) = 1$. Note that $d \mid q$ and $d \neq q$; indeed, $d \mid nd = mq$ and since $(m, d) = 1$, $d \mid q$ and if $d = q$, we would have $m = n$ and $(n, q) = 1$.

We have

$$\begin{aligned}
\sum_{a=1}^{q} \chi(a) e(an/q) &= \sum_{a=1}^{q} \chi(a) e(am/d) \\
&= \sum_{h=1}^{d} \sum_{\substack{a=1 \\ a \equiv h\,(d)}}^{q} \chi(a) e(am/d) \\
&= \sum_{h=1}^{d} e(hm/d) \sum_{\substack{a=1 \\ a \equiv h\,(d)}}^{q} \chi(a) \ = 0
\end{aligned}$$

since χ is primitive (recall Proposition 48).

To prove that $|\tau(\overline{\chi})| = |\overline{\tau}(\chi)| = |\tau(\chi)| = \sqrt{q}$, we use that

$$\sum_{n=1}^{q} |\chi(n)|^2 |\tau(\chi)|^2 = \sum_{n=1}^{q} \left| \sum_{a=1}^{q} \overline{\chi}(a) e(an/q) \right|^2 .$$

The LHS is $= \varphi(q)|\tau(\chi)|^2$ while, by opening the square, the RHS is

$$\sum_{a,b=1,\dots q} \overline{\chi}(a)\chi(b) \sum_{n=1}^{q} e((a-b)n/q) = q \sum_{a=1}^{q} |\chi(a)|^2 = q\varphi(q).$$

Hence $|\tau(\chi)| = \sqrt{q}$. □

From the proof we note that for an arbitrary Dirichlet character χ (q), the identity (3.2) holds whenever $(n, q) = 1$. We used that χ was primitive to extend this identity to all $n \in \mathbf{Z}$ and deduce that $\tau(\chi) \neq 0$. For imprimitive characters, this is not necessarily true; we leave it as an exercise to show that for an imprimitive Dirichlet character χ (mod q) of conductor d induced by χ^* (mod d), we have

$$\tau(\chi) = \mu(q/d)\chi^*(q/d)\tau(\chi^*).$$

THEOREM 19 (Twisted Poisson summation formula). *Let χ be a primitive character mod q and let $f \in \mathscr{S}(\mathbf{R})$. Then*

$$\sum_{n\in\mathbf{Z}} \chi(n)f(n) = \frac{1}{\tau(\overline{\chi})} \sum_{n\in\mathbf{Z}} \overline{\chi}(n)\widehat{f}(n/q).$$

PROOF. The Poisson summation formula can be applied to $g(n) := \chi(n)f(n)$, seen as a function on $\mathbf{R}$. We have

$$\widehat{g}(t) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) \int_{\mathbf{R}} f(x)e((t + a/q)x)dx = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a)\widehat{f}(t + a/q).$$

Then

$$\begin{aligned}
\sum_{n\in\mathbf{Z}} g(n) = \sum_{n\in\mathbf{Z}} \widehat{g}(n) &= \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) \sum_{n\in\mathbf{Z}} \widehat{f}((qn + a)/q) \\
&= \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(qn + a) \sum_{n\in\mathbf{Z}} \widehat{f}((qn + a)/q) \\
&= \frac{1}{\tau(\overline{\chi})} \sum_{n\in\mathbf{Z}} \overline{\chi}(n)\widehat{f}(n/q).
\end{aligned}$$

□

Remark 51. *Note that for $q = 1$ and $\chi = \chi_0$, we recover the usual Poisson summation formula.*

Recall that for a Dirichlet character χ, we have introduced the twisted theta series

$$\tilde{\theta}_\chi(t) = \sum_{n\in\mathbf{Z}} n^\epsilon \chi(n)e^{-\pi n^2 t}$$

where $\epsilon \in \{0, 1\}$ is determined by $\chi(-1) = (-1)^\epsilon$.

THEOREM 20. *Let χ be a primitive character mod q and $t > 0$. Then*

$$\tilde{\theta}_\chi(t) = \frac{(-i)^\epsilon \tau(\chi)}{q^{1+\epsilon}t^{1/2+\epsilon}} \tilde{\theta}_{\overline{\chi}}\left(\frac{1}{q^2 t}\right).$$

PROOF. Apply the twisted Poisson summation formula to $f(x) = x^\epsilon e^{-\pi x^2 t}$. □

3.3. Elements of the analytic theory of $L(s,\chi)$

THEOREM 21. *Let χ be a primitive character mod q, $\chi \neq \chi_0$. The function*

$$\Lambda(s,\chi) = \pi^{-(s+\epsilon)/2}\Gamma\left(\frac{s+\epsilon}{2}\right)L(s,\chi)$$

admits a holomorphic continuation to all $s \in \mathbf{C}$ that satisfies the functional equation

$$\Lambda(s,\chi) = (-i)^{\epsilon}\tau(\chi)q^{-s}\Lambda(1-s,\overline{\chi}).$$

PROOF. The function $\Lambda(s,\chi)$ converges for all $\sigma > 0$ and admits the integral representation

$$\Lambda(s,\chi) = \int_0^{\infty}\left(\sum_{n\geq 1} n^{\epsilon}\chi(n)e^{-\pi n^2 t}\right)t^{(s+\epsilon)/2}\frac{dt}{t} = \frac{1}{2}\int_0^{\infty}\tilde{\theta}_{\chi}(t)t^{(s+\epsilon)/2}\frac{dt}{t}.$$

By definition, the sum in parenthesis is $O(e^{-\pi t})$ as $t \to \infty$. The inversion formula for $\tilde{\theta}_{\chi}(t)$ given by Theorem 20 shows that it is $O\left(e^{-\pi/(q^2 t)}t^{-1/2-\epsilon}\right)$ as $t \to 0$. We conclude that the integral converges for all $s \in \mathbf{C}$ and this yields the holomorphic continuation.

The functional equation also follows from Theorem 20 (and a change of variables):

$$\begin{aligned}\int_0^{\infty}\tilde{\theta}_{\chi}(t)t^{(s+\epsilon)/2}\frac{dt}{t} &= \frac{(-i)^{\epsilon}\tau(\chi)}{q^{1+\epsilon}}\int_0^{\infty}\tilde{\theta}_{\overline{\chi}}\left(\frac{1}{q^2 t}\right)t^{(s-\epsilon-1)/2}\frac{dt}{t}\\ &= \frac{(-i)^{\epsilon}\tau(\chi)}{q^{s}}\int_0^{\infty}\tilde{\theta}_{\overline{\chi}}(t)\,t^{(1-s+\epsilon)/2}\frac{dt}{t}.\end{aligned}$$

□

Remark 52. *Note that if $\chi = \chi_0$, then $L(s,\chi) = \zeta(s)$ and*

$$\sum_{n\geq 1} n^{\epsilon}\chi(n)e^{-\pi n^2 t} = \frac{\tilde{\theta}_{\chi}(t)-1}{2}.$$

From the proof of Theorem 15, one can see that the appearance of the simple poles at $s = 0, 1$ is due to the additive correction factor of $-1/2$.

The Riemann playbook may be applied to Dirichlet L-functions. First note that

$$L(s,\chi) = \exp\left(\sum_p\sum_{r\geq 1}\frac{\chi(p)}{rp^{rs}}\right)$$

for all $s \in \mathbf{C}$ with $\sigma > 1$ — this follows from the existence of the Euler product and adapting the beginning of the proof of Theorem 5. Since the exponential is never zero, $L(s,\chi) \neq 0$ when $\sigma > 1$ and since the gamma-function is nowhere zero, the function $\Lambda(s,\chi)$ has no zeroes in the halfplane $\sigma > 1$. For χ primitive, applying the functional equation, $\Lambda(s,\chi) \neq 0$ when $\sigma < 0$. Hence all zeroes of $\Lambda(s,\chi)$ are contained in the critical strip $0 \leq \sigma \leq 1$. In this context, one also has control on the number of zeroes

of bounded height and an explicit formula, that we will express here as

$$\psi_\chi(x) = \sum_{n\leq x} \chi(n)\Lambda(n) = 1_{\chi=\chi_0} x - \sum_{\substack{L(\sigma+it,\chi)=0 \\ \sigma\in[0,1] \\ |t|\leq T}} \frac{x^s}{s} + O\left(\frac{x\log^2(qx)}{T}\right)$$

for all $x > 0$, $T \geq 2$. The first factor x on the RHS only appears if $\chi = \chi_0 = 1$, i.e., if $\Lambda(s,\chi) = 0$ has a pole at $s = 1$ — which only occurs when $L(s,\chi) = \zeta(s)$.

The **generalized Riemann hypothesis (GRH)** conjectures that if $\Lambda(s,\chi) = 0$ then $\mathrm{Re}(s) = 1/2$. Let $(a,q) = 1$. We leave it to the reader to check (using the proof of Proposition 35) that the following statements are equivalent:

(1) GRH

(2)
$$\sum_{\substack{n\leq x \\ n\equiv a\,(q)}} \Lambda(n) = \frac{x}{\varphi(q)} + O\left(\sqrt{x}(\log qx)^2\right)$$

(3)
$$\pi_{a\,(q)}(x) = \frac{\pi(x)}{\varphi(q)} + O\left(\sqrt{x}\log(qx)\right),$$

where $\pi_{a\,(q)}(x)$ is the number of primes $p \leq x$ such that $p \equiv a\,(q)$.

A central theorem in the theory of Dirichlet L-functions is the **Siegel–Walfisz theorem**, which can be seen as a refinement of both the prime number theorem and Dirichlet's theorem on arithmetic progressions.

THEOREM 22 (Siegel–Walfisz, 1936). *Let $A > 0$. There exists a constant $c > 0$ such that*

$$\pi_{a\,(q)}(x) = \frac{\pi(x)}{\varphi(q)} + O\left(xe^{-c\sqrt{\log x}}\right)$$

for $(a,q) = 1$ and $q \leq (\log x)^A$.

Remark 53. *The restriction on the range of q is due to the possible existence of 'exceptional' (or Landau–Siegel) zeroes for quadratic characters. This is a new phenomenon with respect to the theory of the Riemann zeta function. Roughly we can say that there is a constant $c > 0$ such that $\{s = \sigma + it : \Lambda(s,\chi) = 0\} \cap \left(1 - \frac{c}{q|t|}, 1\right)$ if empty if χ is complex and contains at most one element if χ is quadratic, but we can not rule out that there indeed is a point in this interval (i.e., a zero close to 1). At this point, we refer the interested reader to* [6] *and* [5].

Remark 54. *Let $q = 4$. Then*

$$\frac{\pi_{1\,(4)}(x)}{\pi(x)}, \frac{\pi_{3\,(4)}(x)}{\pi(x)} \sim \frac{1}{\varphi(4)} = \frac{1}{2}$$

indicating that odd primes are roughly equally distributed among the two arithmetic progressions. Chebyshev first observed (1853) that among 'small' primes, primes of the form $p \equiv 3$ (mod 4) seem to appear more frequently. In fact, in 1957, Leech could determine that the first $N \geq 2$ for which $\pi_{1\,(4)}(N) > \pi_{3\,(4)}(N)$ is $N = 26'861$. This does

not yet imply that the bias should persist among larger ranges of primes. For instance, Littlewood (1914) proved that the function $x \mapsto \pi_{1\,(4)}(N) > \pi_{3\,(4)}(N)$ changes sign infinitely often. Under strong (including GRH) but natural assumptions, Rubinstein and Sarnak (1994) showed that the logarithmic density of the set $\{x \geq 2 : \pi_{3\,(4)}(x) > \pi_{1\,(4)}(x)\}$ is 0.9959..., establishing what is called Chebyshev's bias for residue classes mod 4.

3.4. Exercises

Exercise 55. *We have computed in class that $f(x) = e^{-\pi x^2 t}$ has Fourier transform $\widehat{f}(u) = e^{-\pi u^2/t}$. Show that more generally if $f \in L^1(\mathbf{R})$ and $t > 0$, the scaling $g(x) = f(tx)$ has Fourier transform*

$$\widehat{g}(u) = \frac{1}{t}\widehat{f}\left(\frac{u}{t}\right).$$

Exercise 56. *We have seen in class that for a nonprincipal Dirichlet character mod q, we have the (so-called 'trivial') bound*

$$\left|\sum_{n=1}^{N} \chi(n)\right| \leq q$$

for all $N \geq 1$. A stronger bound is given by the **Pólya–Vinogradov inequality**

$$\left|\sum_{n=1}^{N} \chi(n)\right| \leq 2\sqrt{q}\log q,$$

which is used to study the distribution of quadratic residues. In this exercise, we will use Gauss sums and Poisson summation to prove a 'smooth version' of this inequality when χ is primitive. Consider the tent-function

$$f(x) = \begin{cases} 1 - |x| & |x| < 1, \\ 0 & |x| \geq 1 \end{cases}$$

seen in class.

(1) *Prove that for any primitive character χ (mod q), we have*

$$\sum_{n\in\mathbf{Z}} \chi(n) f\left(\frac{n}{N} - 1\right) = \frac{N}{\tau(\overline{\chi})} \sum_{n\in\mathbf{Z}} \overline{\chi}(n)\widehat{f}\left(\frac{nN}{q}\right).$$

(2) *Prove that whenever $N \leq q$, we have the upper bound*

$$\left|\sum_{n\in\mathbf{Z}} \chi(n) f\left(\frac{n}{N} - 1\right)\right| \leq \sqrt{q} - \frac{N}{\sqrt{q}}.$$

Hint: Recall that $\widehat{f}(0) = 1$.

Exercise 57. *Let χ be a Dirichlet character (mod q) of conductor d induced by χ^* (mod d).*

(1) *Show that*
$$L(s,\chi)=\prod_{p|q}(1-\chi^*(p)p^{-s})L(s,\chi^*).$$
(2) *Show that* $\overline{\tau(\chi)}=\chi(-1)\tau(\overline{\chi})$.
(3) *Show that for a principal character* χ_0 *of modulus* n, *we have* $\tau(\chi_0)=\mu(n)$.
(4) *Show that given two characters* χ_1, χ_2 *modulo* q_1, *respectively* q_2, *with* $(q_1,q_2)=1$, *we have*
$$\tau(\chi_1\chi_2)=\tau(\chi_1)\tau(\chi_2)\chi_1(q_2)\chi_2(q_1).$$
(5) *Show that*
$$\tau(\chi)=\mu(q/d)\chi^*(q/d)\tau(\chi^*).$$

Exercise 58. *Let* $(a,q)=1$. *Show that GRH is equivalent to*
$$\sum_{\substack{n\leq x\\ n\equiv a\,(q)}}\Lambda(n)=\frac{x}{\varphi(q)}+O\left(\sqrt{x}\log^2(qx)\right).$$

Exercise 59. *In this exercise, we discuss* **Chebyshev's bias**. *For this, we want to understand how* $\pi_{a\,(q)}(x)$ *depends on* a *(mod* q*).*

(1) *We have seen several instances in which to understand* $\pi(x)$, *it is easier to work with the (Chebyshev) functions* $\psi(x)=\sum_{n\leq x}\Lambda(n)$ *and* $\theta(x)=\sum_{p\leq x}\log p$. *Show that*
$$\sum_{p\leq x}\chi(p)=\frac{\theta_\chi(x)}{\log x}+\int_2^x\frac{\theta_\chi(u)}{u(\log u)^2}du.$$
(2) *Show that*
$$\theta_\chi(x)=\psi_\chi(x)-\sqrt{x}1_{\chi^2=1}+O\left(\frac{\sqrt{x}}{\log x}\right)$$
(3) *Deduce from* (1) *and* (2) *that for* $(a,q)=1$,
$$\pi_{a\,(q)}(x)=-\frac{\sqrt{x}}{\log x\varphi(q)}\sum_{\substack{\chi\,(q)\\ \chi^2=1}}\overline{\chi}(a)$$
$$+\frac{1}{\log x\varphi(q)}\sum_{\chi\,(q)}\overline{\chi}(a)\left(\psi_\chi(x)+\log(x)\int_2^x\frac{\psi_\chi(u)}{u\log^2 u}du\right)+O\left(\frac{\sqrt{x}}{\log^2(x)}\right).$$
(4) *In the study of Chebyshev's bias, it turns out that the first term on the RHS is responsible for the bias. Show that*
$$c(a,q):=\sum_{\substack{\chi\,(q)\\ \chi^2=1}}\overline{\chi}(a)=\sum_{\substack{b\in(\mathbf{Z}/q\mathbf{Z})^\times\\ b^2\equiv a\,(q)}}1.$$
Note that $c(1,4)=1$ *whereas* $c(3,4)=0$.

CHAPTER 4

Modular forms

In the 19th century, the systematic study of modular forms was motivated by the theory of elliptic functions and elliptic curves — until Dedekind eventually provided an independent foundation for the theory of modular forms at the end of the 19th century. Over the first two sections, we sketch a rapid overview of this history, meant to motivate some of the objects we will encounter in the theory of modular forms. We follow the viewpoint of Dedekind afterwards.

4.1. Elliptic functions and Eisenstein series

We consider the ellipse $E : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ and start with the case where $a = b = 1$, i.e., the unit circle given as an algebraic curve by $y = \sqrt{1 - x^2}$. In this case, computing the arc length of a segment of E leads to the elementary integral

$$\int_0^x \sqrt{1 + y'(t)^2}dt = \int_0^x \sqrt{1 + \frac{t^2}{1 - t^2}}dt = \int_0^x \frac{dt}{\sqrt{1 - t^2}} = \operatorname{asin}(x).$$

We may further view this integral for $z \in \mathbf{C}$. Its global inverse is $\sin(z)$ and we have

$$z = \int_0^{\sin(z)} \frac{dt}{\sqrt{1 - t^2}}.$$

Varying the path of integration over the complex plane (in particular with respect to the points ± 1), we see that the multivaluedness of the integral is reflected in the periodicity of the function $\sin(z + 2\pi) = \sin(z)$. For an ellipse, the arc length (or even the circumference) computation does not lead to an elementary integral anymore. The consideration of such (elliptic) integrals was popular in the 18th and 19th century. In 1827, Abel and Jacobi independently observed that (more complicated) 'elliptic integrals' admit a global inverse that is doubly-periodic.

Definition 60. *A function $f : \mathbf{C} \to \mathbf{C}$ is* **doubly-periodic** *if there are two $\mathbf{R}$-linearly independent periods ω_1, $\omega_2 \in \mathbf{C}$ such that*

$$f(z + \omega_1) = f(z + \omega_2) = f(z).$$

The linear independence of ω_1 and ω_2 over $\mathbf{R}$ means that $\omega_1 \notin \mathbf{R}\omega_2$. Let

$$\tau := \frac{\omega_1}{\omega_2}.$$

Then $\tau \in \mathbf{H} = \{z = x + iy \in \mathbf{C} : y > 0\}$ or $\tau \in \overline{\mathbf{H}} = \{x + iy : y < 0\}$.

Convention: Since

$$\mathrm{Im}\left(\frac{1}{\tau}\right) = \mathrm{Im}\left(\frac{\overline{\tau}}{|\tau|^2}\right) = -\frac{\mathrm{Im}(\tau)}{|\tau|^2},$$

we may assume that $\tau \in \mathbf{H}$ (up to exchanging ω_1 and ω_2). In this way, we fix $\{\omega_1, \omega_2\}$ to be an ordered basis for $\mathbf{C}$ (seen as a 2-dimensional vector space over $\mathbf{R}$). To this ordered basis, we associate the **period lattice**

$$\Lambda = \langle \omega_1, \omega_2 \rangle = \mathbf{Z}\omega_1 \oplus \mathbf{Z}\omega_2 = \{m\omega_1 + n\omega_2 : m, n \in \mathbf{Z}\}.$$

It is easy to check that Λ is an abelian group with respect to addition. If f is doubly-periodic with respect to the ordered basis $\{\omega_1, \omega_2\}$, then it is Λ-invariant, i.e., $f(z+\lambda) = f(z)$ for all $\lambda \in \Lambda$. In particular, f is completely determined by the values it takes in the 'fundamental domain'

$$\{s\omega_1 + t\omega_2 : s, t \in [0, 1)\}.$$

If f is holomorphic, then it is bounded and hence, by Liouville's theorem, a constant function.

