An Introduction to Analytic Number Theory

Henry Twiss

2024

Contents

Part I

Background Material

Chapter 1

Preliminaries

A good selection of topics that need some discussion before introducing analytic number theory are the following:

- Asymptotics,
- Dirichlet Characters,
- Special Sums,
- Lattices,
- The Gamma Function,
- Locally Compact Groups,
- Integration Techniques & Transforms,

This is not an extensive list, but it is a decent one. In the interest of keeping this text almost completely self-contained, this chapter is dedicated to the basics of these topics as they are the gadgets that will take center stage in our later investigations. The concepts presented in this chapter are tools in a tool box rather than being pure analytic number theory. In order to improve the readability of the remainder of the text we will use the results presented here without reference unless it is a matter of clarity. As for standard knowledge, we assume familiarity with basic number theory, complex analysis, real analysis, functional analysis, topology, and algebra. We have also outsourced specific subtopics to the appendix and we will reference them when necessary.

1.1 Notational Conventions

Here we make some notational conventions throughout the rest of the text unless specified otherwise:

- The symbols $\subset$ and $\supset$ denote strict containment.
- Any ring is understood to be a commutative ring with 1.
- The finite field with p elements $\mathbb{F}_p$ stands for $\mathbb{Z}/p\mathbb{Z}$.
- A prime $p \in \mathbb{Z}$ is always taken to be positive.

- The symbol ε denotes a small positive constant ($\varepsilon > 0$) that is not necessarily the same from line to line.
- If $a \in (\mathbb{Z}/m\mathbb{Z})^*$, we will always let $\overline{a}$ denote the multiplicative inverse. That is, $a\overline{a} \equiv 1 \pmod{m}$.
- By analytic we mean real analytic or complex analytic accordingly.
- For the complex variables z, s, and u, we write

$$z = x + iy, \quad s = \sigma + it, \quad \text{and} \quad u = \tau + ir,$$

  for the real and imaginary parts of these variables respectively. In some cases we make exceptions and this will always be clear from context. Moreover, in certain expressions we often write $\mathrm{Im}(z)$ for clarity.
- If $r \in \mathbb{Z}$ denotes the order of a possible pole of a complex function, $r \geq 0$ if it is a pole and $r \leq 0$ if it is a zero.
- The nontrivial zeros of an L-function will be denoted by $\rho = \beta + i\gamma$ unless specified otherwise.
- By log we will always mean the principal branch of the logarithm.
- For a sum $\sum$ over integers satisfying a congruence condition, $\sum'$ will denote the sum restricted to relatively prime integers satisfying the same congruence.
- We will write $\int_{(a)}$ for the complex integral over the line whose real part is a and with positive orientation.
- $\delta_{a,b}$ will denote the indicator function for $a = b$. That is, $\delta_{a,b} = 1, 0$ according to if $a = b$ or not.

1.2 Asymptotics

Throughout we assume $f, g : \mathbb{R}^n \to \mathbb{C}$. Much of the language of analytic number theory is given in terms of asymptotics (or estimates or bounds) as they allows us to discuss approximate growth and dispense with superfluous constants. For this reason, asymptotics will be the first material that we will present. The asymptotics that we will cover are listed in the following table:

Asymptotics	Notation
Big O	$f(\mathbf{x}) = O(g(\mathbf{x}))$
Vinogradov's symbol	$f(\mathbf{x}) \ll g(\mathbf{x})$
Order of magnitude symbol	$f(\mathbf{x}) \asymp g(\mathbf{x})$
Little o	$f(\mathbf{x}) = o(g(\mathbf{x}))$
Asymptotic equivalence	$f(\mathbf{x}) \sim g(\mathbf{x})$
Omega symbol	$f(\mathbf{x}) = \Omega(g(\mathbf{x}))$

Implicit in all of these asymptotics is some limiting process $\mathbf{x} \to \mathbf{x}_0$ where $\mathbf{x}_0$ is finite or ∞. If $\mathbf{x}_0$ is finite, then it is understood that the asymptotic is assumed to hold for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$ in norm. If $\mathbf{x}_0$ is infinite, then the asymptotic is assumed to hold for sufficiently large $\mathbf{x}$. If the limiting process is not explicitly mentioned, it is assumed to be as $\mathbf{x} \to \infty$. Often times, asymptotics will hold for all admissible values of $\mathbf{x}$ and this will be clear from context although we still might suppress the specific limiting process.

Remark 1.2.1. *If $f, g : \mathbb{C}^n \to \mathbb{C}$, then the following theory still holds by identifying $\mathbb{C}^n \cong \mathbb{R}^{2n}$. Moreover, if $f, g : \mathbb{N}^n \to \mathbb{C}$ then by extending $f(\mathbf{n})$ and $g(\mathbf{n})$ to $\mathbb{R}^n$ by making them piecewise linear, so that they are piecewise continuous, the following theory still holds with $\mathbf{n}$ in place of $\mathbf{x}$. In particular, we may take $f(\mathbf{x})$ or $g(\mathbf{x})$ to be a constant function.*

Implicit in some asymptotics will there be a constant (such constants are in general not unique and any sufficiently large constant will do). Any such constant is called the **implicit constant** of the asymptotic. The implicit constant may depend on one or more parameters, ε, σ, etc. If we wish to make these dependencies known, we use subscripts. If it is possible to choose the implicit constant independent of a certain parameter then we say that the asymptotic is **uniform** with respect to that parameter. Moreover, we say that an implicit constant is **effective** if the constant is numerically computable and **ineffective** otherwise. Moreover, if we are interested in the dependence of an asymptotic on a certain parameter, say p, we will refer to the **p-aspect** to mean the part of the asymptotic that is dependent upon p.

O-estimates & Symbols

We say $f(\mathbf{x})$ **is of order** $g(\mathbf{x})$ or $f(\mathbf{x})$ is $O(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$ and write $f(\mathbf{x}) = O(g(\mathbf{x}))$ if there is some positive constant c such that

$$|f(\mathbf{x})| \le c|g(\mathbf{x})|,$$

holds as $\mathbf{x} \to \mathbf{x}_0$. We call this an **$O$-estimate** and say that $f(\mathbf{x})$ has **growth at most** $g(\mathbf{x})$. The O-estimate says that for $\mathbf{x}$ close to $\mathbf{x}_0$, the size of $f(\mathbf{x})$ grows like $g(\mathbf{x})$.