Definition 61. *An* **elliptic function** *is a meromorphic function $f : \mathbf{C} \to \mathbf{C}$ that is Λ-periodic for some (period) lattice Λ.*

The first examples of elliptic functions are global inverses of elliptic integrals following Abel and Jacobi. We are going to consider instead 'natural' constructions of elliptic functions.

For instance, we can construct elliptic functions by 'averaging'. Let f be a 'nice' meromorphic function, and consider

$$F(z) = \sum_{\lambda \in \Lambda} f(z + \lambda).$$

Here, 'nice' means that the absolute convergence of F is guaranteed.[1] For example,

$$F(z) = \sum_{\lambda \in \Lambda} \frac{1}{(z - \lambda)^k}$$

for $k > 2$ defines an elliptic function. (Observe that $\exp(F(z))$ is an example of a function that is doubly-periodic but not elliptic.) For $k = 2$, the integral does not converge. An important modification of the previous example is the **Weierstrass $\wp$-function** given by

$$\wp(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda^*} \left(\frac{1}{(z - \lambda)^2} - \frac{1}{\lambda^2} \right),$$

where $\Lambda^* = \Lambda - \{(0, 0)\}$. The function $\wp(z)$ is holomorphic on $\mathbf{C} \setminus \Lambda$, has a double pole at each $z = \lambda \in \Lambda$, and is Λ-invariant. With a bit of complex analysis, one can

[1] This is the same argument at the basis of the Poisson summation formula.

show that the field of elliptic functions for Λ is $\mathbf{C}(\wp, \wp')$, where $\wp'$ can be obtained by differentiating term-by-term, i.e.,

$$\wp'(z) = -2\sum_{\lambda\in\Lambda} \frac{1}{(z-\lambda)^3}.$$

An important family of series related to $\wp$ are the **Eisenstein series**

$$G_k(\Lambda) = \sum_{\lambda\in\Lambda^*} \frac{1}{\lambda^k}.$$

Since $\lambda \in \Lambda \iff -\lambda \in \Lambda$, $G_k(\Lambda) = 0$ whenever k is odd.

Proposition 62. *$G_k(\Lambda)$ is absolutely convergent for $k > 2$.*

PROOF. We give a 'geometry of numbers' proof: We rewrite the series as

$$G_k(\Lambda) = \sum_{n\geq 1}\sum_{\lambda\in\Omega_n} \frac{1}{\lambda^k},$$

with

$$\Omega_n = \{\lambda = l\omega_1 + m\omega_2 \in \Lambda : |l| + |m| = n.\}$$

Note that $|\Omega_1| = 4$ are exactly the 4 lattice points $\pm\omega_1$, $\pm\omega_2$ and $|\Omega_n| = n|\Omega_1| = 4n$.

For each $\lambda \in \Omega_n$, we have $|\lambda| \geq nd$, where d is the shortest distance between 0 and a side of the parallelogram with vertices $\pm\omega_1$, $\pm\omega_2$. Then

$$|G_k(\Lambda)| \leq \sum_{n=1}^{\infty} \frac{|\Omega_n|}{(nd)^k} = \frac{4}{d^k}\sum_{n\geq 1}\frac{1}{n^{k-1}}$$

and this last series converges iff $k > 2$. □

Proposition 63. *At $z = 0$, the Laurent series of $\wp(z)$ is given by*

$$\wp(z) = \frac{1}{z^2} + \sum_{k\geq 1}(k+1)G_{k+2}(\Lambda)z^k.$$

PROOF. Let $\lambda \in \Lambda^* = \Lambda - \{(0,0)\}$. For $|z| < |\lambda|$, we have the geometric series representation

$$\frac{1}{(\lambda - z)} = \frac{1}{\lambda}\sum_{k\geq 0}\left(\frac{z}{\lambda}\right)^k.$$

Differentiating, we find

$$\frac{1}{(z-\lambda)^2} = \sum_{k\geq 1} k\frac{z^{k-1}}{\lambda^{k+1}} = \sum_{k\geq 0}(k+1)\frac{z^k}{\lambda^{k+2}}.$$

Finally, summing over Λ^*, we obtain

$$\sum_{\lambda\in\Lambda^*}\left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\right) = \sum_{k\geq 1}(k+1)\left(\sum_{\lambda\in\Lambda^*}\frac{1}{\lambda^{k+2}}\right)z^k.$$

□

4.2. The moduli space of complex tori

An elliptic function induces a meromorphic function $f : \mathbf{C}/\Lambda \to \mathbf{C}$, where $\mathbf{C}/\Lambda$ can be seen

- algebraically as an abelian (quotient) group,
- geometrically as a compact Riemann surface,
- or yet as a complex elliptic curve, i.e., an abelian variety of dimension 1. (In fact, the mapping $\mathbf{C} \to \hat{\mathbf{C}} \times \hat{\mathbf{C}}$, $z \mapsto (\wp(z), \wp'(z))$ induces a bijection between $\mathbf{C}/\Lambda$ and the algebraic curve
$$y^2 = 4x^3 - g_2 x - g_3,$$
where $g_2 = 60 \cdot G_4(\Lambda)$, $g_3 = 140 \cdot G_6(\Lambda)$ and one moreover has $g_2^3 - 27g_3^2 \neq 0$ meaning that the curve has no cusps, self-intersections, or isolated points.)

As complex tori, one can show that $\mathbf{C}/\Lambda$ and $\mathbf{C}/\Lambda'$ are isomorphic if and only if $\Lambda' = \alpha\Lambda$ for some $\alpha \in \mathbf{C}^*$. If we wish to classify complex tori up to isomorphism, this fact leads us to a natural parametrization via points $\tau \in \mathbf{H}$: following our convention of ordered bases,

$$\Lambda = \langle \omega_1, \omega_2 \rangle = \omega_2 \langle \tau, 1 \rangle =: \omega_2 \Lambda_\tau \tag{4.1}$$

implies that $\mathbf{C}/\Lambda$ is isomorphic to $\mathbf{C}/\Lambda_\tau$ with $\tau \in \mathbf{H}$. Now we further observe that

Proposition 64. *Let $\tau, \tau' \in \mathbf{H}$. The complex tori $\mathbf{C}/\Lambda_\tau$ and $\mathbf{C}/\Lambda_{\tau'}$ are isomorphic iff*

$$\tau' = \frac{a\tau + b}{c\tau + d} \quad \text{for some } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}).$$

Moreover, we then have $\Lambda_\tau = (c\tau + d)\Lambda_{\tau'}$.

Proof. Let $\Lambda_{\tau'} = \alpha\Lambda_\tau$. This leads to the linear relations

$$\begin{cases} \tau' &= \alpha(a\tau + b) \\ 1 &= \alpha(c\tau + d) \end{cases} \qquad \text{or} \qquad \begin{pmatrix} \tau' \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha\tau \\ \alpha \end{pmatrix}$$

for some integers $a, b, c, d \in \mathbf{Z}$, and in fact, we see that this matrix is invertible, i.e., $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbf{Z})$. Further, the above linear relations yield

$$\tau' = \frac{a\tau + b}{c\tau + d}$$

and we can compute that

$$\mathrm{Im}(\tau') = \frac{(ad - bc)\mathrm{Im}(\tau)}{|c\tau + d|^2} = \det(A)^2 \mathrm{Im}(\tau').$$

From the first equality, we see that $\det(A) > 0$ and from the second equality we can conclude that $\det(A) = 1$.

Conversely, we start by observing that

$$\Lambda_{\tau'} = \langle \tau', 1 \rangle = \frac{1}{c\tau + d} \langle a\tau + b, c\tau + d \rangle .$$

It is an easy exercise to check that $\langle a\tau + b, c\tau + d \rangle = \langle \tau, 1 \rangle$. □

Proposition 65. *The map* $\mathrm{SL}_2(\mathbf{Z}) \times \mathbf{H} \to \mathbf{H}$ *given by*

$$(A, z) \mapsto A.z := \frac{az+b}{cz+d}$$

defines a group action of $\mathrm{SL}_2(\mathbf{Z})$ *on* $\mathbf{H}$.

PROOF. From the previous proof, we have the identity

$$\mathrm{Im}(A.z) = \frac{\det(A)\mathrm{Im}(z)}{|cz+d|^2}.$$

Hence if $\det(A) = 1$ and $z \in \mathbf{H}$, $A.z \in \mathbf{H}$. Clearly the identity acts trivially and a direct computation verifies that $(AB).z = A.(B.z)$. □

Clearly, $A.z = -A.z$ and so the action quotients through the group

$$\mathrm{PSL}_2(\mathbf{Z}) = \mathrm{SL}_2(\mathbf{Z})/\langle -I \rangle = \mathrm{SL}_2(\mathbf{Z})/\{I, -I\}.$$

THEOREM 23. *There is a one-to-one correspondence between the* **moduli space of complex tori**, *i.e., the space of isomorphism classes of complex tori and the quotient space* $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}$.

PROOF. Let $\mathcal{M}$ denote the space of isomorphism classes of complex tori. Consider the map $\tau \in \mathbf{H} \mapsto \mathbf{C}/\Lambda_\tau \in \mathcal{M}$. With (4.1), we see that each period lattice can be scaled to a lattice of the form Λ_τ so that this map is surjective. Proposition 64 then shows that the complex $\mathbf{C}/\Lambda_\tau$ and $\mathbf{C}/\Lambda_{\tau'}$ are in the same isomorphism class iff τ and τ' represent the same point in $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}$. □

Accordingly, the group $\mathrm{PSL}_2(\mathbf{Z})$ is called the **modular group**. The quotient space $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}$ can be visualized in terms of a choice of a **fundamental domain** in $\mathbf{H}$, similarly to the situation of the torus $\mathbf{C}/\Lambda$. More precisely, we understand a fundamental domain $D \subset \mathbf{H}$ to be a connected set that contains a unique representative of each orbit $\mathrm{PSL}_2(\mathbf{Z}).\tau$ and that tessellates $\mathbf{H}$ under the action of $\mathrm{PSL}_2(\mathbf{Z})$, i.e., if $\gamma.D \cap D \neq \emptyset$ then $\gamma = e$.

THEOREM 24. *The set*

$$\mathcal{T} = \{x + iy \in \mathbf{H} : -1/2 < x \geq 1/2,\ |z| > 1 \ \text{ or } \ |z| = 1,\ 0 \leq x < 1/2\}$$

is a fundamental domain for $\mathrm{PSL}_2(\mathbf{Z})$ *and* $\mathrm{PSL}_2(\mathbf{Z})$ *is generated by the two transformations*

$$T(z) = z + 1, \quad S(z) = -\frac{1}{z} = \frac{-\overline{z}}{|z|^2} = \frac{-x+iy}{x^2+y^2}.$$

Remark 66. *Note that* S *and* T *correspond to the simplest examples of nontrivial matrices in* $\mathrm{SL}_2(\mathbf{Z})$*: fixing the vector* $\left(\begin{smallmatrix} a \\ c \end{smallmatrix}\right) \in \mathbf{Z}^2$*, one can always find a (nonunique) vector* $\left(\begin{smallmatrix} b \\ d \end{smallmatrix}\right)$ *such that* $ad - bc = 1$*. The matrices* $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ *and* $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ *arise as the simplest solutions of applying this process to the standard vectors* e_1, e_2.

PROOF OF THEOREM 24. For ease of notation, we write $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$. Fix $z = x + iy \in \mathbf{H}$. Recall the identity

$$\mathrm{Im}(\gamma.z) = \frac{y}{|cz+d|^2} = \frac{y}{(cx+d)^2 + (cy)^2}.$$

Then $\mathrm{Im}(\gamma.z) \geq \mathrm{Im}(z)$ iff $|cz+d|^2 \leq 1$ and for fixed z, there are only a finite number of possible $c, d \in \mathbf{Z}$ for which this holds. Hence there exists a point $z_0 \in \Gamma.z$ for which $\mathrm{Im}(z_0) \geq \mathrm{Im}(\gamma.z_0)$. Up to replacing z_0 by $T^n.z_0 = z_0 + n$, for some adequate $n \in \mathbf{Z}$, we may assume that $\mathrm{Re}(z_0) \in (-1/2, 1/2]$. Further, since $\mathrm{Im}(z_0) \geq \mathrm{Im}(S.z_0) = \frac{y_0}{|z_0|^2}$, we have that $|z_0| \geq 1$. If $|z_0| > 1$, we are done, since $z_0 \in \mathcal{T}$, while if $|z_0| = 1$, we may replace z_0 by $S.z_0 \in \mathcal{T}$. We conclude that each orbit $\Gamma.z_0$ contains a point in the 'ideal triangle' $\mathcal{T}$. (And, in fact, we have shown that each orbit $\langle S, T\rangle .z$ does.)

We next show that if $\gamma\mathcal{T} \cap \mathcal{T} \neq \emptyset$ then $\gamma = e$. Let $z \in \mathcal{T}$ and suppose there is $\gamma \in \Gamma$ such that $\gamma.z \in \mathcal{T}$. Up to replacing γ by γ^{-1}, we may assume that

$$\mathrm{Im}(\gamma.z) \geq \mathrm{Im}(z) \iff |cz+d|^2 \leq 1.$$

Since $z \in \mathcal{T}$, we have $y \geq \sqrt{3}/2$ and hence

$$\frac{\sqrt{3}|c|}{2} \leq |c|y \leq |cz+d| \leq 1 \implies |c| \leq \frac{2}{\sqrt{3}} \approx 1.155.$$

If $|c| = 1$, then $|z \pm d| \leq 1$, which is impossible. If $c = 0$, then $\gamma = T^n$ for some $n \in \mathbf{Z}$. But if $n \neq 0$, $\gamma.z \notin \mathcal{T}$ and hence $\gamma = e$. □

4.3. Elliptic modular forms

In this section, we write $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$.

In the previous section, we have seen that the moduli space of complex tori is parametrized by $\Gamma\backslash\mathbf{H}$ building on the fact that every lattice Λ is isomorphic to a 'representative lattice' $\Lambda_\tau = \langle \tau, 1\rangle$ for some $\tau \in \mathbf{H}$. To each lattice Λ, we have seen that we can associate the Eisenstein series

$$G_k(\Lambda) = \sum_{\lambda \in \Lambda^*} \frac{1}{\lambda^k}$$

that converges absolutely whenever $k > 2$. This leads us to define

$$G_k(\tau) := G_k(\Lambda_\tau) = \sum_{\lambda \in \Lambda_\tau^*} \frac{1}{\lambda^k} = \sum_{\substack{m,n \in \mathbf{Z} \\ (m,n) \neq (0,0)}} \frac{1}{(m\tau + n)^k}$$

as a function of one complex variable.

Proposition 67. *The function $G_k(\tau)$ is analytic.*

Proof. Observe from the definition that $G_k(\tau + 1) = G_k(\tau)$. Since

$$\tau = x + iy \mapsto q = e(\tau) = e^{-2\pi y} e(x)$$

gives a biholomorphism between

$$\mathcal{C}_\infty = \{x + iy \in \mathbf{C} : x \in [0,1),\ y > 0\} \longleftrightarrow \mathbf{D}^* = \{|q| < 1\} \setminus \{0\},$$

we see that $G_k(q)$ is analytic in the disk, and the Eisenstein series admits a power series expansion

$$G_k(\tau) = \sum_{n \in \mathbf{Z}} a_n q^n. \tag{4.2}$$

□

We call the coefficients a_n in (4.2) the **Fourier coefficients** of G_k. Indeed, since $G_k(\tau+1) = G_k(\tau)$, the Eisenstein series admits the Fourier series expansion

$$G_k(\tau) = \sum_{n\in\mathbf{Z}} a_n(y)e(nx)$$

and comparing with the generating series above, we see that $a_n(y) = a_n e^{-2\pi ny}$. The Fourier coefficients of G_k encode rich arithmetic information, as illustrated by the following exercise (cf. exercise sheet 6).

Exercise 68. *We will consider the* **divisor functions** $\sigma_s(n) = \sum_{d|n} d^s$.

(1) *For $k > 2$ even, show that*

$$G_k(\tau) = 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{\Gamma(k)}\sum_{n\geq 1}\sigma_{k-1}(n)q^n.$$

(2) *For $\mathrm{Re}(s) > \max\{1, \mathrm{Re}(a)+1\}$, show that*

$$\sum_{n=1}^{\infty}\frac{\sigma_a(n)}{n^s} = \zeta(s)\zeta(s-a).$$

(3) *Extending such identities, Ramanujan proved that*

$$\sum_{n=1}^{\infty}\frac{\sigma_a(n)\sigma_b(n)}{n^s} = \frac{\zeta(s)\zeta(s-a)\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)}.$$

Prove that this implies that $\zeta(1+it) \neq 0$ for all $t \in \mathbf{R}$ (i.e., the prime number theorem!).

The relations from Proposition 64

$$\tau' = \gamma.\tau = \frac{a\tau+b}{c\tau+d} \quad (\gamma\in\Gamma) \iff \mathbf{C}/\Lambda_\tau \cong \mathbf{C}/\Lambda_{\tau'} \quad \text{with } \Lambda_\tau = (c\tau+d)\Lambda_{\tau'}$$

yield the following **modular transformation** of the Eisenstein series:

$$G_k(\tau') = G_k((c\tau+d)^{-1}\Lambda_\tau) = \sum_{\lambda\in(c\tau+d)^{-1}\Lambda_\tau^*}\frac{1}{\lambda^k} = (c\tau+d)^k G_k(\Lambda_\tau) = (c\tau+d)^k G_k(\tau)$$

for each $\gamma\in\Gamma$ and $\tau\in\mathbf{H}$. Moreover, the modular transformation is compatible with the group action: a direct computation shows

$$G_k(\gamma\gamma'.\tau) = (c^*\tau+d^*)^k G_k(\tau) = (c(\gamma'.\tau)+d)^k(c'\tau+d')^k G_k(\tau).$$

Definition 69. *Let $k\in\mathbf{Z}$. A* **modular form of weight** k **for** $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$ *is a holomorphic function $f:\mathbf{H}\to\mathbf{C}$ with modular transformation $f(\gamma z) = (cz+d)^k f(z)$ for all $\gamma\in\Gamma$, $z\in\mathbf{H}$, and which is 'holomorphic at ∞', i.e., for which*

$$f(z) = \sum_{n\in\mathbf{Z}} a_n q^n \quad \text{with } a_n = 0 \text{ for all } n < 0.$$

Remark 70. *Observe that*

$$f(z) = \sum_{n\geq 0} a_n q^n = \sum_{n\geq 0} a_n e^{-2\pi n y} e(nx) \to a_0$$

as $y \to \infty$. This is where the terminology 'holomorphic at ∞' comes from. (Consider what happens if $a_n \neq 0$ with $n < 0$.) Observe also that the modular transformation implies that the only modular form of odd weight k is $f = 0$. Indeed,

$$f((-I).z) = (-1)^k f(z) = f(z) \implies f(z) = 0.$$

The set of all modular forms of weight k forms a vector space (over $\mathbf{C}$) denoted $M_k(\Gamma)$. Clearly, $G_k \in M_k(\Gamma)$ for $k \geq 4$ even.

Proposition 71. *We have the direct sum decomposition*

$$M_k(\Gamma) = \langle G_k \rangle \oplus S_k(\Gamma),$$

where $S_k(\Gamma)$ is the subspace of all modular forms of weight k that 'vanish at ∞', i.e., for which $a_0 = 0$.

Proof. Let $f \in M_k(\Gamma)$ be given by $f(z) = \sum_{n\geq 0} a_n q^n$. Then $g \coloneqq f - \frac{a_0}{2\zeta(k)} G_k \in S_k(\Gamma)$ and the representation $f = g + \frac{a_0}{2\zeta(k)} G_k$ is unique. □

Remark 72. *We will see in the next section that $M_k(\Gamma)$ can be equipped with a natural inner product with respect to which $S_k(\Gamma)$ is the orthogonal complement of $\langle G_k \rangle$.*

A modular form $f \in S_k(\Gamma)$ is called a **cusp form** (*Spitzenform* in German). We exhibit an example of a cusp form. In view of the relation between complex tori and elliptic curves, we consider

$$\Delta(\tau) := g_2^3 - 27 g_3^2 = (60 G_4(\tau))^3 - 27(140 G_6(\tau))^2.$$

Note that by construction $\Delta(\tau) \neq 0$ is guaranteed for all $\tau \in \mathbf{H}$. Moreover, $\Delta(\tau)$ is holomorphic and $\Delta(\gamma.\tau) = (c\tau + d)^{12} \Delta(\tau)$. Further, looking at the expansion of Eisenstein series, we have

$$\Delta(\tau) = (120\zeta(4))^3 - 27(280\zeta(6))^2 + \sum_{n\geq 1} \tau(n) q^n = \sum_{n\geq 1} \tau(n) q^n.$$

(The Fourier coefficients $\tau(n)$ are still actively studied. Ramanujan made conjectures about these coefficients that influenced an important part of the development of number theory in the 20th century.)

Theorem 25. *The vector spaces $M_k(\Gamma)$ are finite dimensional.*

Proof. Recall that by Remark 70, $M_k(\Gamma) = \{0\}$ if k is odd, hence we may assume that k is even.

Let $k = 0$. A modular form $f \in M_0(\Gamma)$ is necessarily holomorphic. Being Γ-invariant, $f(q)$ is completely determined by its values on an ideal triangle of the unit disk, hence bounded, and by Liouville's theorem, constant. The constant is given by

$$f(z) = \sum_{n\geq 0} a_n q^n = a_0.$$

We conclude that $M_k(\Gamma) \cong \mathbf{C}$.

Let $k < 0$ and $f \in M_k(\Gamma)$. Then $f^{12}\Delta^{|k|} = \sum a_n q^n \in M_0(\Gamma)$. Then $f^{12}\Delta^{|k|} = a_0$ and since $\Delta \in S_{12}(\Gamma)$, we have $a_0 = 0$. Hence $M_k(\Gamma) = \{0\}$.

Let $k \geq 4$. In view of the previous theorem, $\dim M_k(\Gamma) = \dim S_k(\Gamma) + 1$. Consider the linear map $M_{k-12} \to S_k$, $f \mapsto f\Delta$. Since $\Delta \neq 0$, this map is invertible with inverse f/Δ and this shows that $S_k(\Gamma) \cong M_{k-12}(\Gamma)$ and $\dim M_k(\Gamma) = \dim M_{k-12}(\Gamma) + 1$ from which we conclude.

For the case $k = 2$, we refer to exercise sheet 6 for an 'elementary' proof that $M_2(\Gamma) = \{0\}$ and include below a sketch of argument using differential geometry. Let $f \in M_2(\Gamma)$. Then $f(z)dz$ is a holomorphic differential closed 1-form on $\Gamma\backslash\mathbf{H}$ and $h(\gamma) = \int_\tau^{\gamma.\tau} f(z)dz$ defines a group homomorphism on Γ. Since $\Gamma = \langle S, T\rangle = \langle S, TS\rangle$, which are both transformations of finite order ($S^2 = (TS)^3 = e$), the modular group can be identified algebraically with the free product $C_2 * C_3$ of finite cyclic groups. Since each homomorphism $C_n \to \mathbf{C}$ is necessarily trivial, we conclude that $f = 0$. □

Corollary 73. *For $k \in 2\mathbf{N}$, we have*

$$\dim M_k(\Gamma) = \begin{cases} \left\lfloor \frac{k}{12} \right\rfloor + 1 & k \not\equiv 2\,(12), \\ \left\lfloor \frac{k}{12} \right\rfloor & k \equiv 2\,(12). \end{cases}$$

Proof. For $k < 12$, we have $\dim(M_k) = 1$ for $k = 0, 4, 6, 8, 10$ and $\dim(M_k) = 0$ for $k = 2$. For $12 \leq k < 24$, the identity $\dim M_k(\Gamma) = \dim M_{k-12}(\Gamma) + 1$ implies that $\dim M_k(\Gamma) = 2$ for $k = 12, 16, 18, 20, 22$ and $\dim M_k(\Gamma) = 1$ for $k = 14$. This pattern then propagates by applying repeatedly this identity. □

Application: identities for divisor functions

Consider the one-dimensional space $M_8(\Gamma)$. Both G_8 and G_4^2 are elements of $M_8(\Gamma)$. Since the space is one-dimensional, there is $c \in \mathbf{C}^*$ such that $G_4^2 = cG_8$. We can determine c by looking at the 0-th Fourier coefficient on both sides. We have on one side

$$G_4^2 = 4\zeta(4)^2 + \sum_{n\geq 1} \dots$$

and on the other side

$$G_8 = 2\zeta(8) + \sum_{n\geq 1} \dots$$

Since the equality of the two power series implies the termwise equality of the coefficients, we conclude that $c = 2\frac{\zeta(4)^2}{\zeta(8)}$. Then equating the higher Fourier coefficients yields

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_3(n-m)$$

for each $n \geq 1$.

Consider next G_{12}, G_4^3, G_6^2, all lying in the 2-dimensional vector space $M_12(\Gamma)$. Then G_{12} can be written as a nontrivial linear combination of G_4^3 and G_6^2 and this will also yield an identity involving the divisor functions σ_{11}, σ_3 and σ_5.

Such identities can also be derived more directly, but the theory of modular forms explains why (and how) identities for σ_{k-1} become more involved for k large; as k grows, so does the dimension of the corresponding vector space.

What we are seeing at play in the previous proof and in these examples is the fact that

$$M_*(\Gamma) = \bigoplus_{k=0}^{\infty} M_k(\Gamma)$$

is a **graded ring**, i.e., if $f \in M_k(\Gamma)$, $g \in M_\ell(\Gamma)$ then $fg \in M_{k+\ell}(\Gamma)$. We call $M_*(\Gamma)$ the **ring of modular forms**.

THEOREM 26. *We have $M_*(\Gamma) \cong \mathbf{C}[G_4, G_6]$, i.e., the ring of modular forms is freely generated by G_4 and G_6.*

SKETCH OF PROOF. Clearly, $G_4^m G_6^n \in M_k(\Gamma)$ whenever $4m + 6n = k$. We show that G_4 and G_6 are algebraically independent; for this it suffices to show that there is no $\alpha \in \mathbf{C}$ for which $G_6^2 = \alpha G_4^3$. Indeed, otherwise the function $f := G_6/G_4$ would transform as a modular form of weight 2, be holomorphic at the cusp and holomorphic on $\mathbf{H}$, since $f^2 = \alpha G_4$ is holomorphic. But we have seen that $M_2(\Gamma) = \{0\}$, hence this is impossible. To conclude, one checks that the number of pairs $(m, n) \in \mathbf{N} \times \mathbf{N}$ such that $4m + 6n = k$ is exactly $\dim M_k(\Gamma)$. □

4.4. Petersson inner product

We can equip $M_k(\Gamma)$ with an inner product, called the **Petersson inner product**: for $f, g \in M_k(\Gamma)$, let

$$\langle f, g \rangle := \int_{\mathcal{T}} f(z)\overline{g(z)}y^{k-2}dxdy.$$

Proposition 74. *$\langle \cdot, \cdot \rangle$ defines a Hermitian inner product on $M_k(\Gamma)$.*

PROOF. We show that $\langle f, g \rangle$ is well defined. It is easy to then check it is a Hermitian inner product. Let $F(z) = f(z)\overline{g(z)}y^k$, with $y = \mathrm{Im}(z)$. Then for each $\gamma \in \Gamma$,

$$F(\gamma.z) = f(\gamma.z)\overline{g(\gamma.z)}\mathrm{Im}(\gamma.z)^k = f(z)(cz+d)^k\overline{g(z)(cz+d)^k}y^k|cz+d|^{-2k} = F(z).$$

It suffices then to check that $d\mu(z) := \frac{dxdy}{y^2}$ is Γ-invariant. For this, we use that $\Gamma = \langle S, T \rangle$ as in Theorem 24. Clearly, $d\mu(z+1) = d\mu(z)$, while

$$d\mu(S.z) = d\mu\left(\frac{-x}{x^2+y^2} + \frac{iy}{x^2+y^2}\right) = d\mu(u+iv) = \frac{dudv}{v^2} = \frac{dx}{dy}y^2$$

as follows from computing the Jacobian associated to $(u, v) \to (x, y)$. □

Aside: Hyperbolic geometry. Petersson's definition is motivated by the fact that the upper half-plane $\mathbf{H}$ is a model of the hyperbolic plane. For context, recall that in the geometry of curved surfaces, the Gaussian curvature, which determines the local shape of the surface, is an intrinsic (i.e., invariant under local isometries)

property of the surface (this is Gauss' celebrated theorema egregium). In the study of higher dimensional shapes (i.e., smooth manifolds), Riemann showed that the intrinsic geometry is fully determined by the choice of a Riemannian metric. Equipping $\mathbf{H}$ with the Riemannian metric $ds^2 = y^{-2}(dx^2 + dy^2)$, one obtains a model for a surface of constant negative curvature (called a hyperbolic surface).

With respect to this metric, the length of nice enough parametrized curve $c : [a, b] \to \mathbf{H}$, $c(t) = x(t) + iy(t)$ is given by

$$L(c) = \int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} dt$$

and the area of a domain $A \subset \mathbf{H}$ is given by

$$|A| = \int_A d\mu(z) = \int_A \frac{dxdy}{y^2}.$$

For example, the hyperbolic area of the ideal triangle $\mathcal{T}$ is given by

$$|\mathcal{T}| = \int_{\mathcal{T}} \frac{dxdy}{y^2} = \int_{-1/2}^{1/2} \int_{\sqrt{1-x^2}}^{\infty} \frac{dy}{y^2}\, dx = \int_{-1/2}^{1/2} \frac{dx}{\sqrt{1-x^2}} = \operatorname{asin}(x)|_{-1/2}^{1/2} = \frac{\pi}{3}.$$

THEOREM 27. *Let $k \geq 4$ even. The space $S_k(\Gamma)$ is the orthogonal complement of $\langle G_k \rangle$ in $M_k(\Gamma)$ with respect to the Petersson inner product.*

In preparation for the proof, we introduce a different normalization of our Eisenstein series G_k. Let $\Gamma_\infty = \langle T \rangle$.

Lemma 75. *There is a bijection between $\Gamma_\infty \backslash \Gamma$ and the set $\{(c, d) \in \mathbf{Z}^2 : (c, d) = 1\}/\{\pm 1\}$.*

PROOF. We obtain a surjection via the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)/\{\pm I\} \to (c, d)$. Note indeed that since the matrix has determinant 1, $(c, d) = 1$. The elements of Γ that leave the lower row invariant are of the form $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$, hence belong to Γ_∞. □

We now set

$$E_k(z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \frac{1}{(cz + d)^k} = \frac{1}{2} \sum_{(c,d)=1} \frac{1}{(cz + d)^k}.$$

We easily see that $G_k = 2\zeta(k)E_k$:

$$\sum_{n \geq 1} \sum_{(c,d)=1} \frac{1}{(cnz + dn)^k} = \sum_{n \geq 1} \sum_{(c,d)=n} \frac{1}{(cz + d)^k} = G_k(z).$$

Remark 76. *Using the Fourier expansion of G_k and properties of Bernoulli numbers (see exercise sheet) one can show that the Fourier coefficients of E_k are all rational. Together with Theorem 26, this says that the ring of modular forms is generated by modular forms with rational coefficients.*

SKETCH OF PROOF OF THEOREM 27. Let $f = \sum a_n q^n \in M_k(\Gamma)$. Then

$$\begin{aligned}\langle f, G_k\rangle &= \int_{\mathcal{T}} f(z)\overline{G_k(z)}y^k d\mu(z)\\ &= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\int_{\mathcal{T}} f(z)\overline{(cz+d)^{-k}}y^k d\mu(z)\\ &= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\int_{\gamma(\mathcal{T})} f(z)y^k d\mu(z)\\ &= \int_{\mathcal{C}_\infty} f(z)y^k\, d\mu(z) = \int_0^\infty\int_0^1 f(x+iy)dx\, y^{k-2}dy = 0\end{aligned}$$

iff $a_0 = 0$. □

4.5. Exercises

Exercise 77. *Suppose that $f : \mathbf{C} \to \mathbf{C}$ is analytic and triply-periodic, i.e.,*

$$f(z+\omega_1) = f(z+\omega_2) = f(z+\omega_3) = f(z)$$

for ω_1, ω_2, $\omega_3 \in \mathbf{C}$ pairwise linear independent over $\mathbf{R}$. We will show that f must be constant. More explicitly, we will show that the set of periods

$$\Lambda = \left\{\omega_{klm} := k\omega_1 + l\omega_2 + m\omega_3 : \substack{k,l,m\in\mathbf{Z}\\ \omega_{klm}=0 \iff k=l=m=0}\right\}$$

has a limit point at the origin.

(1) *Let $n \in \mathbf{N}$ and set $\Omega_n = \{\omega_{klm} \in \Lambda : |k|, |l|, |m| \le n\}$. Show that*

$$|\Omega_n| = (2n+1)^3.$$

(2) *Show that there is a square between two successive positive cubes; i.e., if $a \in \mathbf{N}$ there is $b \in \mathbf{N}$ such that $a^3 \le b^2 < (a+1)^3$.*

(3) *Using $(2n)^3 \le N^2 < (2n+1)^3$ for some $N \in \mathbf{N}$, apply Dirichlet's box principle to show that there are two periods ω_{abc}, $\omega_{klm} \in \Omega_n$ with the property that*

$$0 < |\omega_{abc} - \omega_{klm}| \le n^{-1/2}(|\omega_1| + |\omega_2| + |\omega_3|).$$

(4) *Conclude that the origin is a limit point of Λ.*

Exercise 78. *We will consider the* **divisor functions** $\sigma_s(n) = \sum_{d|n} d^s$.

(1) *For $k > 2$ even, show that*

$$G_k(\tau) = 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{\Gamma(k)}\sum_{n\ge1}\sigma_{k-1}(n)q^n.$$

(2) *For $\mathrm{Re}(s) > \max\{1, \mathrm{Re}(a)+1\}$, show that*

$$\sum_{n=1}^{\infty}\frac{\sigma_a(n)}{n^s} = \zeta(s)\zeta(s-a).$$

(3) *Extending such identities, Ramanujan proved that*

$$\sum_{n=1}^{\infty} \frac{\sigma_a(n)\sigma_b(n)}{n^s} = \frac{\zeta(s)\zeta(s-a)\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)}.$$

Prove that this implies that $\zeta(1+it) \neq 0$ *for all* $t \in \mathbf{R}$ *(i.e., the prime number theorem!).*

Exercise 79. *Show that the only fixed points in* $\mathbf{H}$ *of the action of* $\mathrm{PSL}_2(\mathbf{Z})$ *are* $z = i$ *and* $z = \pm\frac{1}{2} + i\frac{\sqrt{3}}{2}$. *Then show that* $\mathrm{Stab}_\Gamma(i) = \{\gamma \in \Gamma : \gamma.i = i\} = \langle z \mapsto -1/z \rangle$ *and* $\mathrm{Stab}_\Gamma(\rho) = \langle z \mapsto -1/(z+1) \rangle$ *for* $\rho = e^{\pi i/3}$.

Exercise 80. *In this exercise, we will give a simpler proof of* $M_2(\Gamma) = \{0\}$ *than what we have seen in class. We keep the notation form the previous exercise.*

(1) *Let* $f \in M_k(\Gamma)$. *Show that* $f(i) = 0$ *if* $4 \nmid k$ *and* $f(\rho) = 0$ *if* $3 \nmid k$.
(2) *Let* $f \in M_2(\Gamma)$. *Show that* $fG_4 \in M_6(\Gamma)$ *and determine* $c \in \mathbf{C}$ *such that* $fG_4 = cG_6$.
(3) *Let again* $f \in M_2(\Gamma)$. *Use* (1) *and* (2) *to show that if* $f \neq 0$ *then* $\Delta(\rho) = 0$, *a contradiction.*

CHAPTER 5

Theta series

5.1. Introduction

Recall the Jacobi theta series (actually already studied by Euler) given by

$$\theta(z) = \sum_{n \in \mathbf{Z}} q^{n^2} = 1 + 2 \sum_{n \geq 1} q^{n^2},$$

and that the meromorphic continuation of the (completed) zeta-function

$$\pi^{-s/2} \Gamma(s/2) \zeta(s) = \int_0^\infty \frac{\theta(it/2) - 1}{2} t^{s/2} \frac{dt}{t}$$

follows from the relation (Jacobi's inversion formula/Poisson summation)

$$\theta\left(\frac{i}{2t}\right) = \sqrt{t}\theta\left(\frac{it}{2}\right)$$

for all $t > 0$. This relation can be seen as a modular transformation: if we consider the map $\omega_2 : z \mapsto -\frac{1}{4z}$ in $\mathrm{PSL}_2(\mathbf{R})$, then for $z = it/2$, the above formula reads $\theta(\omega_2.z) = \sqrt{-2iz}\theta(z)$, and by analytic continuation, this holds true for all $z \in \mathbf{H}$.

Recall also that for $m \geq 2$,

$$\theta^m(z) = \sum_{n \geq 0} r_m(n) q^n,$$

where $r_m(n)$ is the number of ways n can be represented as a sum of m squares. For the moment we will concern ourselves only with $m = 2k$ even.

We will explain how θ^k can be seen as modular forms for the subgroup

$$\Gamma_0(4) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : 4 \mid c \right\} < \mathrm{SL}_2(\mathbf{Z})$$

We will later (see next section) prove that the group $\Gamma_0(4)$ is generated by the matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Remark 81. *Let T and U be the induced transformations in $\mathrm{PSL}_2(\mathbf{R})$ of the first two matrices above (respectively). As subgroups of $\mathrm{PSL}_2(\mathbf{R})$, we have $\langle T, \omega_2 \rangle = \langle T, U \rangle \cup \omega_2 \langle T, U \rangle$. Note that contrarily to $\langle T, \omega_2 \rangle$, the group $\langle T, U \rangle$ is a subgroup of the modular group.*

Evaluating θ^{2k} on these generators, we have

$$\begin{aligned}\theta^{2k}(T.z) &= \theta^{2k}(z)\\ \theta^{2k}(U.z) &= \theta^{2k}(\omega_2^{-1}T^{-1}\omega_2.z)\\ &= (-2i(T^{-1}\omega_2.z))^k\theta(T^{-1}\omega_2.z)\\ &= (2i(1+\tfrac{1}{4z}))^k(-2iz)^k\theta^{2k}(z) = (4z+1)\theta^{2k}(z)\\ \theta^{2k}(-I.z) &= \theta^{2k}(z)\end{aligned}$$

from which we can deduce that for each $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(4)$,

$$\theta^{2k}(\tfrac{az+b}{cz+d}) = \chi(d)(cz+d)^k\theta^{2k}(z), \tag{5.1}$$

where $\chi(d) = (-1)^{\frac{d-1}{2}}$ is the (unique) odd Dirichlet character mod 4 if k is odd or $\chi = \chi_0$ if k is even.

Remark 82. *Observe that if* $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(4)$ *then* $4 \mid c$ *and* $(c,d) = 1$*, and hence* $(d,4) = 1$ *and hence it suffices to consider* d *(mod 4). Moreover,*

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}\begin{pmatrix} a' & b'\\ c' & d'\end{pmatrix} = \begin{pmatrix} * & *\\ cb'+dd' & *\end{pmatrix}$$

and so if $c, c' \equiv 0$ *(mod 4), then* $cb'+dd' \equiv dd'$ *(mod 4) and hence the modular transformation (5.1) is compatible with multiplication.*

Over the course of the next section, we will see that the study of modular forms for congruence groups reduces to the study of the following type of modular forms, of which the theta series provide important examples:

Definition 83. *A* **modular form of weight** k**, level** N **and Dirichlet character** χ *(mod* N*) is a holomorphic function* $f : \mathbf{H} \to \mathbf{C}$ *that is holomorphic at each cusp and such that*

$$f(\gamma.z) = \chi(d)(cz+d)^k f(z)$$

for all $\gamma \in \Gamma_0(N)$*. The space of all such modular forms is denoted* $M_k(N,\chi)$*. if* $\chi = \chi_0$*, we write* $M_k(N) = M_k(N,\chi_0)$*.*

Observe that for k even and $\chi = \chi_0$, this definition coincides with the definition of modular forms given in the previous chapter. We will explain what 'holomorphic at each cusp' means over the course of the next chapter.