Remark 1.2.2. *Many authors assume that $g(\mathbf{x})$ is a nonnegative function so that the absolute values on $g(\mathbf{x})$ can be dropped. As we require asymptotics that will be used more generally, we do not make this assumption since one could very well replace $O(g(\mathbf{x}))$ with $O(|g(\mathbf{x})|)$. In practice this deviation causes no issue.*

As a symbol, let $O(g(\mathbf{x}))$ stand for a function $f(\mathbf{x})$ that is $O(g(\mathbf{x}))$. Then we may use the O-estimates in algebraic equations and inequalities. Note that this extends the definition of the symbol because $f(\mathbf{x}) = O(g(\mathbf{x}))$ means $f(\mathbf{x})$ is $O(g(\mathbf{x}))$. Moreover, in algebraic equations and inequalities involving O-estimates it is often customary to refer to such an O-estimate as an **error term**. The symbol $\ll$ is known as **Vinogradov's symbol** and it is an alternative way to express O-estimates. We write $f(\mathbf{x}) \ll g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$ if $f(\mathbf{x}) = O(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$. We also write $f(\mathbf{x}) \gg g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$ to mean $g(\mathbf{x}) \ll f(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$. If there is a dependence of the implicit constant on parameters, we use subscripts to denote dependence on these parameters. If both $f(\mathbf{x}) \ll g(\mathbf{x})$ and $g(\mathbf{x}) \ll f(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$, then we say $f(\mathbf{x})$ and $g(\mathbf{x})$ have the **same order of magnitude** and write $f(\mathbf{x}) \asymp g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$. We also say $f(\mathbf{x})$ has **growth** $g(\mathbf{x})$. If there is a dependence of the implicit constant on parameters, we use subscripts to denote dependence on these parameters. From the definition of the O-estimate, this is equivalent to the existence of positive constants c_1 and c_2 such that

$$c_1|g(\mathbf{x})| \le |f(\mathbf{x})| \le c_2|g(\mathbf{x})|.$$

Equivalently, we can interchange $f(\mathbf{x})$ and $g(\mathbf{x})$ in the above equation.

o-estimates & Symbols

We say $f(\mathbf{x})$ **is of smaller order than** $g(\mathbf{x})$ or $f(\mathbf{x})$ is $o(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$ and write $f(\mathbf{x}) = o(g(\mathbf{x}))$ if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| = 0,$$

provided $g(\mathbf{x}) \neq 0$ for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$ in norm. We call this an **o-estimate** and say that $f(\mathbf{x})$ has **growth less than** $g(\mathbf{x})$. The o-estimate says that for $\mathbf{x}$ close to $\mathbf{x}_0$, $g(\mathbf{x})$ dominates $f(\mathbf{x})$. If $f(\mathbf{x}) = o(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$, then $f(\mathbf{x}) = O(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$ where the implicit constant can be taken arbitrarily small by definition of the o-estimate. Therefore, o-estimates are stronger than O-estimates. As a symbol, let $o(g(\mathbf{x}))$ stand for a function $f(\mathbf{x})$ that is $o(g(\mathbf{x}))$. Then we may use the o-estimates in algebraic equations and inequalities. Note that this extends the definition of the symbol because $f(\mathbf{x}) = o(g(\mathbf{x}))$ means $f(\mathbf{x})$ is $o(g(\mathbf{x}))$. We say $f(\mathbf{x})$ **is asymptotic to** $g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$ and write $f(\mathbf{x}) \sim g(\mathbf{x})$ if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| = 1,$$

provided $g(\mathbf{x}) \neq 0$ for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$ in norm. We call this an **asymptotic** and say that $f(\mathbf{x})$ and $g(\mathbf{x})$ are **asymptotically equivalent**. It is useful to think of asymptotic equivalence as $f(\mathbf{x})$ and $g(\mathbf{x})$ being the same size in the limit as $\mathbf{x} \to \mathbf{x}_0$. Immediately from the definition, we see that this is an equivalence relation on functions. In particular, if $f(\mathbf{x}) \sim g(\mathbf{x})$ and $g(\mathbf{x}) \sim h(\mathbf{x})$ then $f(\mathbf{x}) \sim h(\mathbf{x})$. Also, if $f(\mathbf{x}) \sim g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$, then $f(\mathbf{x}) \asymp g(\mathbf{x})$ as $\mathbf{x} \to \mathbf{x}_0$ with $c_1 \leq 1 \leq c_2$. So asymptotic equivalence is stronger than being of the same order of magnitude. Also note that $f(\mathbf{x}) \sim g(\mathbf{x})$ is equivalent to $f(\mathbf{x}) = g(\mathbf{x})(1 + o(1))$ and hence implies $f(\mathbf{x}) = g(\mathbf{x})(1 + O(1))$. We say $f(\mathbf{x})$ **is of larger order than** $g(\mathbf{x})$ or $f(\mathbf{x})$ is $\Omega(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$ and write $f(\mathbf{x}) = \Omega(g(\mathbf{x}))$ if

$$\limsup_{\mathbf{x} \to \mathbf{x}_0} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| > 0.$$

provided $g(\mathbf{x}) \neq 0$ for all $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$ in norm. We call this an **Ω-estimate** and say that $f(\mathbf{x})$ has **growth at least** $g(\mathbf{x})$. Observe that $f(\mathbf{x}) = \Omega(g(\mathbf{x}))$ is precisely the negation of $f(\mathbf{x}) = o(g(\mathbf{x}))$, so that $f(\mathbf{x}) = \Omega(g(\mathbf{x}))$ means $f(\mathbf{x}) = o(g(\mathbf{x}))$ is false. This is weaker than $f(\mathbf{x}) \gg g(\mathbf{x})$ because $f(\mathbf{x}) = \Omega(g(\mathbf{x}))$ means $|f(\mathbf{x})| \geq c|g(\mathbf{x})|$ for some values of $\mathbf{x}$ arbitrarily close to $\mathbf{x}_0$ whereas $f(\mathbf{x}) \gg g(\mathbf{x})$ means $|f(\mathbf{x})| \geq c|g(\mathbf{x})|$ for all values of $\mathbf{x}$ sufficiently close to $\mathbf{x}_0$ in norm.

Algebraic Manipulation for O-estimates and o-estimates

Asymptotics become increasingly more useful when we can use them in equations to represent approximations. We catalogue some of the most useful algebraic manipulations for O-estimates and o-estimates. Most importantly, if an algebraic equation involves a O-estimate or o-estimate then it is understood that the equation is not symmetric and is interpreted to be read from left to right. That is, any function of the form satisfying the estimate on the left-hand side also satisfies the estimate on the right-hand side too. We begin with O-estimates. The trivial algebraic manipulations are collected in the proposition below:

Proposition 1.2.1. *The following O-estimates hold as $\mathbf{x} \to \mathbf{x}_0$:*

(i) If $f(\mathbf{x}) = O(g(\mathbf{x}))$ and $g(\mathbf{x}) = O(h(\mathbf{x}))$, then $f(\mathbf{x}) = O(h(\mathbf{x}))$. Equivalently, $O(O(h(\mathbf{x}))) = O(h(\mathbf{x}))$.

(ii) If $f_i(\mathbf{x}) = O(g_i(\mathbf{x}))$ for $i = 1, 2$, then $f_1(\mathbf{x})f_2(\mathbf{x}) = O(g_1(\mathbf{x})g_2(\mathbf{x}))$.

(iii) *If* $f(\mathbf{x}) = O(g(\mathbf{x})h(\mathbf{x}))$*, then* $f(\mathbf{x}) = g(\mathbf{x})O(h(\mathbf{x}))$.

(iv) *If* $f_i(\mathbf{x}) = O(g_i(\mathbf{x}))$ *for* $i = 1, 2, \ldots, n$*, then* $\sum_{1 \le i \le n} f_i(\mathbf{x}) = O\left(\sum_{1 \le i \le n} |g_i(\mathbf{x})|\right)$.

(v) *If* $f_n(\mathbf{x}) = O(g_n(\mathbf{x}))$ *for* $n \ge 1$*, then* $\sum_{n \ge 1} f_n(\mathbf{x}) = O\left(\sum_{n \ge 1} |g_n(\mathbf{x})|\right)$ *provided both* $\sum_{n \ge 1} f_n(\mathbf{x})$ *and* $\sum_{n \ge 1} |g_n(\mathbf{x})|$ *converge.*

(vi) *If* $f(\mathbf{x}) = O(g(\mathbf{x}))$ *as* $\mathbf{x} \to \mathbf{x}_0$ *and* $T(\mathbf{x})$ *is such that* $T(\mathbf{x}) \to \mathbf{x}_0$ *as* $\mathbf{x} \to \mathbf{x}_0$*, then* $(f \circ T)(\mathbf{x}) = O((g \circ T)(\mathbf{x}))$.

(vii) *If* $f(\mathbf{x}) = O(g(\mathbf{x}))$*, then* $\mathrm{Re}(f(\mathbf{x})) = O(g(\mathbf{x}))$ *and* $\mathrm{Im}(f(\mathbf{x})) = O(g(\mathbf{x}))$.

Proof. Statements (i)-(iii) and (vi) follow immediately from the definition of the O-estimate. Statement (iv) follows from the definition and the triangle inequality. Statement (v) follows in the same way as (iv) given that both sums converge. Statement (vii) follows from the definition the O-estimate and the bounds $|x| \le |z|$ and $|y| \le |z|$ for any complex z. $\square$

The most common application of Proposition 1.2.1 (vi) will be in the single variable case when $z \asymp w$ or $z \sim w$ (the latter case implying the former) where w is a function of z (usually one that is more simple than x itself). Taking $h(z) = w$, Proposition 1.2.1 (vi) says that if $f(z) = O(g(z))$ then $f(w) = O(g(w))$. In terms of Vinogradov's symbol, $f(z) \ll g(z)$ implies $f(w) \ll g(w)$. O-estimates also behave well with respect to integrals provided the functions involved are of a real variable:

Proposition 1.2.2. *Suppose* $f(\mathbf{x}) = O(g(\mathbf{x}))$ *as* $\mathbf{x} \to \infty$*,* $f(\mathbf{x})$ *and* $g(\mathbf{x})$ *are integrable on a domain where this estimate holds, and let* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *belong to this region. Then*

$$\int_{\mathbf{x}_1}^{\mathbf{x}_2} f(\mathbf{x})\, d\mathbf{x} = O\left(\int_{\mathbf{x}_1}^{\mathbf{x}_2} |g(\mathbf{x})|\, d\mathbf{x}\right).$$

Proof. This follows immediately from the definition of the O-estimate. $\square$

The next proposition is a collection of some useful expressions for simplifying equations involving O-estimates:

Proposition 1.2.3. *Let* $f(\mathbf{x})$ *be a function such that* $f(\mathbf{x}) \to 0$ *as* $\mathbf{x} \to \mathbf{x}_0$*. The following* O*-estimates hold as* $\mathbf{x} \to \mathbf{x}_0$*:*

(i) $\frac{1}{1+O(f(\mathbf{x}))} = 1 + O(f(\mathbf{x}))$.

(ii) $(1 + O(f(\mathbf{x})))^w = 1 + O(f(\mathbf{x}))$ *for any complex* w.

(iii) $\log(1 + O(f(\mathbf{x}))) = O(f(\mathbf{x}))$.

(iv) $e^{1+O(f(\mathbf{x}))} = 1 + O(f(\mathbf{x}))$.

Proof. Taking the Taylor series truncated after the first term and applying Taylor's theorem gives the following O-estimates as $z \to 0$:

(i) $\frac{1}{1+z} = 1 + O(z)$.

(ii) $(1 + z)^z = 1 + O(z)$.

(iii) $\log(1 + z) = O(z)$.

(iv) $e^z = 1 + O(z)$.