Remark 84. *With more work, one can show that for each* $\gamma \in \left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(4)$*, we have the modular transformation*

$$\theta(\gamma.z) = \overline{\varepsilon}_d\left(\frac{c}{d}\right)(cz+d)^{1/2}\theta(z),$$

where

$$\varepsilon_d = \begin{cases} 1 & d \equiv 1\,(4),\\ i & d \equiv 3\,(4),\end{cases}$$

and $\left(\frac{c}{d}\right)$ *is the extended Legendre symbol, that is the usual Legendre symbol for* d *odd and positive, and otherwise:*

$$\left(\frac{c}{d}\right) = \frac{c}{|c|}\left(\frac{c}{-d}\right) \quad (c \neq 0), \qquad \left(\frac{0}{d}\right) = \begin{cases} 1 & d = \pm 1, \\ 0 & \textit{otherwise.} \end{cases}$$

5.2. Congruence modular groups...

Let $N \in \mathbf{N}$. Besides

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : N \mid c \right\},$$

we will also consider the group

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

and the **principal congruence group of level** N

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

By definition,

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \mathrm{SL}_2(\mathbf{Z}).$$

It follows from the following result that these subgroups have finite index in $\mathrm{SL}_2(\mathbf{Z})$.

THEOREM 28. *The projection* $\mathbf{Z} \to \mathbf{Z}/N\mathbf{Z}$ *induces a surjective homomorphism* $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. *Its kernel is* $\Gamma(N)$.

PROOF. We will only prove that the homomorphism is surjective. For this, let $A \in \mathrm{SL}_2(\mathbf{Z}/N\mathbf{Z})$. We can always choose integers a, b, c, d such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \pmod{N} = A$$

but this only guarantees that

$$ad - bc \equiv 1 \pmod{N}. \tag{5.2}$$

We show that we can choose

$$\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} \pmod{N} = A$$

with $a'd' - b'c' = 1$.

Let $c' = c$. By (5.2), $(c', d, N) = 1$. We first show that we can choose $d' \equiv d \pmod{N}$ such that $(c', d') = 1$. Observe that if $c' = N$, then $(c', d) = 1$ and so we can simply choose $d' = d$. Suppose then that $c' \neq N$ and define

$$P = \prod_{\substack{p \mid c' \\ p \nmid N}} p > 1.$$

Then $(P, N) = 1$ so that by the Chinese remainder theorem, the system of congruence equations

$$\begin{cases} x \equiv 1 \pmod{P} \\ x \equiv d \pmod{N} \end{cases}$$

has a solution $d' \equiv d \pmod{N}$. Then since $d' \equiv 1 \pmod{P}$ and $c' \equiv 0 \pmod{P}$, we conclude that $(c', d') = 1$.

Since $(c', d') = 1$, there exist integers $u, v \in \mathbf{Z}$ such that $ud' - vc' = 1$. Observe that

$$\begin{pmatrix} a & b \\ c' & d' \end{pmatrix} \begin{pmatrix} u & v \\ c' & d' \end{pmatrix}^{-1} = \begin{pmatrix} ad' - bc' & * \\ 0 & ud' - vc' \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N}$$

or equivalently, there is some integer n for which

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u & v \\ c' & d' \end{pmatrix} = \begin{pmatrix} u + nc' & v + nd' \\ c' & d' \end{pmatrix} \pmod{N}.$$

Since the matrix on the right has determinant 1, we choose $a' = u + nc'$, $b' = v + nd'$. □

Congruence modular groups are finite index subgroups of $\mathrm{SL}_2(\mathbf{Z})$ defined by congruence relations on their matrix entries:

Definition 85. *A subgroup $\Gamma < \mathrm{SL}_2(\mathbf{Z})$ is called a* **congruence subgroup** *(of level N) if there is a positive integer N such that $\Gamma \supset \Gamma(N)$ (with N minimal).*

Remark 86. *Observe that if $A \equiv I \pmod{N}$ and $M \mid N$, then $A \equiv I \pmod{M}$ so that $\Gamma(N) \subset \Gamma(M)$ whenever $M \mid N$.*

Case study: $\Gamma(2)$.

Proposition 87. *The group $\Gamma(2)$ is generated by the matrices $\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.*

Proof. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(2)$. Observe that if the lower left entry is $c = 0$, then $a = d = 1$ or $= -1$ and $2 \mid b$ so that A is a product of powers of $-I$ and $\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$, and we are done.

Suppose that $c \neq 0$. We will suppose that $|a| > |c|$ (the procedure is similar otherwise) and use Euclidean division: There exist $q_0, r_0 \in \mathbf{Z}$ such that $a = 2cq_0 + r_0$ with $|r_0| < |c|$. Then

$$\begin{pmatrix} 1 & -2q_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & * \\ c & * \end{pmatrix} = \begin{pmatrix} r_0 & * \\ c & * \end{pmatrix}.$$

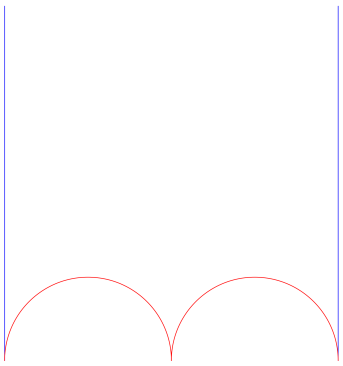
Now $|r_0| < |c|$ and we write $c = 2r_0q_1 + r_1$ with $|r_1| < |r_0|$. Then

$$\begin{pmatrix} 1 & 0 \\ -2q_1 & 1 \end{pmatrix} \begin{pmatrix} r_0 & * \\ c & * \end{pmatrix} = \begin{pmatrix} r_0 & * \\ r_1 & * \end{pmatrix}.$$

Since we are multiplying matrices in the group $\Gamma(2)$, the top entry is always odd and the bottom entry is always even. Applying these two steps in alternation, we eventually obtain a matrix with lower left entry 0 and we are done. □

Remark 88. *Let* $\alpha_2 = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}$. *We leave it to the reader to check that* $\alpha_2^{-1}\Gamma(2)\alpha_2 = \Gamma_0(4)$. *The claim from the previous section follows:* $\Gamma_0(4)$ *is generated by* $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$, *and* $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

The matrix group $\Gamma(2)$ induces an action of $\overline{\Gamma}(2) := \Gamma(2)/\{\pm I\}$ on $\mathbf{H}$. Using the set of generators given by Proposition 87, we can check that the following domain $\mathcal{F}$, with vertices at $-1, 0, 1, \infty$ is a fundamental domain for $\overline{\Gamma}(2)$:



The blue sides are identified by the transformation $z \mapsto z + 2$ while the red sides are identified via the transformation $z \mapsto \frac{z}{2z+1}$. Folding the fundamental domain into a topological surface, we obtain a thrice-punctured sphere, where each puncture is a cusp, corresponding in the fundamental domain picture to the points at infinity 0, 1, and ∞. Observe that since $\Gamma_0(4) = \alpha_2^{-1}\Gamma(2)\alpha_2$, $\alpha_2^{-1}\mathcal{F}$ is a fundamental domain for $\Gamma_0(4)$, with the same side-pairings, and cusps coming from the points $0, \frac{1}{2}, \infty$.

From here on, let Γ be a finite-index subgroup of $\mathrm{SL}_2(\mathbf{Z})$. From the case study of $\Gamma(2)$, we have informally seen that a cusp is both a vertex of the (selected) fundamental domain and a point on the boundary $\mathbf{R} \cup \{\infty\}$ of the upper half-plane $\mathbf{H}$.

Let $\overline{\Gamma}$ denote the image of Γ under the canonical projection $\mathrm{SL}_2(\mathbf{R}) \to \mathrm{PSL}_2(\mathbf{R})$. To give a precise definition, we first need to extend the action of $\overline{\Gamma}$ on $\mathbf{H}$ by Möbius

transformation to an action on $\mathbf{H} \cup \mathbf{R} \cup \{\infty\}$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.z = \begin{cases} \dfrac{a}{c} & z = \infty, \\ \infty & z = -\dfrac{d}{c}, \\ \dfrac{az+b}{cz+d} & \text{otherwise.} \end{cases}$$

Definition 89. *A* **cusp for** Γ *is an orbit* $\overline{\Gamma}.x \subset \overline{\Gamma}\backslash(\mathbf{Q} \cup \{\infty\})$ *with a representative* $x \in \mathbf{Q} \cup \{\infty\}$ *that is the unique fixed point for some* $\gamma \in \overline{\Gamma}$.

If

$$\frac{ax+b}{cx+d} = x \iff cx^2 + (d-a)x - b = 0, \tag{5.3}$$

then a unique solution arises iff the discriminant of this quadratic equation is 0, moreover this unique solution is necessarily in $\mathbf{Q} \cup \{\infty\}$. We will usually refer to a cusp $\overline{\Gamma}.x$ by its representative x, understanding that two cusps x, x' are equivalent if they are in the same $\overline{\Gamma}$-orbit.

In $\overline{\Gamma}(2)$, observe that ∞ is the unique fixed point of $z \mapsto z+2$ in $\overline{\Gamma}(2)$, while 0 is the unique fixed point of $z \mapsto \frac{z}{2z+1}$.

Proposition 90. *Let* x *be a cusp for* Γ *and let* $\overline{\Gamma}$ *denote the image of* Γ *in* $\mathrm{PSL}_2(\mathbf{R})$. *Then* $\Gamma_x = \{\gamma \in \overline{\Gamma} : \gamma.x = x\} \cong \mathbf{Z}$.

Proof. Suppose first that $x = \infty$. Then elements of Γ_∞ have the form $z \mapsto z+t$ for some $t \in \mathbf{Z}$. Since the set of possible values of t is discrete, we can choose a minimal such $t > 0$, which we call t_0. We claim that

$$\langle z \mapsto z + t_0 \rangle = \{z \mapsto z + mt_0 : m \in \mathbf{Z}\} = \Gamma_\infty.$$

Suppose for contradiction that there is some translation $z \mapsto z+s$ in Γ_∞ with s not an integer multiple of t_0. There is some integer m such that the composition $z \mapsto z + (s + mt_0)$ is a translation by $s + mt_0 \in (0, t_0)$ in Γ_∞, contradicting the minimality of t_0.

Suppose that $x \neq \infty$. Since $x \in \mathbf{Q}$, we can always choose $\sigma_x \in \mathrm{PSL}_2(\mathbf{Z})$ such that $\sigma_x(\infty) = x$. Then the subgroup $\sigma_x^{-1}\Gamma\sigma_x$ of $\mathrm{PSL}_2(\mathbf{Z})$ has a cusp at infinity and in fact

$$\sigma_x^{-1}\Gamma_x\sigma_x = (\sigma_x^{-1}\Gamma\sigma_x)_\infty.$$

Since the RHS is isomorphic to $\mathbf{Z}$ by the first part of this proof, we conclude. □

It is usually convenient to consider the situation at the cusp at infinity. We note that when Γ is congruence, then it always has a cusp at infinity (and it will usually have many more!). Indeed, $\Gamma \supset \Gamma(N)$ and $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma(N)$ fixes the point at infinity.

Coming back to Definition 83, we say that a modular form $f \in M_k(N, \chi)$ is **holomorphic at each cusp** x **of** $\Gamma_0(N)$ if

$$\lim_{\substack{z \in \mathcal{F} \\ z \to x}} f(z) = \text{ constant},$$

where $\mathcal{F}$ is a fundamental domain for $\Gamma_0(N)$.

5.3. ...and their modular forms

In this section, let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$. We say that f is a **modular form for Γ of weight** k if $f : \mathbf{H} \to \mathbf{C}$ if f is holomorphic on $\mathbf{H}$, holomorphic at each cusp for Γ and $f(\frac{az+b}{cz+d}) = (cz+d)^k f(z)$. The space of all such forms is denoted, as usual, by $M_k(\Gamma)$.

Proposition 91. *If Γ is of level N, $M_k(\Gamma) \subset M_k(\Gamma(N))$.*

Proof. Let $f \in M_k(\Gamma)$. Then f is a holomorphic function and has the modular transformation of a form for $\Gamma(N)$, since $\Gamma(N) \subset \Gamma$. Observe that cusps of $\Gamma(N)$ are cusps of Γ. Hence f is holomorphic at each cusp of $\Gamma(N)$ and we conclude that $f \in M_k(\Gamma(N))$. □

Proposition 92. *If $f \in M_k(\Gamma(N))$, then $g(z) = f(Nz) \in M_k(\Gamma_1(N^2))$.*

Proof. From the previous proposition, observe that if we had $\Gamma_1(N^2) \subset \Gamma(N)$ (which is clearly false) we would have $M_k(\Gamma(N)) \subset M_k(\Gamma_1(N^2))$. We claim instead that $\alpha_N \Gamma_1(N^2) \alpha_N^{-1} \subset \Gamma(N)$, where

$$\alpha_N = \begin{pmatrix} \sqrt{N} & \\ & 1/\sqrt{N} \end{pmatrix},$$

and deduce that $g = f \circ \alpha_N \in M_k(\Gamma_1(N^2))$.

A direct matrix computation shows

$$\alpha_N \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} a & bN \\ c/N & d \end{pmatrix}$$

and from this we see that $\alpha_N \Gamma_1(N^2) \alpha_N^{-1} \subset \Gamma(N)$.

In particular, if x is a cusp for $\Gamma_1(N^2)$, then $\alpha_N.x = Nx$ is a cusp for $\Gamma(N)$ and hence the function g is automatically holomorphic at each cusp of $\Gamma_1(N^2)$. Further, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N^2)$, $\gamma' = \left(\begin{smallmatrix} a & bN \\ c/N & d \end{smallmatrix}\right) \in \Gamma(N)$. Then

$$g(\gamma.z) = f(\alpha_N \gamma.z) = f(\gamma' \alpha_N z) = (\tfrac{c}{N} Nz + d)^k f(Nz) = (cz+d)^k g(z)$$

and this proves that $g \in M_k(\Gamma_1(N^2))$. □

Theorem 29. *For each $N \in \mathbf{N}$, we have*

$$M_k(\Gamma_1(N)) = \bigoplus_\chi M_k(N, \chi),$$

where χ runs over all Dirichlet characters mod N.

For this we will need the intermediate lemma

Lemma 93. *$\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ with $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbf{Z}/N\mathbf{Z})^\times$.*

Proof. Consider the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N) \mapsto d \in (\mathbf{Z}/N\mathbf{Z})^\times$, which is well-defined since $(d, N) = 1$. This is a homomorphism, since

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} * & * \\ * & cb' + dd' \end{pmatrix}$$

and $cb' + dd' \equiv dd' \pmod N$. It is surjective, as follows from the proof of Theorem 28. Its kernel contains all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ with $d \equiv 1 \pmod N$ and their inverses, forcing $a \equiv 1 \pmod N$ as well. Conversely, each $\gamma \in \Gamma_1(N)$ lies in the kernel of the above map. □

Proof of Theorem 29. There is an action of $\Gamma_0(N)$ on $M_k(\Gamma_1(N))$ given by

$$(f|_k \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right))(z) := (cz+d)^{-k} f(\tfrac{az+b}{cz+d}).$$

(Checking that this is actually an action is left as exercise — note that the action is well-defined requires $\Gamma_1(N)$ to be normal in $\Gamma_0(N)$.) Observe that the action restricted to $\Gamma_1(N)$ is trivial. Indeed, for each $\gamma \in \Gamma_1(N)$, $f|_k\gamma = f$ since $f \in M_k(\Gamma_1(N))$. Hence we have an induced action of $\Gamma_0(N)/\Gamma_1(N) \cong G := (\mathbf{Z}/N\mathbf{Z})^\times$ on $M_k(\Gamma_1(N))$, which depends only on the value of $\left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix}\right)$. Accordingly we may write $f|_k d = f|_k\gamma$ for $\gamma = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix}\right) \in \Gamma_0(N)/\Gamma_1(N)$.

For each $\chi \in \widehat{G}$, set

$$\pi(\chi)f = \frac{1}{\varphi(N)} \sum_{d\,(N)} \overline{\chi}(d) f|_k d \in M_k(\Gamma_1(N)).$$

We have $\pi(\chi)f \in M_k(\Gamma_1(N))$ and in fact

$$\pi(\chi)f \in M_k(N,\chi) = \{f \in M_k(\Gamma_1(N)) : f|_k\gamma = \chi(d)f \text{ for all } \gamma \in \Gamma_0(N)\}$$

as follows from

$$(\pi(\chi)f)|_k h = \frac{\chi(h)}{\varphi(N)} \sum_{d\,(N)} \overline{\chi}(dh) f|_k(dh) = \chi(h)\pi(\chi)f$$

for each $h \in G$. Moreover, observe that $\pi(\chi)f = f$ if $f \in M_k(N,\chi)$. Hence $\pi(\chi)$ is the projection of $M_k(\Gamma_1(N))$ onto $M_k(N,\chi)$.

We leave as exercise to check that $M_k(N,\chi_1) \cap M_k(N,\chi_2) = \{0\}$ if $\chi_1 \neq \chi_2$. Finally, we show that $\sum \pi(\chi)$ is the identity on $M_k(\Gamma_1(N))$. Indeed, by the orthogonality relations, we have

$$\sum_{\chi\,(N)} \pi(\chi)f = \frac{1}{\varphi(N)} \sum_{d\,(N)} \sum_{\chi\,(N)} \overline{\chi}(d) f|_k d = f$$

for all $f \in M_k(\Gamma_1(N))$. This proves the decomposition $f = \sum f_\chi$, with $f_\chi \in M_k(N,\chi)$. □

Remark 94. *In the language of representation theory, we have a representation of $(\mathbf{Z}/N\mathbf{Z})^\times$ on the vector space $M_k(\Gamma_1(N))$. Each representation decomposes into a direct sum of irreducible representations, and each irreducible representation is induced by a Dirichlet character as above.*

Although we will not prove this, it is important to remark that the spaces $M_k(N,\chi)$ are again finite-dimensional and that their dimensions can be computed.

5.4. Twisted Eisenstein series and sums of squares

We can construct examples of forms in $M_k(N,\chi)$ by 'averaging.' Recall the Eisenstein series $E_k(z)$ defined by the scaling

$$G_k(z) = 2\zeta(k)E_k(z), \quad E_k(z) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \frac{1}{(cz+d)^k}$$

for $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, $\Gamma_\infty = \langle z\mapsto z+1\rangle$. We define the Eisenstein series (at the cusp at ∞) for (N,χ) to be

$$E_{k,\chi}(z) := E_{k,\chi}^{(\infty)}(z) = \sum_{\gamma\in\Gamma_\infty\backslash\overline{\Gamma}_0(N)} \frac{\overline{\chi}(d)}{(cz+d)^k}$$

where $\Gamma_\infty = \langle z\mapsto z+N\rangle$.

We can make this series more explicit by using that there is a bijection

$$\Gamma_\infty\backslash\overline{\Gamma}_0(N) \to \{(c,d) : N \mid c,\ (c,d)=1\}/\{\pm 1\}$$

Then

$$E_{k,\chi}^{(\infty)}(z) = \frac{1}{2}\sum_{\substack{(c,d)=1\\ N|c}} \frac{\overline{\chi}(d)}{(cz+d)^k}.$$

Note that if k and χ do not have the same parity then $E_{k,\chi}=0$. We will assume both k and χ are either even or odd.

Thus

$$G_{k,\chi} := G_{k,\chi}^{(\infty)} = 2L(k,\overline{\chi})E_{k,\chi}^{(\infty)}, \quad G_{k,\chi}^{(\infty)}(z) = \sum_{(c,d)\neq(0,0)} \frac{\overline{\chi}(d)}{(cNz+d)^k}.$$

Proposition 95. *For k and χ of the same parity, and $k\geq 1$, we have $G_{k,\chi}^{(\infty)}\in M_k(N,\chi)$ with for each cusp p,*

$$\lim_{z\to p} G_{k,\chi}^{(\infty)}(z) = \begin{cases} 2L(k,\overline{\chi}) & p=\infty,\\ 0 & \textit{otherwise.}\end{cases}$$

Proof. We leave it as an exercise to check that $G_{k,\chi}$ has the correct modular transformation and that — applying the twisted Poisson summation and proceeding as in Exercise [BLAH] — its Fourier expansion is given by

$$G_{k,\chi}^{(\infty)}(z) = 2L(k,\overline{\chi}) + \frac{2(2\pi i)^k N^{1-k}}{\tau(\chi)\Gamma(k)}\sum_{n\geq 1}\left(\sum_{d|n}\chi(d)d^{k-1}\right)q^n.$$

To conclude we prove that $G_{k,\chi}(z)\to 0$ as $z\to p$ for a cusp that does not have the point at infinity as representative.