Now apply Proposition 1.2.1 (v) to each of these O-estimates, and use Proposition 1.2.1 (i). $\square$

For o-estimates, the following properties are useful:

Proposition 1.2.4. *The following o-estimates hold as $\mathbf{x} \to \mathbf{x}_0$:*

(i) If $f(\mathbf{x}) = o(g(\mathbf{x}))$ and $g(\mathbf{x}) = o(h(\mathbf{x}))$, then $f(\mathbf{x}) = o(h(\mathbf{x}))$. Equivalently, $o(o(h(\mathbf{x}))) = o(h(\mathbf{x}))$.

(ii) If $f_i(\mathbf{x}) = o(g_i(\mathbf{x}))$ for $i = 1, 2$, then $f_1(\mathbf{x})f_2(\mathbf{x}) = o(g_1(\mathbf{x})g_2(\mathbf{x}))$.

(iii) If $f(\mathbf{x}) = o(g(\mathbf{x})h(\mathbf{x}))$, then $f(\mathbf{x}) = g(\mathbf{x})o(h(\mathbf{x}))$.

(iv) If $f_i(\mathbf{x}) = o(g_i(\mathbf{x}))$ for $i = 1, 2, \ldots, n$, then $\sum_{1\le i\le n} f_i(\mathbf{x}) = o\left(\sum_{1\le i\le n} |g_i(\mathbf{x})|\right)$.

(v) If $f(\mathbf{x}) = o(g(\mathbf{x}))$ as $\mathbf{x} \to \mathbf{x}_0$ and $h(\mathbf{x})$ is such that $h(\mathbf{x}) \to \mathbf{x}_0$ as $\mathbf{x} \to \mathbf{x}_0$, then $(f \circ h)(\mathbf{x}) = o((g \circ h)(\mathbf{x}))$.

Proof. Statements (i)-(iii) and (v) follow immediately from the definition of the o-estimate. Statement (iv) follows from the definition and that $\sum_{1\le i\le n} |g_i(\mathbf{x})| \ge |g_i(\mathbf{x})|$. $\square$

Growth & Decay of Functions

We will also be interested in the growth rate of functions. There are many types of growth rates, but we will only recall the ones that are standard. Throughout let $c \ge 1$. First suppose $\mathbf{x} \to \mathbf{x}_0$. If $f(\mathbf{x}) \asymp \log^c ||\mathbf{x}||$, we say that $f(\mathbf{x})$ is of **logarithmic growth**. If $f(\mathbf{x}) \asymp ||\mathbf{x}||^c$, we say that $f(\mathbf{x})$ is of **polynomial growth**. If $f(\mathbf{x}) \asymp e^{c||\mathbf{x}||}$, we say that $f(\mathbf{x})$ is of **exponential growth**. Now suppose $\mathbf{x} \to \infty$. If $f(\mathbf{x}) \asymp \log^{-c} ||\mathbf{x}||$, we say that $f(\mathbf{x})$ is of **logarithmic decay**. If $f(\mathbf{x}) \asymp ||\mathbf{x}||^{-c}$ for some $c \ge 1$, we say that $f(\mathbf{x})$ is of **polynomial decay**. If $f(\mathbf{x}) \asymp e^{-c||\mathbf{x}||}$, we say that $f(\mathbf{x})$ is of **exponential decay**. In all of these cases, we refer to the constant c as the **order** of growth or decay respectively. If $f(\mathbf{x}) = \Omega(||\mathbf{x}||^n)$ for all $n \ge 0$, then we say $f(\mathbf{x})$ is of **rapid growth**. Alternatively, if $f(\mathbf{x}) = o(||\mathbf{x}||^{-n})$ for all $n \ge 0$, then we say $f(\mathbf{x})$ is of **rapid decay**.

1.3 Dirichlet Characters

The most important multiplicative periodic functions for an analytic number theorist are the Dirichlet characters. A **Dirichlet character** χ modulo $m \ge 1$ is an m-periodic completely multiplicative function $\chi : \mathbb{Z} \to \mathbb{C}$ such that $\chi(a) = 0$ if and only if $(a, m) > 1$. We call m the **modulus** of χ. Sometimes we will also write χ_m to denote a Dirichlet character modulo m if we need to express the dependence upon the modulus. For any $m \ge 1$, there is always the **principal Dirichlet character** modulo m which we denote by $\chi_{m,0}$ (sometimes also seen as $\chi_{0,m}$ or the ever more confusing χ_0) and is defined by

$$\chi_{m,0}(a) = \begin{cases} 1 & (a, m) = 1, \\ 0 & (a, m) > 1. \end{cases}$$

When $m = 1$, the principal Dirichlet character is identically 1 and we call this the **trivial Dirichlet character**. This is also the only Dirichlet character modulo 1, so $\chi_1 = \chi_{1,0}$. In general, we say a Dirichlet character χ is **principal** if it only takes values 0 or 1. We now discuss some basic facts of Dirichlet characters. Since $a^{\varphi(m)} \equiv 1 \pmod{m}$ by Euler's little theorem, where φ is Euler's totient function, the multiplicativity of χ implies $\chi(a)^{\varphi(m)} = 1$. Therefore the nonzero values of χ_m are $\varphi(m)$-th roots of unity. In particular, there are only finitely many Dirichlet characters of any fixed modulus m. Given two Dirichlet character χ and ψ modulo m, we define $\chi\psi$ by $\chi\psi(a) = \chi(a)\psi(a)$. This is also a Dirichlet character modulo m, so the Dirichlet characters modulo m form an abelian group denoted by X_m. If we have a Dirichlet character χ modulo m, then $\overline{\chi}$ defined by $\overline{\chi}(a) = \overline{\chi(a)}$ is also a Dirichlet character modulo m and is called the **conjugate Dirichlet character** of χ. Since the nonzero values of χ are roots of unity, if $(a, m) = 1$ then $\overline{\chi}(a) = \chi(a)^{-1}$. So $\overline{\chi}$ is the inverse of χ. This is all strikingly similar to characters on $(\mathbb{Z}/m\mathbb{Z})^*$ (see Appendix C.2), and there is a connection. To see it, by the periodicity of χ, it's nonzero values are uniquely determined by $(\mathbb{Z}/m\mathbb{Z})^*$. Then since χ is multiplicative, it descends to a character χ of $(\mathbb{Z}/m\mathbb{Z})^*$ (we abuse notation here). Conversely, if we are given a character χ of $(\mathbb{Z}/m\mathbb{Z})^*$ we can extend it to a Dirichlet character by defining it to be m-periodic and declaring $\chi(a) = 0$ if $(a, m) > 1$. We call this extension the **zero extension**. So in other words, Dirichlet characters modulo m are the zero extensions of group characters on $(\mathbb{Z}/m\mathbb{Z})^*$. Clearly zero-extension respects multiplication of characters. As groups are isomorphic to their character groups (see Proposition C.2.1), we deduce that the group of Dirichlet characters modulo m is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^*$. That is, $X_m \cong \widehat{(\mathbb{Z}/m\mathbb{Z})^*} \cong (\mathbb{Z}/m\mathbb{Z})^*$. In particular, there are $\varphi(m)$ Dirichlet characters modulo m. From now on we identify Dirichlet characters modulo m with their corresponding group characters of $(\mathbb{Z}/m\mathbb{Z})^*$. We now state two very useful relations called **orthogonality relations** for Dirichlet characters (this follows from the more general orthogonality relations in Appendix C.2 but we wish to give a direct proof):

Proposition 1.3.1.

(i) For any two Dirichlet characters χ and ψ modulo m,

$$\frac{1}{\varphi(m)} \sideset{}{'}\sum_{a \pmod{m}} \chi(a)\overline{\psi}(a) = \delta_{\chi,\psi}.$$

(ii) For any $a, b \in (\mathbb{Z}/m\mathbb{Z})^*$,

$$\frac{1}{\varphi(m)} \sum_{\chi \pmod{m}} \chi(a)\overline{\chi}(b) = \delta_{a,b}.$$

Proof. We will prove the statements separately.

(i) Denote the left-hand side by S and let b be such that $(b, m) = 1$. Then $a \to ab^{-1}$ is a bijection on $(\mathbb{Z}/m\mathbb{Z})^*$ so that

$$\frac{\chi(b)\overline{\psi}(b)}{\varphi(m)} \sideset{}{'}\sum_{a \pmod{m}} \chi(a)\overline{\psi}(a) = \frac{1}{\varphi(m)} \sideset{}{'}\sum_{a \pmod{m}} \chi(ab)\overline{\psi}(ab) = \frac{1}{\varphi(m)} \sideset{}{'}\sum_{a \pmod{m}} \chi(a)\overline{\psi}(a).$$

Consequently $\chi(b)\overline{\psi}(b)S = S$ so that $S = 0$ unless $\chi(b)\overline{\psi}(b) = 1$ for all b such that $(b, m) = 1$. This happens if and only if $\psi = \chi$ in which case $S = 1$. This proves (i).

(ii) Denote the left-hand side by S. Let ψ be any Dirichlet character modulo m. As $\chi \to \chi\overline{\psi}$ is a bijection on X_m, we have

$$\frac{\psi(a)\overline{\psi}(b)}{\varphi(m)} \sum_{\chi \pmod{m}} \chi(a)\overline{\chi}(b) = \frac{1}{\varphi(m)} \sum_{\chi \pmod{m}} \psi\chi(a)\overline{\psi\chi}(b) = \frac{1}{\varphi(m)} \sum_{\chi \pmod{m}} \chi(a)\overline{\chi}(b).$$

Therefore $\psi(a)\overline{\psi}(b)S = S$ so that $S = 0$ unless $\psi(a)\overline{\psi}(b) = \psi(a\overline{b}) = 1$ for all Dirichlet characters ψ modulo m. If this happens, then $a\overline{b} = 1 \pmod{m}$, or equivalently, $a \equiv b \pmod{m}$. Indeed, let $m = p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$ be the prime factorization of m. By the structure theorem for finite abelian groups,

$$(\mathbb{Z}/m\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{r_1}\mathbb{Z})^* \times (\mathbb{Z}/p_2^{r_2}\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_k^{r_k}\mathbb{Z})^*.$$

Now let n_i be a generator for the cyclic group $(\mathbb{Z}/p_i^{r_i}\mathbb{Z})^*$ and let ω_i be a primitive $p_i^{k_i}$-th root of unity for $1 \le i \le k$. Writing $a\overline{b} = n_1^{f_1}n_2^{f_2}\cdots n_k^{f_k}$, consider the Dirichlet character ψ modulo m defined by

$$\psi(n_1^{e_1}n_2^{e_2}\cdots n_k^{e_k}) = \omega_1^{e_1f_1}\omega_2^{e_2f_2}\cdots\omega_r^{e_rf_r}.$$

We have

$$\psi(1) = \omega_1^{f_1}\omega_2^{f_2}\cdots\omega_r^{f_r}.$$

As w_i has order $p_i^{k_i}$ and $0 \le f_i < p_i^{k_i} - 1$ for all i, the only way $\psi(1) = 1$ is if $f_i = 0$ for all i. Therefore $a\overline{b} = 1 \pmod{m}$. In this case $S = 1$. This proves (ii). □

In many practical settings, the orthogonality relations are often used in the following form:

Corollary 1.3.1.

(i) For any Dirichlet character χ modulo m,

$$\frac{1}{\varphi(m)}\sideset{}{'}\sum_{a \pmod{m}} \chi(a) = \delta_{\chi,\chi_{m,0}}.$$

(ii) For any $a \in (\mathbb{Z}/m\mathbb{Z})^*$,

$$\frac{1}{\varphi(m)}\sum_{\chi \pmod{m}} \chi(a) = \delta_{a,1}.$$

Proof. For (i), take $\psi = \chi_{m,0}$ in Proposition 1.3.1 (i). For (ii), take $b \equiv 1 \pmod{m}$ in Proposition 1.3.1 (ii). □

Now that we understand the basics of Dirichlet characters, we might be interested in computing them. This is not hard to do by hand for small m. If the modulus is large this is of course more difficult. However, there is a way to build Dirichlet characters modulo m_2 from those modulo m_1. Let χ_{m_1} be a Dirichlet character modulo m_1. If $m_1 \mid m_2$ then $(a, m_2) = 1$ implies $(a, m_1) = 1$. Therefore we can define a Dirichlet character χ_{m_2} by

$$\chi_{m_2}(a) = \begin{cases} \chi_{m_1}(a) & \text{if } (a, m_2) = 1, \\ 0 & \text{if } (a, m_2) > 1. \end{cases}$$