Choose some σ_p such that $\sigma_p(\infty)=p$. Then

$$E_{k,\chi}^{(p)}(\sigma_p z) = \sum_{\gamma\in\Gamma_p\backslash\Gamma_0(N)} \frac{1}{j(\gamma,\sigma_p z)^k} = j(\sigma_p,z)^k \sum_{\gamma\in\Gamma_p\backslash\Gamma_0(N)} \frac{1}{j(\gamma\sigma_p,z)^k}$$

As $y \to \infty$, the only contribution in the sum comes from those γ's for which $\gamma\sigma_p = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \in \sigma_p^{-1}\Gamma_p\sigma_p$. We claim that if p and ∞ are not representatives of the same cusp, this contribution is zero. Assume otherwise; there is an element $\gamma \in \Gamma_0(N)$ such that $\gamma \in \sigma_p^{-1}\Gamma_p$ and $\gamma(p) = \infty$, i.e., p and ∞ lie in the same orbit, a contradiction. □

Consider $M_1(4, \chi)$ where χ is the odd Dirichlet character mod 4. This space is known to have dimension 1 and contains both

$$G_{1,\chi}(z) = \frac{\pi}{2} + 2\pi \sum_{n \geq 1} \left(\sum_{d|n} \chi(d) \right) q^n$$

and

$$\theta^2(z) = 1 + \sum_{n \geq 1} r_2(n) q^n.$$

Comparing Fourier coefficients, we have $\theta^2 = \frac{2}{\pi} G_{1,\chi}$ and hence

$$r_2(n) = 4 \sum_{d|n} \chi(d) = 4 \sum_{\substack{d|n \\ d \text{ odd}}} (-1)^{\frac{d-1}{2}}$$

for each $n \geq 1$. As a corollary, we obtain Fermat's theorem on primes of the form $4k + 1$:

THEOREM 30 (Fermat). *Every prime of the form $p \equiv 1$ (mod 4) can be written (uniquely up to signs and reordering) as a sum of two squares.*

PROOF. It follows directly from the previous formula that $r_2(p) = 4(\chi(1) + \chi(p)) = 4 \cdot 2 = 8$. □

Remark 96. *Along the same lines, one can prove that*

$$r_4(n) = 8(2 + (-1)^n) \sum_{\substack{d|n \\ d \text{ odd}}} d$$

and deduce Lagrange's theorem: every $n \geq 1$ can be written as a sum of 4 squares. See [8, Chapter 3.1].

5.5. Theta functions and quadratic forms

The theorems of Fermat and Lagrange (and much of the work developed by Gauss in his *Disquisitiones arithmeticae*) are early instances of the more general representation problem for quadratic forms: given an integral positive definite quadratic form Q in $d \geq 2$ variables, which $n \in \mathbf{N}$ are represented by Q? Or, more explicitly,

for which $n \in \mathbf{N}$ does $Q(\xi) = n$ have a solution $\xi \in \mathbf{Z}^d$?

***

Review of some basic definitions concerning quadratic forms. A **(real) quadratic form** $Q \in \mathbf{R}[x_1, \ldots, x_d]$ in $d \geq 2$ variables is a polynomial with real coefficients where each term has degree 2. Recall from linear algebra that a matrix $A \in M_d(\mathbf{R})$ is symmetric if $A^T = A$, and that in that case it is diagonalizable. More precisely, there exists an orthogonal matrix $U \in O(d) = \{U \in M_d(\mathbf{R}) : U^T U = I\}$ such that $A = U^T D U$, where D is diagonal.

Proposition 97. *There is a bijection between real quadratic forms and the symmetric matrices $A \in M_d(\mathbf{R})$ that determine them.*

PROOF. The coefficients of Q can be arranged in a $d \times d$ matrix A such that $Q(x) = x^T A x$. Since $Q(x) = x^T(\frac{A+A^T}{2})x$, we may assume that A is symmetric. □

We will say that a quadratic form is **integral**[1] if $A \in M_d(\mathbf{Z})$. Using that $2Q(x) = x^T(A + A^T)x$, we will assume not only that A is symmetric but also that it is even, i.e., for each diagonal entry $a_{ii} \equiv 0 \pmod 2$, and use the convention

$$Q(x) = \frac{1}{2} x^T A x$$

for integral quadratic forms, with A integral, symmetric, and even. The form Q is **positive definite** if all its eigenvalues (which are the diagonal elements in the diagonal matrix $D = UAU^T$) are positive. We write $|A| = \det(A)$ and note that $|A| > 0$ when Q is positive definite.

***

One reason for the prominence of theta series in the theory of modular forms is that they help answer the question above. Consider

$$\Theta_Q(z) := \sum_{\xi \in \mathbf{Z}^d} q^{Q(\xi)}.$$

If we take Q to be integral and positive definite, then Θ_Q converges for all $z \in \mathbf{H}$, and

$$\Theta_Q(z) = \sum_{n \geq 0} r_Q(n) q^n, \quad r_Q(n) = \#\{\xi \in \mathbf{Z}^d : Q(\xi) = n\}.$$

In fact, the modular theory of these theta functions is used to show

THEOREM 31 (Kloosterman, 1924–1926). *If $d \geq 4$ and $n \in \mathbf{N}$ is a sufficiently large integer such that the congruence equation $Q(\xi) \equiv n \pmod{2^7|A|}$ has a solution $\xi \in \mathbf{Z}^d$ then n is represented by Q.*

We will not include a proof of this result, but refer the interested reader to [4, Chapter 11] and [7]. In this section, we will restrict ourselves to illustrating some new aspects relative to the modular behavior of these theta functions. As usual, to show that Θ_Q is a modular form, one builds on Poisson summation to obtain an inversion formula.

[1] This is Gauss' definition, but beware that it clashes with the other common definition of an integral quadratic form as a quadratic form with integer coefficients.

Proposition 98. *Let Q be an even integral positive definite form. Then*

$$\Theta_Q(z) = \sum_{\xi\in\mathbf{Z}^d} e(Q_A(\xi)z) = |A|^{-1/2}\left(\frac{i}{z}\right)^{d/2}\sum_{\xi\in\mathbf{Z}^d} e(-Q_{A^{-1}}(\xi)/z).$$

Proof. The idea is to apply Poisson summation (in $\mathbf{R}^d$) to the function $f(x) = q^{Q_A(x)}$. Its Fourier transform is given by

$$\widehat{f}(\xi) = \int_{\mathbf{R}^d} f(x)e(-x^T\xi)dx = \int_{\mathbf{R}^d} e(Q_A(x)z - x^T\xi)\,dx$$

for $\xi\in\mathbf{Z}^d$. By a change of basis, we can diagonalize the form Q_A: let $A = U^TDU$ as above and set $B = \sqrt{D}U$ so that $A = B^TB$. Then for $y = Bx$, $Q_A(x) = \frac{1}{2}\|y\|^2$. Set also $\eta := B^{-T}\xi$. Then

$$\begin{aligned}\int_{\mathbf{R}^d} e(Q_A(x)z - x^T\xi)dx &= \frac{1}{|B|}\int_{\mathbf{R}^d} e(\|y\|^2z/2 - y^T\eta)dy\\ &= \frac{1}{|A|^{1/2}}\prod_{k=1}^d\int_{\mathbf{R}} e(y_k^2z/2 - y_k\eta_k)dy_k\\ &= |A|^{-1/2}\prod_{k=1}^d\left(\frac{i}{z}\right)^{1/2} e\left(-\eta_k^2/2z\right)\\ &= |A|^{-1/2}\left(\frac{i}{z}\right)^{d/2} e\left(-\|\eta\|^2/2z\right).\end{aligned}$$

Finally, note that $\|\eta\|^2 = \xi^TB^{-1}B^{-T}\xi = \xi^T(B^TB)^{-1}\xi = \xi^TA^{-1}\xi = Q_{A^{-1}}(\xi)$ □

Definition 99. *The* **level** *of an even integral positive definite Q is the smallest integer $N\in\mathbf{N}$ such that NA^{-1} is again even and integral.*

Under the change of variable $\eta = NA^{-1}\xi$, i.e.,

$$Q_{A^{-1}}(\xi) = N^{-2}Q_A(\eta),$$

the previous formula becomes

$$\Theta_Q(z) = |A|^{-1/2}\left(\frac{i}{z}\right)^{d/2}\sum_{\substack{\eta\in\mathbf{Z}^d\\ A\eta\equiv 0\,(N)}} e\left(\frac{Q_A(\eta)}{N^2z}\right) = |A|^{-1/2}\left(\frac{i}{z}\right)^{d/2}\sum_{\substack{\eta\,(N)\\ A\eta\equiv 0\,(N)}}\Theta_Q(-1/z;\eta),$$

where $\Theta_Q(z;\eta)$ is the **congruent theta function**

$$\Theta_Q(z;\eta) = \sum_{\substack{\xi\in\mathbf{Z}^d\\ \xi\equiv\eta\,(N)}} q^{Q(\xi)/N^2}.$$

Thus establishing the modularity of Θ_Q requires understanding the transformation of each congruent theta function $\Theta_Q(z;\eta)$. We will not pursue this here, but we take note of the final result:

THEOREM 32. [4, Theorem 10.9] *Let d be even, and let Q be an even integral positive definite quadratic form in d variables. Then $\Theta_Q \in M_{d/2}(N, \chi_D)$, where $D = (-1)^{d/2}|A|$ and $\chi_D(d) = \left(\frac{D}{d}\right)$ is the quadratic Dirichlet character given by the Kronecker symbol.*

A special case of this theorem is θ^d, where θ is the classical theta series and d is even. Indeed, let $A = 2I_d$. Clearly A is symmetric even integral and positive definite, with level $N = 4$, and $Q_A(\xi) = \frac{1}{2}\|\xi\|^2$. Thus $\Theta_Q = \theta^d \in M_{d/2}(4, \chi)$.

***

The simple case of unimodular forms. Let Q be even integral positive definite. We say that Q is unimodular if its associated symmetric matrix A has determinant $|A| = 1$. Unimodular forms are precisely level 1 forms, as follows from the following proposition.

Proposition 100. *Let Q be an even integral positive definite quadratic form in d variables with symmetric matrix A and let N be the level of Q. Then $N \mid |A|$ and $|A| \mid N^d$.*

PROOF. One can check by linear algebra that $|A|A^{-1}$ is again even integral positive definite. Hence, by definition $N \mid |A|$. On the other hand, if N is the level of Q, then $\det(NA^{-1}) = N^d|A|^{-1}$ is an integer and $|A| \mid N^d$. □

THEOREM 33. *Let Q be an even integral positive definite and unimodular quadratic form in d variables. If $8 \mid d$, then $\Theta_Q \in M_{d/2}(\Gamma)$ for $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$.*

PROOF. If Q is unimodular, then $Q_{A^{-1}} = Q_A$ and the inversion formula in Proposition 98 reduces to

$$\Theta_Q(z) = \left(\frac{i}{z}\right)^{d/2} \Theta_Q\left(\frac{-1}{z}\right).$$

Hence if $8 \mid d$, $\Theta_Q(-1/z) = z^{d/2}\Theta_Q(z)$. Moreover, by definition, $\Theta_Q(z+1) = \Theta_Q(z)$. Since $z \mapsto -1/z$ and $z \mapsto z+1$ generate $\mathrm{PSL}_2(\mathbf{Z})$ (Theorem 24), we conclude that

$$\Theta_Q \in M_{d/2}(\mathrm{PSL}_2(\mathbf{Z}))$$

if $8 \mid d$. □

5.6. Equidistribution of lattice points on the sphere

We will apply the modular theory of the theta functions to discuss the following statement: *When $d \geq 3$, the sets*

$$\Omega_n = \left\{ \frac{\xi}{\|\xi\|} \in \mathbf{S}^{d-1} : \xi \in \mathbf{Z}^d,\ \|\xi\|^2 = n \right\}$$

become equidistributed on $\mathbf{S}^{d-1}$ as $n \to \infty$ along an appropriate sequence of integers.

***

Although the case $d = 2$ is not included, we will first discuss the statement in that situation to introduce the notion of equidistribution. For convenience, using the diffeomorphism $\theta \in \mathbf{R}/\mathbf{Z} \mapsto e^{2\pi i\theta} \in \mathbf{S}^1$, we reformulate this problem in $\mathbf{R}/\mathbf{Z}$.

Definition 101 (Proposition-definition)**.** *We say that a sequence* $(a_n)_{n\geq 1} \subset \mathbf{R}/\mathbf{Z}$ **becomes equidistributed in** $\mathbf{R}/\mathbf{Z}$ *if either of the following equivalent properties hold*

(1)
$$\lim_{N\to\infty} \frac{1}{N} \#\{n \leq N : a_n \in [a,b)\} = b - a$$
for each subinterval $[a,b) \subset \mathbf{R}/\mathbf{Z}$*;*

(2)
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(a_n) = \int_0^1 f(x)dx$$
for each $f \in C(\mathbf{R}/\mathbf{Z})$.

Sketch of proof. Observe that (1) coincides with the formulation of (2) if we choose f to be the characteristic function $\chi_{[a,b)}$ supported on $[a,b)$. We can pass from (1) to (2) by using that every continuous function is a finite linear combination of characteristic functions supported on intervals and from (2) to (1) by approximating the characteristic function by continuous functions. □

Remark 102. *For a real sequence* $(a_n)_{n\in\mathbf{N}} \subset \mathbf{R}$*, we say that* (a_n) *becomes uniformly distributed mod 1 if the associated sequence* $b_n := a_n \pmod 1$ *becomes equidistributed in* $\mathbf{R}/\mathbf{Z}$.

In fact, these approximation arguments allow to further restrict (2) to the dense subset of smooth functions $C^\infty(\mathbf{R}/\mathbf{Z})$. Recall that each $f \in C^\infty(\mathbf{R}/\mathbf{Z})$ admits a Fourier series expansion
$$f(x) = \sum_{m\in\mathbf{Z}} \widehat{f}(m) e(mx),$$
with
$$\widehat{f}(m) = \int_0^1 f(x)e(-mx)dx. \tag{5.4}$$
Observe that
$$\widehat{f}(0) = \int_0^1 f(x)dx$$
and that when $m \neq 0$, integration by parts shows that
$$|\widehat{f}(m)| \ll |m|^{-k}$$
for any $k \geq 1$. Thus, by uniform convergence,
$$\frac{1}{N}\sum_{n=1}^{N} f(a_n) = \int_0^1 f(x)dx + \sum_{m\neq 0} \widehat{f}(m) \left(\frac{1}{N}\sum_{n=1}^{N} e(ma_n)\right)$$
for any $f \in C^\infty(\mathbf{R}/\mathbf{Z})$ and thus it follows that (a_n) becomes equidistributed in $\mathbf{R}/\mathbf{Z}$ if and only if
$$\sum_{m\in\mathbf{Z}} |\widehat{f}(m)| \left|\frac{1}{N}\sum_{n=1}^{N} e(ma_n)\right| \to 0$$
as $N \to \infty$. Considering the rapid decay of the Fourier coefficients (5.4), we arrive to

THEOREM 34 (Weyl's criterion). *A sequence $(a_n)_{n\in\mathbf{N}} \subset \mathbf{R}/\mathbf{Z}$ becomes equidistributed in $\mathbf{R}/\mathbf{Z}$ if and only if*

$$\sum_{n=1}^{N} e(ma_n) = o(N)$$

for each $m \in \mathbf{Z} \setminus \{0\}$.

Example 103. *Consider the sequence $a_n := n\alpha$ (mod 1) for a fixed $\alpha \in \mathbf{R}/\mathbf{Z}$. Clearly if $\alpha = p/q$ is rational, then the sequence a_n takes only finitely many values with $a_{q+k} = a_k$. (If we think of these points on the unit circle, they correspond to the q-th roots of unity.) Suppose instead that α is irrational. In that case, Weyl's criterion can be applied to show that (a_n) becomes equidistributed:*

$$\left|\frac{1}{N}\sum_{n=1}^{N} e(mn\alpha)\right| = \frac{|e(mN\alpha)-1|}{N\,|e(m\alpha)-1|} \leq \frac{2}{N|e(m\alpha)-1|} \to 0$$

as $N \to \infty$ whenever $m \neq 0$.

Example 104. *Weyl famously put his criterion to use to show the following theorem*

THEOREM 35 (Weyl, 1916). *Let $f(x) \in \mathbf{R}[x]$ be a polynomial with at least one irrational coefficient. Then $(f(n))_{n\in\mathbf{N}}$ becomes uniformly distributed mod 1.*

The special case corresponding $a_n = n^2\alpha$ will be studied in the exercises.

***

From here on, let $d \geq 3$. In analogy to 1-dimensional situation, we say that the sets $\Omega_n \subset \mathbf{S}^{d-1}$ become equidistributed in $\mathbf{S}^{d-1}$ as $n \to \infty$ if

$$\lim_{n\to\infty} \frac{1}{|\Omega_n|} \sum_{x\in\Omega_n} f(x) = \int_{\mathbf{S}^{d-1}} f(x)d\lambda(x),$$

for every $f \in C^\infty(\mathbf{S}^{d-1})$, where λ is the normalized Lebesgue measure on the sphere. Harmonic analysis on the sphere yields for each such test-function f an expansion

$$f(x) = \sum_{\nu\geq 0} c_\nu P_\nu(x),$$

with c_ν rapidly decaying and where the P_ν are spherical harmonics, i.e., homogeneous polynomials on $\mathbf{R}^d$ of degree ν (i.e., $P_\nu(ax) = a^\nu P(x)$) that are harmonic on $\mathbf{R}^d$. Seen as restricted to $\mathbf{S}^{d-1}$, the space of all spherical harmonics is dense in the space of smooth functions on $\mathbf{S}^{d-1}$ (with respect to the uniform norm). Hence by Weyl's criterion, equidistribution takes place if and only if

$$\lim_{n\to\infty} \frac{1}{|\Omega_n|} \sum_{x\in\Omega_n} P_\nu(x) = 0 \tag{5.5}$$

for each $\nu > 0$. We specialize the criterion to the sets Ω_n introduced at the beginning of this section. Then $|\Omega_n| = r_d(n)$ and

$$\sum_{x\in\Omega_n} P_\nu(x) = \sum_{\substack{\xi\in\mathbf{Z}^d\\ \|\xi\|^2=n}} P_\nu\left(\frac{\xi}{\|\xi\|}\right) = n^{-\nu/2} \sum_{\substack{\xi\in\mathbf{Z}^d\\ \|\xi\|^2=n}} P_\nu(\xi).$$

The generating function for the latter sum:

$$\sum_{n\geq 0} r_d(n, P_\nu) q^n := \sum_{n\geq 0} \left(\sum_{\substack{\xi\in\mathbf{Z}^d\\ \|\xi\|^2=n}} P_\nu(\xi) \right) q^n = \sum_{\xi\in\mathbf{Z}^d} P_\nu(\xi) q^{\|\xi\|^2} =: \Theta(z; P_\nu)$$

is a cusp form in $S_{d/2+\nu}(4, \chi)$ (since as a homogeneous function $P_\nu(0) = 0$ when $\nu > 0$). Thus (5.5) is equivalent to showing

$$\lim_{n\to\infty} \frac{n^{-\nu/2} r_d(n, P_\nu)}{r_d(n)} = 0 \tag{5.6}$$

for all $\nu > 0$.

We want lower and upper bounds on Fourier coefficients of modular forms. For cusp forms, the first available upper bound is the following.