In this case, we say χ_{m_2} is **induced** from χ_{m_1} or that χ_{m_1} **lifts** to χ_{m_2}. All that is happening is χ_{m_2} is a Dirichlet character modulo m_2 whose values are given by those that χ_{m_1} takes. Clearly every Dirichlet character is induced from itself. On the other hand, provided there is a prime p dividing m_2 and not m_1 (so m_2 is a larger modulus), χ_{m_2} will be different from χ_{m_1}. For instance, $\chi_{m_2}(p) = 0$ but $\chi_{m_1}(p) \neq 0$. In general, we say a Dirichlet character is **primitive** if it is not induced by any character other than itself and **imprimitive** otherwise. Notice that the principal Dirichlet characters are precisely those Dirichlet characters induced from the trivial Dirichlet character, and the only primitive one is the trivial Dirichlet character. In any case, we can determine when Dirichlet characters are induced:

Proposition 1.3.2. *A Dirichlet character χ_{m_2} is induced from a Dirichlet character χ_{m_1} if and only if χ_{m_2} is constant on the residue classes in $(\mathbb{Z}/m_2\mathbb{Z})^*$ that are congruent modulo m_1. When this happens, χ_{m_1} is uniquely determined.*

Proof. For the forward implication, if χ_{m_2} is induced from χ_{m_1}, then χ_{m_2} is constant on the residue classes in $(\mathbb{Z}/m_2\mathbb{Z})^*$ that are congruent modulo m_1 because χ_{m_1} is. For the reverse implication, note that the surjective homomorphism $\mathbb{Z}/m_2\mathbb{Z} \to \mathbb{Z}/m_1\mathbb{Z}$ given by reduction modulo m_1 induces a surjective homomorphism $(\mathbb{Z}/m_2\mathbb{Z})^* \to (\mathbb{Z}/m_1\mathbb{Z})^*$ (because reduction modulo m_1 preserve inverses). Now suppose χ_{m_2} is constant on the residue classes in $(\mathbb{Z}/m_2\mathbb{Z})^*$ that are congruent modulo m_1. Surjectivity of the previously mentioned map implies χ_{m_2} induces a unique group character on $(\mathbb{Z}/m_1\mathbb{Z})^*$ and hence a unique Dirichlet character modulo m_1. By construction χ_{m_2} is induced from χ_{m_1}. $\square$

We are interested in primitive Dirichlet characters because they are the building blocks for all Dirichlet characters:

Theorem 1.3.1. *Every Dirichlet character χ is induced from a primitive Dirichlet character $\widetilde{\chi}$ that is uniquely determined by χ.*

Proof. Let the modulus of χ be m. Define a partial ordering on the set of Dirichlet characters where $\psi \le \chi$ if χ is induced from ψ. This ordering is clearly reflexive, and it is transitive by Proposition 1.3.2. Set

$$X = \left\{ \psi \in \bigcup_{d \mid m} X_d : \psi \le \chi \right\}$$

This set is nonempty and finite by Proposition 1.3.2. Now suppose $\chi_{m_1}, \chi_{m_2} \in X$. Set $m_3 = (m_1, m_2)$. Also from Proposition 1.3.2, χ is constant on the residue classes of $(\mathbb{Z}/m\mathbb{Z})^*$ that are congruent modulo m_1 or m_2 and hence also m_3. Therefore Proposition 1.3.2 implies there is a unique Dirichlet character χ_{m_3} modulo m_3 that lifts to χ_{m_1} and χ_{m_2}. We have now shown that every pair $\chi_{m_1}, \chi_{m_2} \in X$ has a lower bound χ_{m_3}. Hence X contains a primitive Dirichlet character $\widetilde{\chi}$ that is minimal with respect to this partial ordering. There is only one such element. Indeed, since $m_3 \le m_1, m_2$ the partial ordering is compatible with the total ordering by modulus. Thus $\widetilde{\chi}$ is unique. $\square$

In light of Theorem 1.3.1, we define **conductor** q of a Dirichlet character χ modulo m to be the modulus of the unique primitive character $\widetilde{\chi}$ that induces χ. This is the most important data of a Dirichlet character since it tells us how χ is built. Note that χ is primitive if and only if its conductor and modulus are equal. Also observe that if χ has conductor q, then χ is actually q-periodic (necessarily $q \mid m$), and the nonzero values of χ are all q-th roots of unity because those are the nonzero values of $\widetilde{\chi}$. Moreover, $\chi = \widetilde{\chi}\chi_{\frac{m}{q},0}$ by the definition of induced Dirichlet characters. Moreover, primitive characters behave well with respect to multiplication if the conductors are relatively prime as the following proposition shows:

Proposition 1.3.3. *Suppose $(q_1, q_2) = 1$ and where χ_1 and χ_2 are Dirichlet characters modulo q_1 and q_2 respectively. Set $\chi = \chi_1\chi_2$ so that χ is a Dirichlet character modulo q_1q_2. Then χ is a primitive if and only if χ_1 and χ_2 are both primitive.*

Proof. First suppose χ is primitive of conductor q. If d_1 and d_2 are the conductors of χ_1 and χ_2 respectively, then χ is d_1d_2-periodic and primitivity further implies that $q \mid d_1d_2$. But as $d_1 \mid q_1$, $d_2 \mid q_2$, and $q = q_1q_2$, we must have $q = d_1d_2$ and hence $d_1 = q_1$ and $d_2 = q_2$. It follows that χ_1 and χ_2 are both primitive. Conversely, suppose χ_1 and χ_2 are both primitive. If d is the conductor of χ, set $d_1 = (d, q_1)$ and $d_2 = (d, q_2)$. As $(q_1, q_2) = 1$ and $q = q_1q_2$, we must have $(d_1, d_2) = 1$ and $d_1d_2 = q$. But then $d_1 = q_1$ and $d_2 = q_2$. Hence $d = q_1q_2$ which implies that χ is primitive. $\square$

We would now like to distinguish Dirichlet characters whose nonzero values are either real or imaginary. We say χ is **real** if it is real-valued. Hence the nonzero values of χ are 1 or -1 since they must be roots of unity. We say χ is an **complex** if it is not real. More commonly, we distinguish Dirichlet characters modulo m by their order as an element of $(\mathbb{Z}/m\mathbb{Z})^*$. If χ is of order 2, 3, etc. in $(\mathbb{Z}/m\mathbb{Z})^*$ then we say it is **quadratic**, **cubic**, etc. In particular, a Dirichlet character is quadratic if and only if it is real. For any Dirichlet character χ, $\chi(-1) = \pm 1$ because $\chi(-1)^2 = 1$. We would like to distinguish this parity. Accordingly, we say χ is **even** if $\chi(-1) = 1$ and **odd** if $\chi(-1) = -1$. Clearly even Dirichlet characters are even functions and odd Dirichlet characters are odd functions. Moreover, χ and $\overline{\chi}$ have the same parity and any lift of χ has the same parity as χ. Also note that

$$\frac{\chi(1) - \chi(-1)}{2} = \begin{cases} 0 & \text{if } \chi \text{ is even,} \\ 1 & \text{if } \chi \text{ is odd.} \end{cases}$$

Lastly, we would like to discuss quadratic Dirichlet characters. We can construct quadratic Dirichlet characters using Jacobi symbols. If $m \geq 1$ is odd, consider

$$\chi_m(n) = \left(\tfrac{n}{m}\right).$$

Clearly χ_m a quadratic Dirichlet character modulo m because the Jacobi symbol is multiplicative, nonzero if and only if $(n, m) = 1$, and determined modulo m. However, quadratic Dirichlet characters given by Jacobi symbols do not exhaust all possible quadratic Dirichlet characters. For this, we need to use Kronecker symbols. We say that $D \in \mathbb{Z}$ is a **fundamental discriminant** if D is of the form

$$D = \begin{cases} d & \text{if } D \equiv 1 \pmod 4, \\ 4d & \text{if } \frac{D}{4} \equiv 2, 3 \pmod 4, \end{cases}$$

for some square-free $d \in \mathbb{Z}$. Necessarily $d \equiv 1 \pmod 4$ or $d \equiv 2, 3 \pmod 4$ respectively and thus nonzero. We define the **quadratic Dirichlet character** χ_D associated to the fundamental discriminant D by

$$\chi_D(m) = \left(\frac{D}{m}\right).$$

It turns out that χ_D defines a primitive quadratic Dirichlet character, and exhausts all primitive quadratic Dirichlet characters, as the following theorem shows:

Theorem 1.3.2. *If D is a fundamental discriminant and $D \neq 1$, then χ_D is a primitive quadratic Dirichlet character of conductor $|D|$. Moreover, all primitive quadratic Dirichlet characters are of this form.*

Proof. We first show that χ_D is a quadratic Dirichlet character of conductor $|D|$. If $D \equiv 1 \pmod 4$, the sign in quadratic reciprocity is always 1 so that

$$\chi_D(m) = \left(\frac{m}{|D|}\right),$$

and hence is a quadratic Dirichlet character modulo $|D|$ because it is given by the Jacobi symbol. If $\frac{D}{4} \equiv 3 \pmod 4$, the sign in quadratic reciprocity is $\left(\frac{-1}{m}\right)$ which is the primitive quadratic Dirichlet character modulo 4 (there are only two Dirichlet characters modulo 4 since $\varphi(4) = 2$ and clearly $\left(\frac{-1}{m}\right)$ is not principal) so that

$$\chi_D(m) = \left(\frac{-1}{m}\right)\left(\frac{m}{\left|\frac{D}{4}\right|}\right),$$

and hence is a Dirichlet character modulo $|D|$. If $\frac{D}{4} \equiv 2 \pmod{16}$, first observe that $\left(\frac{D}{m}\right) = \left(\frac{8}{m}\right)\left(\frac{\frac{D}{8}}{m}\right)$ where $\left(\frac{8}{m}\right)$ is one of the two primitive quadratic Dirichlet character modulo 8 (the other is $\left(\frac{-8}{m}\right)$ as there are four Dirichlet character modulo 8 because $\varphi(8) = 4$ and the other two are the principal Dirichlet character and the Dirichlet character induced from $\left(\frac{-1}{m}\right)$ as mentioned previously). As $\frac{D}{8} \equiv 1, 3 \pmod 4$, the sign in quadratic reciprocity is either 1 or $\left(\frac{-1}{m}\right)$ according to these two cases. Thus

$$\chi_D(m) = \left(\frac{8}{m}\right)\left(\frac{m}{\left|\frac{D}{8}\right|}\right) \quad \text{or} \quad \chi_D(m) = \left(\frac{-8}{m}\right)\left(\frac{m}{\left|\frac{D}{8}\right|}\right),$$

according to if $\frac{D}{8} \equiv 1, 3 \pmod 4$ respectively, and hence is a quadratic Dirichlet character modulo $|D|$. We can compactly express all of these cases as follows:

$$\chi_D(m) = \begin{cases} \left(\frac{m}{|D|}\right) & \text{if } D \equiv 1 \pmod 4, \\ \left(\frac{-1}{m}\right)\left(\frac{m}{\left|\frac{D}{4}\right|}\right) & \text{if } \frac{D}{4} \equiv 3 \pmod 4, \\ \left(\frac{8}{m}\right)\left(\frac{m}{\left|\frac{D}{8}\right|}\right) & \text{if } \frac{D}{8} \equiv 1 \pmod 4, \\ \left(\frac{-8}{m}\right)\left(\frac{m}{\left|\frac{D}{8}\right|}\right) & \text{if } \frac{D}{8} \equiv 3 \pmod 4. \end{cases}$$