Proposition 105 (Hecke's bound)**.** *Let $f = \sum a_n q^n$ is a cusp form of weight k, then $a_n \ll n^{k/2}$.*

Proof. We first claim that $F(z) = |f(z)|y^{k/2}$ is Γ-invariant. Indeed,

$$|f(\gamma z)|y(\gamma z)^{k/2} = |f(z)|y^{k/2}|cz + d|^k |cz + d|^{-k} = |f(z)|y^{k/2}.$$

Since f vanishes at each cusp, it is bounded on the fundamental domain, and hence uniformly bounded. Then

$$|a_n| \leq e^{2\pi n y} \int_0^1 |f(x + iy)|dx \ll e^{-2\pi n y} y^{-d/2}$$

for any fixed $y > 0$ and we optimize by choosing $y = 1/n$. □

Corollary 106. *For each $\nu > 0$, $n^{-\nu/2} r_d(n, P_\nu) \ll n^{d/4}$.*

Proof. From the discussion above, we have that $r_d(n, P_\nu)$ are Fourier coefficients of a cusp form of weight $d/2 + \nu$. Thus, applying Hecke's bound, we have $r_d(n, P_\nu) \ll n^{d/4+\nu/2}$. □

Proposition 107. *Let $d \geq 5$ such that $4 \mid d$. Then $r_d(n) \gg n^{d/2-1}$.*

Proof. First note that for all $d \geq 4$, we have $r_d(n) > 0$ by Lagrange's theorem (every positive integer can be written as a sum of four squares). Consider the theta function $\theta^d = \sum r_d(n) q^n \in M_{d/2}(4, \chi)$. Since $4 \mid d$, we have $\chi = \chi_0$ and $M_{d/2}(4, \chi) = M_{d/2}(\Gamma_0(4))$. Each $f \in M_{d/2}(4, \chi)$ can be written uniquely as a linear combination of Eisenstein series and cusp forms in $M_{d/2}(\Gamma_0(4))$. Hence, each $r_d(n)$ is a linear

combination of Fourier coefficients of Eisenstein series and Fourier coefficients of cusp forms of weight $k = d/2$; we write

$$r_d(n) = \delta_d(n) + h_d(n).$$

Then by Hecke's bound $|h_d(n)| = O(n^{d/4})$.

Looking at the Fourier coefficients of the Eisenstein series to estimate $\delta_d(n)$, we see that the main contribution comes from the divisor function $\sigma_{d/2-1}(n)$. Since the divisor function $\sigma_k(n)$ is multiplicative, we first observe that for prime powers $n = p^\ell$,

$$\sigma_k(p^\ell) = \sum_{d|p^\ell} d^k = \sum_{j=0}^{\ell} p^{jk} = \frac{p^{(\ell+1)k} - 1}{p^k - 1} \asymp p^{\ell k}$$

and hence, by multiplicativity, the divisor function $\sigma_k(n) \asymp n^k$ for n sufficiently large. We conclude that $\delta_d(n) \asymp n^{d/2-1}$. □

Remark 108. *A stronger result is in fact true: for all $d \geq 4$, we have $\delta_d(n) \asymp n^{d/2-1}$ and hence $r_d(n) \asymp n^{d/2+1} + O(n^{d/4})$ via Hecke's bound. Note that if $d \leq 4$, then $r_d(n) = O(n^{d/4})$. This is not sufficient to prove equidistribution according to Weyl's criterion (5.6). Luckily, for cusp forms of integer weight, a stronger bound than Hecke's is known, namely that*

$$|a_n| \ll n^{\frac{k-1}{2}+\varepsilon}$$

for any arbitrarily small $\varepsilon > 0$. This bound is usually referred to as **Ramanujan's conjecture** *and follows from the work of Deligne on the Weil conjectures in the 1970s. We will come back to this important result in the next chapter.*

Proof of the equidistribution of lattice points on $\mathbf{S}^{d-1}$ for $d \geq 4$ even. By Weyl's equidistribution criterion, equidistribution is equivalent to

$$\lim_{n\to\infty} \frac{n^{-\nu/2} r_d(n, P_\nu)}{r_d(n)} = 0$$

for any homogeneous polynomial P of degree $\nu > 0$. These functions appear as the Fourier coefficients of certain modular form; we use that

$$\Theta(\cdot, P) = \sum_{n\geq 1} r_d(n, P) q^n \in S_{d/2+\nu}$$

to deduce via Ramanujan's conjecture that $r_d(n, P) \ll n^{\frac{d}{4}-\frac{1}{2}+\frac{\nu}{2}+\varepsilon}$, and that

$$\theta^d = \sum_{n\geq 0} r_d(n) q^n \in M_{d/2}$$

together with Remark 108 to deduce that $r_d(n) \asymp n^{d/2-1}$. Thus indeed

$$\lim_{n\to\infty} \frac{n^{-\nu/2} r_d(n, P_\nu)}{r_d(n)} \ll \lim_{n\to\infty} \frac{n^{\frac{d}{4}-\frac{1}{2}+\varepsilon}}{n^{d/2-1}} = \lim_{n\to\infty} n^{-\frac{d}{4}+\frac{1}{2}+\varepsilon} = 0.$$

□

***

What happens when...

...$d \geq 4$ is odd? The theta functions θ^d and $\Theta(\cdot, P_\nu)$ are modular (resp. cusp) forms of half-integer weight $d/2$ (resp. $d/2+\nu$). Although their modular transformation is more complicated, Hecke's bound and the asymptotic $\delta_d(n) \asymp n^{d/2-1}$ apply as well, and equidistribution again follows.

...$d = 3$? Several complications arise:

(1) First of all, not every positive integer can be expressed as the sum of three squares. We have the following results:

THEOREM 36 (Legendre). *$r_3(n) = 0$ iff $n = 4^a(8^b+7)$.*

THEOREM 37 (Siegel). *If n is squarefree, $r_3(n) \gg n^{1/2+\varepsilon}$ for any $\varepsilon > 0$.*

(2) We are again dealing with modular forms of half-integer weight — θ^3 has weight $3/2$ and $\Theta(\cdot, P_\nu)$ has weight $3/2+\nu$.

(3) As in Remark 108, since $d = 3$, Hecke's bound does not suffice anymore, and Ramanujan's conjecture does not generally hold for forms of half-integer weight. In fact, a counter-example is given by the following theta function. Let χ be an odd Dirichlet character (mod 4), and consider

$$\sum_{n \in \mathbf{Z}} n\chi(n) q^{n^2} = 2 \sum_{n \geq 1} n\chi(n) q^{n^2} \in M_{3/2}(4, \chi)$$

with Fourier coefficients $a_{n^2} = n\chi(n)$. If Ramanujan's conjecture were true, we would have $|a_{n^2}| = n \ll n^{1/4+\varepsilon}$, which is absurd since $|a_{n^2}| = n$. Nonetheless Ramanujan's conjecture is expected to be true for n squarefree, and available (highly nontrivial) bounds in this direction yield:

THEOREM 38 (Duke 1990, via Iwaniec 1987). *The sets Ω_n become equidistributed on $\mathbf{S}^3$ as $n \to \infty$ along squarefree integers.*

5.7. Exercises

Exercise 109. *Let χ be a Dirichlet character mod N.*

(1) *Show that $M_k(N, \chi) = \{0\}$ if χ and k are not of the same parity;*
(2) *Show that $M_k(N, \chi_1) \cap M_k(N, \chi_2) = \{0\}$ if $\chi_1 \neq \chi_2$;*
(3) *Show that there is an isomorphism $M_k(N, \chi) \cong M_k(N, \overline{\chi})$.*

Exercise 110. *Let $j(A, z) = cz + d$, where $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{R})$ and $z \in \mathbf{H}$, and let χ be a Dirichlet character mod N.*

(1) *Show that $j(A, z)$ is a multiplicative 1-cocycle, i.e., $j(AB, z) = j(A, Bz)j(B, z)$ for all A, $B \in \mathrm{SL}_2(\mathbf{R})$.*
(2) *Consider the Eisenstein series at the cusp p given by*

$$E_{k,\chi}^{(p)}(z) = \sum_{\gamma \in \Gamma_p \backslash \Gamma_0(N)} \frac{1}{j(\sigma_p \gamma, z)^k}$$

with $\sigma_p(\infty) = p$. (Take $\sigma_\infty = 1$.) Show that $E_{k,\chi}^{(p)} \in M_k(N, \chi)$.

(3) *Show that* $\gamma \mapsto f|_k\gamma$ *induces an action of* $\Gamma_0(N)$ *on* $M_k(\Gamma_1(N))$.

Exercise 111. *Let* χ *be a primitive character mod* N. *Use that the identity* $\sum_{n\geq 1}\frac{x^n}{n} = -\log(1-x)$ *remains valid if* $|x|=1$ *and* $x\neq 1$ *to deduce that*

$$L(1,\chi) = \begin{cases} \dfrac{-\tau(\chi)}{N}\displaystyle\sum_{n=1}^{N}\overline{\chi}(n)\log|1-e(n/N)| & \chi(-1)=1, \\ \dfrac{i\pi\tau(\chi)}{N^2}\displaystyle\sum_{n=1}^{N} n\overline{\chi}(n) & \chi(-1)=-1. \end{cases}$$

Deduce that for χ *the odd Dirichlet character mod 4, we have*

$$G_{1,\chi}^{(\infty)}(z) = \frac{\pi}{2} + 2\pi\sum_{n\geq 1}\left(\sum_{d|n}\chi(d)\right)q^n.$$

Exercise 112. *Show that a positive integer* $n\geq 1$ *is a sum of two squares if and only if* n *is of the form* $n = d^2\cdot 2^\ell\cdot p_1\cdots p_k$, *where* $d\geq 1$, $\ell\in\{0,1\}$, *and* $p_1,\dots,p_k\equiv 1$ *(mod 4).*

Exercise 113. *In this exercise, we consider the Weyl sums*

$$S_N(\alpha) = \sum_{n=1}^{N} e(\alpha n^2)$$

with α *irrational. By Weyl's equidistribution criterion, it suffices to show* $S_N(\alpha) = o(N)$ *to deduce that the sequence* $(n^2\alpha)_{n\in\mathbf{N}}$ *becomes uniformly distributed mod 1.*

(1) *Show that*

$$|S_N(\alpha)|^2 = N + 2\mathrm{Re}\left(\sum_{n=1}^{N-1} e(\alpha n^2)\sum_{m=1}^{N-n} e(2\alpha nm)\right).$$

(This is the key insight: by squaring, the polynomial in the argument is reduced to one degree less — here we pass from αn^2 *to* αmn *— and this procedure can be further iterated to prove Weyl's theorem on polynomial equidistribution.*[2]

(2) *Show that*

$$|e(x)-1| \asymp \|x\|,$$

where $\|x\|$ *is the distance from* x *to the nearest integer, and deduce that*

$$|S_N(\alpha)|^2 \ll N + \sum_{n=1}^{N-1}\min\left\{N, \frac{1}{\|2\alpha n\|}\right\}.$$

[2] See, e.g., Chapter 3 in Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities* for a full proof or Chapter 4.4.3 in Einsiedler, Ward, *Ergodic theory with a view towards number theory* for an ergodic proof.

(3) *We need to understand how often* $\|2\alpha n\| < 1/N$. *By Dirichlet's approximation theorem, for every irrational* α, *there is a* $q \leq N$ *and* $(a,q) = 1$ *such that*

$$\left|2\alpha - \frac{a}{q}\right| < \frac{1}{qN}.$$

(Or in other words, $\|2\alpha q\| < 1/N$.*) Breaking the interval* $\{1, ..., N-1\}$ *into blocks* $\{k, k_1, \ldots, k+q-1\}$ *of length* q *(and possibly one block of smaller size), show that for each* $n \in \{k, \ldots, k+q-1\}$,

$$\|2\alpha n\| = \left\|\frac{r+b}{q} + O\left(\frac{1}{q}\right)\right\|$$

where $r \equiv an \pmod q$ *and* $b = \|2\alpha kq\|$. *It follows that there are only* $O(1)$ n*'s such that* $\|2\alpha n\| < 1/N$.

(4) *Deduce that*

$$\begin{aligned}|S_N(\alpha)|^2 &\ll N + \left(\frac{N}{q}+1\right)\left(N + \sum_{n=1}^{N-1}\frac{1}{\|2\alpha n\|}\right)\\ &\ll N + \left(\frac{N}{q}+1\right)\left(N + \sum_{s=1}^{q/2}\frac{q}{s}\right)\\ &\ll N + \left(\frac{N}{q}+1\right)(N + q\log q) \ll \frac{N^2}{q} + N\log q.\end{aligned}$$

(5) *Show that* $q = q_N \to \infty$ *as* $N \to \infty$ *to conclude that* $|S_N(\alpha)| = o(N)$.

CHAPTER 6

L-functions attached to modular forms

In the first part of this chapter, take $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$ and let $f \in M_k(\Gamma)$ with q-expansion

$$f(z) = \sum_{n \geq 0} a_n q^n.$$

Proposition 114. *The Dirichlet series*

$$L(s, f) = \sum_{n \geq 1} \frac{a_n}{n^s}$$

converges absolutely when $\mathrm{Re}(s) > k$.

Proof. Since $f \in M_k(\Gamma)$, the Fourier coefficients a_n are linear combinations of the divisor sum $\sigma_{k-1}(n)$ (appearing in the Fourier coefficients of Eisenstein series) and Fourier coefficients of cusp forms; hence, as we have seen last chapter,

$$a_n \asymp n^{k-1} + O(n^{\frac{k-1}{2}+\varepsilon})$$

and it follows that

$$|L(s, f)| \leq \sum_{n \geq 1} \frac{|a_n|}{n^\sigma} \ll \sum_{n \geq 1} n^{k-\sigma-1} < +\infty$$

if and only if $\sigma > k$. □

For example, recall that the classical Eisenstein series $E_k(z)$, $k > 2$, have Fourier expansion

$$E_k(z) = 1 + \frac{(-1)^{k/2}(2\pi)^k}{\Gamma(k)\zeta(k)} \sum_{n \geq 1} \sigma_{k-1}(n) q^n.$$

Hence their attached L-functions are constant multiples of

$$\sum_{n=1}^{\infty} \frac{\sigma_{k-1}(n)}{n^s} = \zeta(s)\zeta(s - k + 1)$$

converging absolutely whenever $\mathrm{Re}(s) > k$.

6.1. Hecke's converse theorem (1936)

THEOREM 39. *The function*

$$\Lambda(s,f) = (2\pi)^{-s}\Gamma(s)L(s,f)$$

admits a meromorphic continuation to all $s \in \mathbf{C}$ that is analytic except for simple poles at $s = 0$ and $s = k$ of residue a_0 and admits the functional equation

$$\Lambda(s,f) = (-1)^{k/2}\Lambda(k-s,f).$$

PROOF. As in Riemann's proof of the analytic continuation of the completed zeta-function, we first observe that when $\mathrm{Re}(s) > k$, $\Lambda(s,f)$ has integral representation

$$\Lambda(s,f) = \int_0^\infty (f(iy) - a_0)\, y^s \frac{dy}{y}.$$

The modular transformation

$$f(-1/iy) = (iy)^k f(iy)$$

yields

$$\begin{aligned}\int_0^1 (f(iy)-a_0)y^s\frac{dy}{y} &= i^{-k}\int_0^1 f(-1/iy)y^{s-k}\frac{dy}{y} - a_0\int_0^1 y^s\frac{dy}{y}\\ &= i^{-k}\int_1^\infty f(iy)y^{k-s}\frac{dy}{y} - \frac{a_0}{s}\\ &= i^{-k}\int_1^\infty (f(iy)-a_0)\, y^{k-s}\frac{dy}{y} - \frac{a_0}{k-s} - \frac{a_0}{s}.\end{aligned}$$

Thus

$$\Lambda(s,f) = \int_1^\infty (f(iy)-a_0)\left(y^s + (-1)^{k/2}y^{k-s}\right)\frac{dy}{y} - \frac{a_0}{k-s} - \frac{a_0}{s}.$$

Since

$$f(iy) - a_0 = \sum_{n\geq 1} a_n e^{-2\pi n y} = O(e^{-2\pi y}),$$

we conclude that the integral converges for all s and the meromorphic continuation follows, as does the functional equation. □

Remark 115. *In particular, if $f \in S_k(\Gamma(1))$, then $\Lambda(s,f) = (2\pi)^{-s}\Gamma(s)L(s,f)$*

(i) *has analytic continuation to the whole $\mathbf{C}$-plane*
(ii) *that is bounded in every vertical strip $\sigma_1 \leq \mathrm{Re}(s) \leq \sigma_2$,*
(iii) *and satisfies $\Lambda(s,f) = i^k\Lambda(k-s,f)$.*

THEOREM 40 (Hecke, 1936). *Let $(a_n)_{n\geq 1} \subset \mathbf{C}$ be a sequence for which $|a_n| \ll n^K$ for some $K > 0$. Let*

$$L(s) = \sum_{n\geq 1}\frac{a_n}{n^s}, \quad \Lambda(s) = (2\pi)^{-s}\Gamma(s)L(s).$$

If $\Lambda(s)$ satisfies (i), (ii), (iii) as above, then

$$f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma(1)).$$

Moreover, if $L(s) = \prod(1 - a(p)p^{-s} - p^{k-1-2s})^{-1}$ *holds, then* f *is (up to normalization) a Hecke form.*

PROOF. The bound $|a_n| \ll n^K$ guarantees the convergence of the q-series f. We only need to show that $f(-1/z) = z^k f(z)$ and in fact it suffices to do this for $z = iy$, $y > 0$: if the holomorphic function $f(i/y) - (iy)^k f(iy) = 0$ for all $y > 0$, then it vanishes identically on $\mathbf{H}$ by the open mapping theorem.

By Mellin inversion and using the functional equation for Λ, we may write

$$f(iy) = \frac{1}{2\pi i}\int_{(\sigma)} \Lambda(s) y^{-s} ds = \frac{i^k}{2\pi i}\int_{(\sigma)} \Lambda(k-s) y^{-s} ds.$$

Our goal is to show that the latter integral is equal to

$$\tfrac{i^k}{2\pi i}\int_{(k-\sigma)} \Lambda(k-s) y^{-s} ds = \tfrac{i^k}{2\pi i} y^{-k} \int_{(\sigma)} \Lambda(s) y^{s} ds = (iy)^{-k} f(i/y).$$

In other words, we need to justify shifting the line of integration from (σ) to $(k-\sigma)$. For this we rely on the Phragmén–Lindelöf principle. In complex analysis, we say that a holomorphic function f has **finite order** if $|f(z)| = O\left(e^{|z|^\rho}\right)$ for some $\rho > 0$ and all $|z| > R$. The order ρ is given by

$$\rho = \lim_{R\to\infty} \sup_{r \geq R} \frac{\log\log\left(\max\{|f(z)| : |z| = r\}\right)}{\log r}.$$

Nonvanishing holomorphic functions of finite order can be expressed as infinite products (the Hadamard factorization); this plays an important role in the proof of the explicit formula we left out in Chapter 2. The **Phragmén–Lindelöf principle** states that if $f(s)$ is a holomorphic function of finite order in $\sigma_1 \leq \mathrm{Re}(s) \leq \sigma_2$, $\mathrm{Im}(s) > c$ for some $c > 0$ but that $|f(s)| \ll \mathrm{Im}(s)^M$ for $\mathrm{Re}(s) = \sigma_1, \sigma_2$, then $|f(s)| = O(t^M)$ uniformly in the vertical strip $\sigma_1 \leq \mathrm{Re}(s) \leq \sigma_2$. □

The precursor of Hecke's converse theorem is the following result of Hamburger (1921): If $F(s) = \sum a_n n^{-s}$, $G(s) = \sum b_n n^{-s}$ are Dirichlet series such that $(s-1)F(s)$, $(s-1)G(s)$ are entire of finite order and

$$\pi^{-s/2}\Gamma(s/2)F(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)G(1-s)$$

then $F(s) = G(s) = a_1\zeta(s)$, where ζ is the Riemann zeta function.

In other words, the Riemann zeta function is completely characterized by its functional equation and regularity properties. It should be pointed out that there are other proofs of the meromorphic continuation of ζ — even going back to Riemann's original paper — that do not build on the functional equation as the one we saw does; Hamburger's theorem is not a priori obvious.

6.2. Ramanujan's memoir (1916)

Recall that $\theta^{24} \in M_{12}(\Gamma) = \langle G_{12}\rangle \oplus \langle \Delta\rangle$ and hence the sum-of-squares function $r_{24}(n)$ can be expressed as a linear combination of an explicit divisor sum and the

more mysterious $\tau(n)$, which are the Fourier coefficients of the modular discriminant

$$\Delta = \sum_{n\geq 1} \tau(n) q^n.$$

Studying these functions, Ramanujan made three conjectures based on his numerical computations:

(1) τ is multiplicative, i.e., $\tau(mn) = \tau(m)\tau(n)$ when $(m,n) = 1$;
(2) $\tau(p^{j+1}) = \tau(p^j)\tau(p) - p^{11}\tau(p^{j-1})$, i.e., τ is not completely multiplicative;
(3) $|\tau(p)| \leq 2p^{11/2}$.

***

The first two conjectures were proven by Mordell a year later, in 1917. The idea is to introduce functions $T(n)$, $n \geq 1$, given by

$$T(n)\Delta = \sum_{m\geq 1} \left(\sum_{d|(m,n)} d^{11}\tau(mnd^{-2}) \right) q^m,$$

based on

Fact 116. *The conjectures (1) and (2) hold iff* $\tau(m)\tau(n) = \sum_{d|(m,n)} d^{11}\tau(mnd^{-2})$.

Hence it suffices to show

$$T(n)\Delta = \tau(n)\Delta.$$

If $T(n)$ can be seen to be a linear operator on the one-dimensional vector space $S_{12}(\Gamma)$ (we will prove this in the next section), then $T(n)\Delta = \lambda(n)\Delta$ for some $\lambda(n) \in \mathbf{C}$. At the level of the Fourier coefficients, this translates to the relations

$$\lambda(n)\tau(m) = \sum_{d|(m,n)} d^{11}\tau(mnd^{-2})$$

for all $m, n \in \mathbf{N}$. Taking $m = 1$, we have

$$\lambda(n)\tau(1) = \tau(n)$$

and we can verify by direct computation[1] that $\tau(1) = 1$. Thus $\lambda(n) = \tau(n)$.

***

The multiplicative relations (1) and (2) imply (in fact, are equivalent to) the **Euler product expansion**

$$L(s,\Delta) = \sum_{n\geq 1} \frac{\tau(n)}{n^s} = \prod_p (1 - \tau(p)p^{-s} + p^{11-2s})^{-1} \tag{6.1}$$

[1] Recall that Δ is defined by a linear combination of powers of Eisenstein series; this can be used to show

$$\tau(n) = \frac{65}{756}\sigma_{11}(n) + \frac{691}{756}\sigma_5(n) - \frac{691}{3}\sum_{0<m<n} \sigma_5(m)\sigma(n-m).$$

in the half-plane of convergence of $L(s,\Delta)$. Indeed, by the fundamental theorem of arithmetic and (1), we have

$$L(s,\Delta) = \prod_p L_p(s,\Delta),$$

where

$$L_p(s,\Delta) := \sum_{j\geq 0} \frac{\tau(p^j)}{p^{js}}.$$

Applying (2) yields

$$\begin{aligned} L_p(s,\Delta) &= 1 + \tau(p)p^{-s} + \sum_{j\geq 1} \frac{\tau(p^{j+1})}{p^{(j+1)s}} \\ &= 1 + \tau(p)p^{-s} + \sum_{j\geq 1} \frac{\tau(p^j)}{p^{js}}\tau(p)p^{-s} - \sum_{j\geq 0} \frac{\tau(p^j)}{p^{js}}p^{11-2s} \\ &= 1 + L_p(s,\Delta)(\tau(p)p^{-s} - p^{11-2s}) \end{aligned}$$

and hence

$$L_p(s,\Delta) = (1 - \tau(p)p^{-s} + p^{11-2s})^{-1}.$$

The local factors $L_p(s,f)$ can be used to reframe (3) (referred to as the **Ramanujan conjecture**) as a 'local Riemann hypothesis.'

Fact 117. *The Ramanujan conjecture* (3) *holds iff for each prime* p*,*

$$L_p(s,\Delta)^{-1} = 1 - \tau(p)p^{-s} + p^{11-2s}$$

does not vanish to the right of $\mathrm{Re}(s) = 11/2$*.*

Proof. Consider the factorization

$$1 - \tau(p)p^{-s} + p^{11-2s} = (1 - \alpha(p)p^{11/2-s})(1 - \beta(p)p^{11/2-s}),$$

where the new parameters $\alpha(p)$, $\beta(p) \in \mathbf{C}$ need to satisfy

$$\alpha(p)\beta(p) = 1, \quad \alpha(p) + \beta(p) = \tau(p)p^{-11/2}.$$

We will show that the following assertions are equivalent.

(i) $L_p(s,\Delta)^{-1}$ is nonvanishing for $\mathrm{Re}(s) > 11/2$ for each prime p;
(ii) $|\alpha(p)| = |\beta(p)| = 1$ for each prime p;
(iii) (3), i.e., $|\tau(p)| \leq 2p^{11/2}$ for each prime p.

The nonvanishing (i) implies that $|\alpha(p)|, |\beta(p)| \neq p^{\sigma-11/2}$ for all $\sigma > 11/2$, which in turns implies that $|\alpha(p)|, |\beta(p)| \leq 1$. Since $\alpha(p)\beta(p) = 1$, this yields (ii) $|\alpha(p)| = |\beta(p)| = 1$. (iii) follows immediately by the triangle inequality:

$$|\tau(p)| = |\alpha(p) + \beta(p)|p^{11/2} \leq (|\alpha(p)| + |\beta(p)|)p^{11/2} = 2p^{11/2}.$$

Finally, $|\tau(p)| \leq 2p^{11/2}$ implies that $1 - \tau(p)p^{11/2-s} + p^{11-2s}$ has zeros only on the line $\mathrm{Re}(s) = 11/2$. □

***

Notice that we can define the maps $T(n)$ more generally: Let $f = \sum a(n)q^n \in M_k(\Gamma)$ and set

$$T(n)f = \sum_{m\geq 0}\left(\sum_{d|(m,n)} d^{k-1}a(mnd^{-2})\right)q^m. \tag{6.2}$$

We can also extend Ramanujan's third conjecture to weight k forms; this is the **Ramanujan–Petersson conjecture** (now: Deligne's theorem):

THEOREM 41 (Deligne). *If $f \in S_k(\Gamma)$, then $|a(p)| \leq 2p^{\frac{k-1}{2}}$ for each prime p.*

6.3. Hecke operators

With some (computational) effort, one can show that

(a) (1) and (2) also hold for $T(n)$ as given by (6.2);
(b) For all $n \geq 1$, we have the closed formula

$$T(n)f(z) = \frac{1}{n}\sum_{ad=n} a^k \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right);$$

(c) $T(n)f \in M_k(\Gamma)$ and if $f \in S_k(\Gamma)$, $T(n)f \in S_k(\Gamma)$.

Instead of presenting in full these (somewhat tedious) computations, we will reframe the operators $T(n)$ from a more conceptual point of view from which these properties are easily derived.

We have an action of $\mathrm{GL}_2^+(\mathbf{Z}) = \{A \in \mathrm{GL}_2(\mathbf{Z}) : |A| > 0\}$ on $M_k(\Gamma)$ via the **slash operator**

$$f|_A(z) = |A|^{k/2}(cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right).$$

Further let

$$\Delta_n = \left\{\begin{pmatrix} a & b \\ & d\end{pmatrix} : ad = n,\ 0 \leq b < d\right\} \subset G_n = \{A \in \mathrm{GL}_2^+(\mathbf{Z}) : |A| = n\}.$$

Remark 118. *With some matrix algebra, one can show that the set Δ_n parametrizes $\Gamma\backslash G_n$, i.e., we have the disjoint coset decomposition*

$$G_n = \bigcup_{\rho\in\Delta_n} \Gamma\rho.$$

Definition 119. *For each $n \geq 1$, $f \in M_k(\Gamma)$, we have the* **Hecke operator**

$$T(n)f(z) = n^{k/2-1}\sum_{\rho\in\Delta_n} f|_\rho.$$

A direct computation shows that this definition coincides with (b) above:

$$\begin{aligned} T(n)f(z) &= n^{k-1}\sum_{ad=n} d^{-k}\sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right) \\ &= \frac{1}{n}\sum_{ad=n} a^k \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right) \end{aligned}$$

and replacing f by its q-expansion, we recover (6.2):

$$\begin{aligned} T(n)\left(\sum_{m\geq 0} a(m)q^m\right) &= n^{k-1}\sum_{ad=n} d^{-k}\sum_{m\geq 0} a(m)q^{ma/d}\sum_{b=0}^{d-1} e(mb/d) \\ &= n^{k-1}\sum_{ad=n} d^{1-k}\sum_{\substack{m\geq 0\\ d|m}} a(m)q^{ma/d} \\ &= n^{k-1}\sum_{d|n} d^{1-k}\sum_{m\geq 0} a(md)q^{mn/d} \\ &= \sum_{d|n} d^{k-1}\sum_{m\geq 0} a(mn/d)q^{md} \\ &= \sum_{m\geq 0}\left(\sum_{d|(m,n)} d^{k-1}a(mnd^{-2})\right)q^m. \end{aligned}$$

THEOREM 42. *$T(n)$ is a linear operator on $M_k(\Gamma)$ and further keeps the subspace $S_k(\Gamma)$ invariant.*

PROOF. Let $f \in M_k(\Gamma)$. In particular, $f|_\gamma = f$ for all $\gamma \in \Gamma$. Looking at (b), we see that for the first assertion, it suffices to prove that $T(n)f|_\gamma = T(n)f$.

Let $\gamma \in \Gamma$. For each $\rho \in \Delta_n$, we have $\rho\gamma = \gamma'\rho'$ for some $\gamma' \in \Gamma$, $\rho' \in \Delta_n$ by Remark 118. Then

$$T(n)f|_\gamma = n^{k/2-1}\sum_{\rho\in\Delta_n} f|_{\rho\gamma} = n^{k/2-1}\sum_{\rho'\in\Delta_n} f|_{\gamma'}|_{\rho'} = n^{k/2-1}\sum_{\rho'\in\Delta_n} f|_{\rho'} = T(n)f.$$

Further, by definition, if $f = \sum a(m)q^m$, then

$$T(n)f = \sum_{m\geq 0} b(m)q^m \quad \text{with } b(0) = a(0)\sigma_{k-1}(n) = 0.$$

Hence if $f \in S_k(\Gamma)$, $a(0) = 0$ and $b(0) = 0$, i.e., $T(n)f \in S_k(\Gamma)$. □

***

Suppose that $f \in M_k(\Gamma)$ is an eigenfunction of **all** Hecke operators $T(n)$, that is, for each $n \geq 1$, $T(n)f = \lambda(n)f$. At the level of Fourier coefficients, this gives

$$\sum_{d|(m,n)} d^{k-1}a(mnd^{-2}) = \lambda(n)a(m)$$

for all $m, n \in \mathbf{N}$, whereby $\lambda(n)$ is given (taking $m = 1$ above) by

$$\lambda(n)a(1) = a(n)$$

for each $n \geq 1$. In particular, $f = a(0)$ if $a(1) = 0$. Suppose that $a(1) \neq 0$, then (up to replacing f by $f/a(1)$) we may assume that $a(1) = 1$. Then $\lambda(n) = a(n)$ and so

$$a(n)a(m) = \sum_{d|(m,n)} d^{k-1} a(mnd^{-2})$$

and hence

$$a(nm) = a(n)a(m) \quad (m, n) = 1$$
$$a(p^{j+1}) = a(p^j)a(p) - p^{k-1}a(p^{j-1})$$

for all $j \geq 2$.

Definition 120. *We say that $f = \sum a(m)q^m \in M_k(\Gamma)$ is a* **normalized Hecke eigenform** *(or Hecke form) if $T(n)f = \lambda(n)f$ for all $n \geq 1$ and $a(1) = 1$.*

***

THEOREM 43. *The space $M_k(\Gamma)$ of modular forms of weight k admits an orthonormal basis of normalized Hecke eigenforms.*

PROOF. We will use the orthogonal decomposition $M_k(\Gamma) = \langle E_k \rangle \oplus S_k(\Gamma)$, where we recall that for $k > 2$, the first subspace is spanned by the Eisenstein series

$$E_k(z) = 1 + \frac{(-1)^{k/2}(2\pi)^k}{\Gamma(k)\zeta(k)} \sum_{n \geq 1} \sigma_{k-1}(n)q^n.$$

The divisor function $\sigma_{k-1}(n)$ verifies (1) and (2), hence

$$\sigma_{k-1}(m)\sigma_{k-1}(n) = \sum_{d|(m,n)} d^{k-1}\sigma_{k-1}(mnd^{-2}),$$

and it follows that E_k is a normalized Hecke eigenform. Hence it suffices to check that $S_k(\Gamma)$ admits an orthonormal basis of Hecke forms.

We view $S_k(\Gamma)$ as an inner product space with respect to the Petersson inner product. Then

$$\langle T(n)f, g \rangle = n^{k/2-1} \sum_{\rho \in \Delta_n} \int_{\Gamma\backslash\mathbf{H}} f|_\rho \overline{g} d\mu = n^{k/2-1} \sum_{\rho \in \Delta_n} \int_{\Gamma\backslash\mathbf{H}} f\overline{g}|_\rho d\mu = \langle f, T(n)g \rangle ,$$

i.e. $T(n)$ is self-adjoint and we can further check by direct computation that $T(m)T(n) = T(n)T(m)$ for all $m, n \geq 1$. By linear algebra, since the linear operators $T(n)$ commute and are self-adjoint, there exists a basis of simultaneous eigenfunctions for all $T(n)$. Since $T(n)$ are self-adjoint the eigenvalues are real, and for $f = \sum a(n)q^n \neq g = \sum b(n)q^n$, we have

$$a(n) \langle f, g \rangle = \langle T(n)f, g \rangle = \langle f, T(n)g \rangle = b(n) \langle f, g \rangle$$

hence $\langle f, g \rangle = 0$. □

We now can conclude that if $f \in S_k$ is a normalized Hecke eigenform, then

$$L(s,f) = \prod_p (1 - a(p)p^{-s} + p^{k-1-2s})^{-1}.$$

Remark 121. *Note in passing that Ramanujan's conjecture does not extend to non-cuspidal forms; recall that* $\sigma_{k-1}(n) \asymp n^{k-1}$.

6.4. Maass forms and nonholomorphic/spectral Eisenstein series

For simplicity, we will again assume that $\Gamma = \mathrm{PSL}(2,\mathbf{Z})$. With respect to the hyperbolic metric on $\mathbf{H}$, the Laplace operator is given by

$$\Delta = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right).$$

Definition 122. *A* **Maass form** *is a smooth function f on $\mathbf{H}$ such that*

- *f is Γ-invariant, i.e., $f(\gamma z) = f(z)$;*
- *f is an eigenfunction of Δ;*
- *$f(x+iy) = O(y^N)$ as $y \to \infty$ for some $N > 0$.*

Once again we can construct examples by forming **nonholomorphic Eisenstein series**

$$E(z,s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \mathrm{Im}(\gamma z)^s = \frac{1}{2}\sum_{(c,d)=1} \frac{y^s}{|cz+d|^{2s}}.$$

By comparison with the holomorphic Eisenstein series

$$E_k = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \frac{1}{(cz+d)^k},$$

$E(z,s)$ can be seen to converge absolutely whenever $\mathrm{Re}(s) > 1$, and its Fourier expansion is derived similarly. In fact — in anticipation of Theorem 44 we will use the following normalization to express the Fourier expansion. Let

$$E^*(z,s) = \xi(2s)E(z,s) = \pi^{-s}\Gamma(s)\zeta(2s)E(z,s).$$

Then its Fourier expansion is given by

$$E^*(z,s) = \xi(2s)y^s + \xi(2-2s)y^{1-s} + 2\sum_{n\neq 0} |n|^{s-1/2}\sigma_{1-2s}(|n|)\sqrt{y}K_{s-1/2}(2\pi|n|y)e(nx),$$

where the K-Bessel function is given by the everywhere converging Mellin transform

$$K_s(y) = \frac{1}{2}\int_0^\infty e^{-y(t+t^{-1})/2}t^s\frac{dt}{t}$$

for which we have the simple estimate

Lemma 123. *We have $K_{s-1/2}(2\pi|n|y) = O(e^{-\pi y})$ as $y \to \infty$.*

Proof. It is easy to see that if $a, b > 2$, then $ab > a + b$. Hence $e^{-ab} < e^{-a}e^{-b}$. The estimate now follows from the definition of $K_{s-1/2}(2\pi|n|y)$ using $a = \pi|n|y$ and $y > 2/\pi$. □

It follows immediately from the Fourier expansion that

THEOREM 44. *$E^*(z,s)$ has meromorphic continuation to all $s \in \mathbf{C}$ with simple poles at $s=0$ and $s=1$ that satisfies $E^*(z,1-s) = E^*(z,s)$.*

We now check that $E(z,s)$, for $\mathrm{Re}(s) > 1$, is indeed a Maass form. The Γ-invariance follows from the definition of $E(z,s)$ by averaging over cosets in $\Gamma_\infty \backslash \Gamma$. The polynomial growth can be seen from the Fourier expansion above. Finally, we leave to the reader to check that $(\Delta + s(1-s))y^s = 0$.

Exercise 124. *Let f be a smooth function on $\mathbf{H}$ and let $A \in \mathrm{SL}_2(\mathbf{R})$. Check that $\Delta f(Az) = (\Delta f)(Az)$.*

Since the action of Δ commutes with the action of $\mathrm{PSL}(2,\mathbf{R})$ by Möbius transformation, we conclude that $(\Delta + s(1-s))E(z,s) = 0$. That $E(z,s)$ is an eigenfunction of the Laplacian explains the appearance of the K-Bessel functions in its Fourier expansion. Indeed, if we write

$$E(z,s) = \sum_{n \in \mathbf{Z}} a_n(y,s) e(nx),$$

then $(\Delta + s(1-s))E(z,s) = 0$ yields second order linear differential equations

$$y^2 a_n''(y,s) + \left(s(1-s) - 4\pi^2 n^2 y^2\right) a_n(y,s) = 0 \tag{6.3}$$

with two independent solutions. When $n = 0$, it is easy to check these are y^s and y^{1-s}. When $n \neq 0$, (6.3) is a Bessel differential equation for which the K-Bessel function is a solution.

***

Factsheet

- The spectral theorem says that the space is given by the direct decomposition $\overline{\mathcal{C}} \oplus \overline{\mathcal{E}}$, where $\mathcal{E}$ is the space spanned by the Eisenstein series $E(\cdot, \frac{1}{2} + it)$, $(t \in \mathbf{R})$, and $\mathcal{C}$ is the space spanned by **Maass cusp forms**, i.e., Maass forms that vanish in the cusp with
$$\int_0^1 f(x+iy)dx = 0.$$
- The space $\mathcal{C}$ admits an orthonormal basis of normalized Hecke–Maass eigenforms, that is of functions that are simultaneously eigenfunctions of the Laplacian and of all Hecke operators. For such forms, the associated L-functions have Euler product
$$L(s,f) = \prod_p (1 - a(p)p^{-s} + p^{-2s})^{-1}.$$
- While the existence of such a basis is known — first shown via Selberg's trace formula in the 1950s but see also the argument of Lindenstrauss and Venkatesh exposed in [1] — there is no known explicit example of a Maass cusp form for $\mathrm{PSL}(2,\mathbf{Z})$. There are however known explicit examples for certain finite index subgroups of the modular groups, due to Maass.

- The Ramanujan conjecture for Maass cusp forms, i.e., the assertion that
$$1 - a(p)p^{-s} + p^{-2s}$$
is nonvanishing for $\mathrm{Re}(s) > 0$, is still not known (but is believed to be true).

6.5. Rankin–Selberg L-functions

Let $f = \sum a_n q^n, g = \sum b_n q^n \in S_k(\Gamma)$. We will assume moreover that f and g are Hecke eigenforms. Around the same time (1939–1940), Rankin and Selberg (independently) studied the analytic properties of the Dirichlet series

$$L(s, f \times g) = \sum_{n \geq 1} \frac{a_n b_n}{n^s}$$
$$L(s, f \otimes g) = \zeta(2s - 2k + 2)L(s, f \times g).$$

The Rankin–Selberg construction builds on the following computation:

Proposition 125. *Let Φ be a smooth Γ-invariant function on $\mathbf{H}$ for which $\Phi(x+iy) = O(y^{-N})$ for all $N > 0$ as $y \to \infty$. Then*

$$\Lambda(s) := \int_0^\infty \int_0^1 \Phi(x + iy)\, dx\, y^{s-1} \frac{dy}{y} = \int_{\Gamma\backslash\mathbf{H}} \Phi(z) E(z, s) d\mu(z)$$

Proof. Let $\mathrm{Re}(s) > 1$. Then

$$\begin{aligned}\int_{\Gamma\backslash\mathbf{H}} \Phi(z)E(z,s)d\mu(z) &= \sum_{\gamma \in \Gamma_\infty\backslash\Gamma} \int_{\Gamma\backslash\mathbf{H}} \Phi(\gamma z)\mathrm{Im}(\gamma z)^s d\mu(z) \\ &= \int_{\Gamma_\infty} \Phi(x+iy) y^s \frac{dxdy}{y^2} \\ &= \int_0^\infty \int_0^1 \Phi(x+iy) dx y^{s-1} \frac{dy}{y},\end{aligned}$$

where for the last equality we used the fundamental domain for $\Gamma_\infty\backslash\mathbf{H}$ given by $\{(x, y) : 0 \leq x < 1,\ y > 0\}$. □

Theorem 45. *Let $f, g \in S_k(\Gamma)$ as above. Then*

$$\Lambda(s, f \otimes \overline{g}) = (2\pi)^{-2s}\Gamma(s)\Gamma(s - k + 1)L(s, f \otimes \overline{g})$$

admits a meromorphic continuation to $s \in \mathbf{C}$ with at most simple poles at $s = k, k-1$, satisfying the functional equation

$$\Lambda(s, f \otimes \overline{g}) = \Lambda(2k - 1 - s, f \otimes \overline{g}).$$

Proof. For f, g as above, the function $\Phi(z) = f(z)\overline{g(z)}y^k$ satisfies the conditions of Proposition 125 and

$$\int_0^1 \Phi(x + iy)dx = \sum_{m,n \geq 1} a_m \overline{b_n} e^{-2\pi(m+n)y)} y^k \int_0^1 e((m - n)x)dx = \sum_{n \geq 1} a_n \overline{b_n} e^{-4\pi n y} y^k.$$

so that

$$\Lambda(s) = \sum_{n\geq 1} a_n\overline{b_n} \int_0^\infty e^{-4\pi ny} y^{k+s-1} \frac{dy}{y} = \frac{\Gamma(s+k-1)}{(4\pi)^{k+s-1}} \sum_{n\geq 1} \frac{a_n\overline{b_n}}{n^{s+k-1}}$$

or equivalently

$$\Lambda(s-k+1) = \frac{\Gamma(s)}{(4\pi)^s} L(s, f\times \overline{g}).$$

Now let

$$\Lambda^*(s) = \xi(2s)\Lambda(s) = \int_{\Gamma\backslash \mathbf{H}} \Phi(z) E^*(z,s) d\mu(z),$$

which has a meromorphic continuation with simple poles at $s = 0, 1$ and satisfies the functional equation $\Lambda^*(1-s) = \Lambda^*(s)$ — since the completed Eisenstein series do. Then

$$\Lambda^*(s-k+1) = \pi^{-s+k-1} \frac{\Gamma(s-k+1)\Gamma(s)}{(4\pi)^s} L(s, f\otimes \overline{g}) = \pi^{k-1}\Lambda(s, f\otimes \overline{g}).$$

The analytic properties follow and

$$\Lambda(2k-s-1, f\otimes \overline{g}) = \pi^{1-k}\Lambda^*(k-s) = \pi^{1-k}\Lambda^*(s-k+1) = \Lambda(s, f\otimes \overline{g}).$$

□

Remark 126. *$L(s, f\times g)$ is not as nice an analytic function as $L(s, f\otimes g)$; its meromorphic continuation has infinitely many poles at the zeros of the Riemann zeta function.*

We take note (without proof) of the following result:

THEOREM 46. *The Rankin–Selberg L-function has Euler product expansion*

$$L(s, f\otimes \overline{g}) = \prod_p \prod_{i,j=1,2} \left(1 - \alpha_i(p)\alpha_j(p)p^{-s}\right)^{-1},$$

where $1 - a(p)p^{-s} + p^{k-1-2s} = (1-\alpha_1(p)p^{-s})(1-\alpha_2(p)p^{-s})$.

***

Application 1: Rankin's estimate towards the Ramanujan conjecture

Lemma 127. *The residue of $E(z,s)$ at $s=1$ is $3/\pi$.*

PROOF. Because $(\Delta + s(1-s))E(z,s) = 0$, the residue is a harmonic function in z, and hence a constant, which we denote by c. We now show how to compute c.

Consider the truncated Eisenstein series

$$E(z,s) - y^s = \sum_{\substack{\gamma\in\Gamma_\infty\backslash\Gamma \\ \gamma\notin\Gamma_\infty}} \mathrm{Im}(\gamma z)^s$$

Applying the Rankin–Selberg unfolding:

$$\begin{aligned}\int_{\Gamma\backslash\mathbf{H}} (E(z,s) - y^s)\, d\mu(z) &= \int_{\Gamma_\infty\backslash\mathbf{H}-\mathcal{T}} y^s d\mu(z) \\ &= \int_0^1 \int_0^{A(x)} y^{s-1} \frac{dy}{y}\, dx = \int_0^1 \frac{A(x)^{s-1}}{s-1} dx.\end{aligned}$$

Now we consider the residue at $s = 1$ on both sides; this gives (see p. 57)

$$c|\mathcal{T}| = c\frac{\pi}{3} = 1.$$

□

THEOREM 47. *Let $f = \sum a(n)q^n \in S_k(\Gamma)$. Then $|a(n)| \ll n^{\frac{k}{2}-\frac{1}{5}}$.*

SKETCH OF PROOF. The simpler L-function

$$L(s, f \times \overline{f}) = \sum_{n\geq 1} \frac{|a(n)|^2}{n^s}$$

has a meromorphic continuation to $\mathrm{Re}(s) > k-1$ with a simple pole at $s = k$ by Theorem 45. By Perron's formula, we have

$$\sum_{n\leq X}{}' |a(n)|^2 = \frac{1}{2\pi i}\int_{(\sigma)} L(s, f \times \overline{f})\frac{X^s}{s}ds$$

for $\sigma > k$. With a careful application of contour shifting Rankin showed that

$$\sum_{n\leq X} |a(n)|^2 = \mathrm{Res}_{s=k} L(s, f \times \overline{f})\frac{X^s}{s} + O\left(X^{k-2/5}\right). \tag{6.4}$$

To compute the residue (and see that it is nonzero), we use that

$$L(s, f \times \overline{f}) = \frac{(4\pi)^s}{\Gamma(s)}\int_{\Gamma\backslash\mathbf{H}} |f(z)|^2 y^k E(z, s-k+1)d\mu(z).$$

Then

$$\mathrm{Res}_{s=k} L(s.f \times \overline{f}) = \frac{(4\pi)^k}{\Gamma(k)}\langle f, f\rangle \,\mathrm{Res}_{s=1} E(z, s),$$

where $\langle f, f\rangle$ is the Petersson inner product and the (constant) residue of the Eisenstein series is given by the previous lemma.

To obtain an estimate on individual Fourier coefficients, we only need that the summands on the LHS of (6.4) are positive:

$$|a(n)|^2 = \sum_{m\leq n} |a(m)|^2 - \sum_{m\leq n-1} |a(m)|^2 = c(n^k - (n-1)^k) + O\left(n^{k-2/5}\right) = O\left(n^{k-2/5}\right)$$

□

***

Application 2: Ramanujan's identity for divisor functions. At the beginning of this chapter, we have seen that

$$\sum_{n=1}^{\infty} \frac{\sigma_{k-1}(n)}{n^s} = \zeta(s)\zeta(s-k+1)$$

is essentially the L-function of the Eisenstein series E_k. Consider this time the L-function associated to the nonholomorphic Eisenstein series $E(z,\nu)$. By analogy, we define this L-function to be given by

$$\begin{aligned} L(s,E_{s'}) &= \sum_{n\geq 1} \frac{\sigma_{1-2s'}(n) n^{s'-1/2}}{n^s} \\ &= \sum_{n\geq 1} \left(\sum_{d|n} d^{1-2s'} \right) n^{s'-s-1/2} \\ &= \zeta(s-s'+1/2)\zeta(s+s'-1/2) =: \zeta(s-\mu_1)\zeta(s-\mu_2), \end{aligned}$$

with

$$\mu_1 = s' - 1/2, \quad \mu_2 = 1/2 - s'.$$

Its Rankin–Selberg convolution is given by

$$\begin{aligned} L(s, E_{s'} \times E_{s''}) &\coloneqq \sum_{n\geq 1} \sigma_{1-2s'}(n) n^{s'-s-1/2} \\ L(s, E_{s'} \otimes E_{s''}) &\coloneqq \zeta(2s) L(s, E_{s'} \times E_{s''}). \end{aligned}$$

By Theorem 46, the latter has Euler product

$$L(s, E_{s'} \otimes E_{s''}) = \prod_p \prod_{i,j=1,2} (1 - p^{\mu_i + \mu_j} p^{-s})^{-1}.$$

Comparing the two expressions for $L(s, E_{s'} \otimes E_{s''})$ yields the Ramanujan identity

$$\sum_{n\geq 1} \frac{\sigma_a(n)\sigma_b(n)}{n^s} = \frac{\zeta(s)\zeta(s-a)\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)}.$$

6.6. Twisting and Weil's converse theorem (1967)

Weil extended Hecke's theorem to holomorphic modular forms of higher level. One may first wonder whether the L-function attached to a holomorphic cusp form of higher level also enjoys analytic properties

(i) has analytic continuation to the whole $\mathbf{C}$-plane
(ii) that is bounded in every vertical strip $\sigma_1 \leq \mathrm{Re}(s) \leq \sigma_2$,
(iii) and satisfies $\Lambda(s,f) = i^k \Lambda(k-s,f)$.

Recall how in level 1, the proof of these two properties builds on two elements: (1) the Mellin integral representation $\Lambda(s,f) = \int_0^\infty f(iy) y^s \frac{dy}{y}$ (*à la Riemann*) and (2) that $\Gamma(1) = \langle S, T \rangle$ so that the modularity implies $f = \sum a_n q^n = O(e^{-2\pi y})$ as $y \to \infty$ and $f(i/y) = (iy)^k f(iy)$. (See Section 1 of this chapter.)

For higher level, the first missing element is that $S \notin \Gamma_0(N)$ if $N \geq 2$. In its place, one uses the **Fricke involutions**

$$\omega_N = \begin{pmatrix} & -1 \\ N & \end{pmatrix}.$$

Let $f \in S_k(N,\psi)$, for some Dirichlet character $\psi \pmod N$. Then for $g := f|_{\omega_N}$, we have

$$f(iy) = N^{-k/2}(i/y)^k g(i/(Ny))$$

and thus

$$\text{(iii}') \qquad \Lambda(s,f) = N^{-k/2}i^k \int_0^\infty g(i/(Ny))y^{s-k}\frac{dy}{y} = N^{k/2-s}i^k\Lambda(k-s,g).$$

We can also write

$$\Lambda(s,f) = \int_1^\infty f(iy)y^s\frac{dy}{y} + N^{k/2-s}i^k\int_{1/N}^\infty g(iy)y^{k-s}\frac{dy}{y}$$

from which we can then deduce (i) and (ii) for $\Lambda(s,f)$ (as well as for $\Lambda(s,g)$) using that

Lemma 128. *If $f \in S_k(N,\psi)$, $g := f|_{\omega_N} \in S_k(N,\overline{\psi})$.*

PROOF. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$ and note that

$$\omega_N \begin{pmatrix} a & b \\ c & d\end{pmatrix} = \begin{pmatrix} d & -c/N \\ bN & a\end{pmatrix}\omega_N,$$

i.e., ω_N normalizes $\Gamma_0(N)$. We immediately recover

$$g|_\gamma = f|_{\omega_N\gamma} = \left(f|_{\left(\begin{smallmatrix} d & -c/N \\ bN & a\end{smallmatrix}\right)}\right)|_{\omega_N} = (\psi(a)f)\,|_{\omega_N} = \psi(a)f|_{\omega_N} = \overline{\psi(d)}g.$$

Let p be a cusp for $\Gamma_0(N)$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$ such that $\gamma p = p$. Then $\left(\begin{smallmatrix} d & -c/N \\ bN & a\end{smallmatrix}\right)\omega_N p = \omega_N p$, i.e., $\omega_N p$ is also a cusp of $\Gamma_0(N)$. We conclude that g is holomorphic on $\mathbf{H}$ and at each cusp of $\Gamma_0(N)$. Thus $g \in S_k(N,\overline{\psi})$. □

To summarize, for $f \in S_k(N,\psi)$, the completed L-function $\Lambda(s,f)$

(i) has analytic continuation to the whole $\mathbf{C}$-plane
(ii) that is bounded in every vertical strip $\sigma_1 \le \mathrm{Re}(s) \le \sigma_2$,
(iii)' and satisfies $\Lambda(s,f) = N^{k/2-s}i^K\Lambda(k-s,g)$ for $g = f|_{\omega_N}$.

***

For the converse, one needs to prove that f has the correct modular transformation for each generator of $\Gamma_0(N)$. In level 1, this reduces to checking the modular transformation under the generator S and this essentially amounts to the functional equation. Because the number of generators of $\Gamma_0(N)$ grows as N grows, a single functional equation cannot suffice to deduce the modularity of f for (general) higher level. Weil approached the problem by **twisting**: Fix $D > 0$ and a *primitive* Dirichlet character $\chi \pmod D$ and define, for $f = \sum a(n)q^n \in S_k(N,\psi)$, the twisted form

$$f_\chi = \sum_{n\ge 1}\chi(n)a(n)q^n.$$

The next lemma attests that modular forms behave nicely under twisting.

Lemma 129. *Let χ be primitive mod D. If $f \in S_k(N,\psi)$, then $f_\chi \in S_k(D^2N, \psi\chi^2)$. Moreover, $f_\chi|_{\omega_{D^2N}} = w(\chi) g_{\overline{\chi}}$, where*

$$w(\chi) = D^{-1}\chi(N)\psi(D)\tau(\chi)^2.$$

PROOF. Since χ is primitive we have that $\tau(\overline{\chi})\chi(n) = \sum \overline{\chi}(u)e(nu/D)$ for each $n \in \mathbf{Z}$ and where $|\tau(\overline{\chi})| = \sqrt{D}$. Then

$$\tau(\overline{\chi}) f_\chi = \sum_{u\,(D)} \overline{\chi}(u) f|_{\left(\begin{smallmatrix} 1 & u/D \\ & 1 \end{smallmatrix}\right)}.$$

Check that

$$\begin{pmatrix} 1 & u/D \\ & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+cu/D & b - bcdu/D - cd^2u^2/D^2 \\ c & d - cd^2u/D \end{pmatrix}\begin{pmatrix} 1 & d^2u/D \\ & 1 \end{pmatrix}.$$

Then if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(ND^2)$, we have $\gamma' = \left(\begin{smallmatrix} 1 & u/D \\ & 1 \end{smallmatrix}\right)\gamma\left(\begin{smallmatrix} 1 & -d^2u/D \\ & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$ and

$$\begin{aligned}
\tau(\overline{\chi}) f_\chi|_\gamma &= \sum_{u\,(D)} \overline{\chi}(u) f|_{\left(\begin{smallmatrix} 1 & u/D \\ & 1 \end{smallmatrix}\right)\gamma} \\
&= \sum_{u\,(D)} \overline{\chi}(u) f|_{\gamma'\left(\begin{smallmatrix} 1 & d^2u/D \\ & 1 \end{smallmatrix}\right)} \\
&= \psi(d) \sum_{u\,(D)} \overline{\chi}(u) f|_{\left(\begin{smallmatrix} 1 & d^2u/D \\ & 1 \end{smallmatrix}\right)} \\
&= (\psi\chi^2)(d) \sum_{u\,(D)} \overline{\chi}(ud^2) f|_{\left(\begin{smallmatrix} 1 & d^2u/D \\ & 1 \end{smallmatrix}\right)} \\
&= \tau(\overline{\chi})(\psi\chi^2)(d) f_\chi,
\end{aligned}$$

which is what we had to show to prove the first part of the statement. The second statement is left as an exercise. □

Proceeding as above, we arrive to the conclusion that $\Lambda(s, f_\chi)$

(i) has analytic continuation to the whole $\mathbf{C}$-plane
(ii) that is bounded in every vertical strip $\sigma_1 \le \mathrm{Re}(s) \le \sigma_2$,
(iii)' and satisfies $\Lambda(s, f_\chi) = (ND^2)^{k/2-s} w(\chi) i^k \Lambda(k-s, g_{\overline{\chi}})$ for $g = f|_{\omega_N}$.

We can now state Weil's converse theorem.

THEOREM 48 (Weil). *Fix N positive, and ψ a Dirichlet character (mod N). Let (a_n), $(b_n) \subset \mathbf{C}$ be sequences such that a_n, $b_n = O(n^K)$ for some $K > 0$ and set*

$$\Lambda_1(s,\chi) = (2\pi)^{-s}\Gamma(s)\sum_{n\ge 1}\frac{\chi(n)a_n}{n^s}, \quad \Lambda_2(s,\chi) = (2\pi)^{-s}\Gamma(s)\sum_{n\ge 1}\frac{\chi(n)b_n}{n^s}$$

for any primitive character χ (mod D), with $(D,N) = 1$.

Let $\mathcal{S}$ be a finite set of prime numbers such that $p \mid N \implies p \in S$. If for every primitive character χ of conductor $D = 1$ or $D \notin \mathcal{S}$ prime, $\Lambda_1(s,\chi)$ and $\Lambda_2(s,\overline{\chi})$

(i) *have analytic continuation to the whole $\mathbf{C}$-plane,*
(ii) *are bounded in every vertical strip $\sigma_1 \le \mathrm{Re}(s) \le \sigma_2$,*

(iii) *and satisfy* $\Lambda_1(s,\chi) = (ND^2)^{k/2-s} w(\chi) i^k \Lambda_2(k-s,\overline{\chi})$,

then $f \in M_k(N,\psi)$, $g \in M_k(N,\overline{\psi})$ *and* $g = f|_{\omega_N}$.

We ran out of time to cover the proof in class; the interested reader can consult Chapter 1.5 in [2].

Weil's converse theorem had important ramifications for the **modularity conjecture**. In 1955, Taniyama conjectured that L-functions attached to elliptic curves were related to modular forms. For instance, if $L(s,E) = \sum a_n n^{-s}$ is the L-series attached to an elliptic curve E, then $f = \sum a_n q^n$ should be a Hecke weight 2 form of some level N. Weil used his theorem to precise the expected level. Following work of Frey, Serre, and Ribet in the 1980s, the modularity conjecture was shown to imply Fermat's last theorem. The conjecture was eventually settled by Wiles (and Taylor–Wiles) in 1995.

Bibliography

[1] Bergeron, Spectral Theory of Hyperbolic Surfaces, Springer UniText 2016.
[2] Bump, Automorphic Forms and Representations.
[3] Iwaniec, Lectures on the Riemann zeta function. American Mathematical Society University Lecture Series, 2014.
[4] Iwaniec, Topics in classical automorphic forms. Graduate Studies in Mathematics, AMS, 1997.
[5] Iwaniec, Kowalski, Analytic Number Theory. American Mathematical Society Colloquium Publications, 2004.
[6] Montgomery, Vaughan, Multiplicative Number Theory: I. Classical Theory. Cambridge studies in advanced mathematics, 2006.
[7] Sarnak, *Kloosterman, quadratic forms and modular forms* NAW (2000), 140–145.
[8] Zagier, Elliptic modular forms, in: 1-2-3 of Modular Forms.