This shows that χ_D is a quadratic Dirichlet characters modulo $|D|$. It easily follows from the above that χ_D is primitive. Indeed, we have already mentioned that the characters $\left(\frac{-1}{m}\right)$, $\left(\frac{8}{m}\right)$, and $\left(\frac{-8}{m}\right)$ are all primitive. Therefore, since D, $\frac{D}{4}$, and $\frac{D}{8}$ are square-free according to their equivalences modulo 4 as given above, and $D \neq 1$, it suffices to show by Proposition 1.3.3 that χ_p is primitive for all primes p with $p \neq 2$. This is immediate since p is prime and clearly χ_p is not principal. We now show that every primitive quadratic Dirichlet character is of the form χ_D for some fundamental discriminant D. By Proposition 1.3.3, it suffices to consider primitive quadratic Dirichlet character modulo $q = p^m$ for some prime p and $m \geq 1$. First suppose that $p \neq 2$. Then $(\mathbb{Z}/q\mathbb{Z})^*$ is cyclic and so every $n \in (\mathbb{Z}/p^m\mathbb{Z})^*$ is of the form $n = v^\nu$ for some $\nu \in (\mathbb{Z}/\varphi(p^m)\mathbb{Z})$ and where v is a generator of $(\mathbb{Z}/p^m\mathbb{Z})^*$. Then every Dirichlet character χ modulo p^m is of the form

$$\chi(n) = e^{\frac{2\pi i k\nu}{\varphi(p^m)}},$$

where $0 \leq k \leq \varphi(q) - 1$. Indeed, this is a unique Dirichlet character for every such k and there are $\varphi(p^m)$ Dirichlet characters modulo p^m as well as the same amount of choices for k. Moreover, χ is primitive if and only if $p \nmid k$ for otherwise χ is a Dirichlet character modulo p^{m-1}. Similarly, χ is quadratic if and only if $\frac{k}{\varphi(p^m)}$ has at most 2 in its denominator which is equivalent to $k \equiv \frac{\varphi(p^m)}{2} \pmod{\varphi(p^m)}$ and hence such a k exists and is unique because $p \neq 2$. We also see that if χ is quadratic, it is imprimitive unless $m = 1$ for then $\varphi(p) = p - 1$ is not a multiple of p. All of this is to say that there is a unique quadratic Dirichlet character modulo q and it is primitive if and only if $q = p$. Necessarily, this unique primitive quadratic Dirichlet character modulo p is given by χ_D for the fundamental discriminant $D = p$ if $p \equiv 1 \pmod 4$ and $D = -p$ if $p \equiv 3 \pmod 4$. Now suppose $p = 2$ so that $q = 2^m$ for some $m \geq 1$. If $m = 1$, $\varphi(2) = 1$ and there are no primitive quadratic Dirichlet characters as the only Dirichlet character is principal. If $m = 2$, $\varphi(4) = 2$ so that there are two Dirichlet characters. They are both quadratic but only one is primitive, namely the principal Dirichlet character as well as the aforementioned primitive quadratic Dirichlet character $\left(\frac{-1}{m}\right)$. For $m \geq 3$, $(\mathbb{Z}/2^m\mathbb{Z})^* \cong C_2 \otimes C_{2^{m-2}}$ where C_2 and $C_{2^{m-2}}$ are the cyclic groups of order 2 and 2^{m-2} respectively. Therefore every $n \in (\mathbb{Z}/2^m\mathbb{Z})^*$ is of the form $n = (-1)^\mu 5^\nu$ for $\mu \in \mathbb{Z}/2\mathbb{Z}$ and $\nu \in \mathbb{Z}/2^{m-2}\mathbb{Z}$ (because the orders of -1 and 5 modulo 2^m are 2 and 2^{m-2} respectively, with the latter case following by induction for $m \geq 3$, and that $\langle -1 \rangle \cap \langle 5 \rangle = \{1\}$). Then every Dirichlet character χ modulo 2^m, for $m \geq 3$, is of the form

$$\chi(n) = e^{\frac{2\pi i j\mu}{2}} e^{\frac{2\pi i k\nu}{2^{m-2}}},$$

where $0 \leq j \leq 1$ and $0 \leq k \leq 2^{m-2} - 1$. Similarly to the case for $p \neq 2$, χ is primitive if and only if $2^{m-2} \nmid k$, or equivalently, k is odd. Moreover, χ is quadratic if and only if $\frac{k}{2^{m-2}}$ has at most 2 in its denominator which is to say that $2^{m-3} \mid k$. Therefore for a primitive quadratic Dirichlet to exist we must have k odd and $2^{m-3} \mid k$ which can happen if and only if $m = 3$. Then $\phi(8) = 4$, so that there are four Dirichlet characters. They are all quadratic but only two are primitive, namely the principal Dirichlet character, the Dirichlet character induced from $\left(\frac{-1}{m}\right)$, and the two aforementioned primitive quadratic Dirichlet characters given by $\left(\frac{8}{m}\right)$ and $\left(\frac{-8}{m}\right)$. These three primitive quadratic Dirichlet characters are given by χ_D for the fundamental discriminants $D = -4$, $D = 8$, and $D = -8$ respectively. We have now shown that all primitive quadratic Dirichlet characters of prime power modulus are given by χ_D for some fundamental discriminant D and thus the same follows for all primitive quadratic Dirichlet characters by Proposition 1.3.3. This completes the proof. $\square$

It follows from Theorem 1.3.2 that all quadratic Dirichlet characters are induced from some χ_D (including $D = 1$ since this corresponds to the trivial Dirichlet character). In particular, so too are the quadratic Dirichlet characters given by Jacobi symbols.

1.4 Special Sums

Analytic number theory comes with its class of special sums that appear naturally. They play the role of discrete counterparts to continuous objects (there is a rich underpinning here). Without a sufficient understanding of these sums, they would cause a discrete obstruction to an analytic problem that we wish to solve.

Ramanujan & Gauss Sums

Let's begin with the Ramanujan sum. For $m \geq 1$ and $b \in \mathbb{Z}$, the **Ramanujan sum** $r(b, m)$ is defined by

$$r(b, m) = \sideset{}{'}\sum_{a \pmod{m}} e^{\frac{2\pi i ab}{m}}.$$

Note that the Ramanujan sum is a finite sum of m-th roots of unity on the unit circle. Clearly $r(0, m) = \varphi(m)$. Ramanujan sums can be computed explicitly by means of the Möbius function (see Appendix A.1):

Proposition 1.4.1. *For any $m \geq 1$ and any nonzero $b \in \mathbb{Z}$,*

$$r(b, m) = \sum_{\ell \mid (b, m)} \ell \mu\left(\frac{m}{\ell}\right).$$

Proof. This is a computation:

$$\begin{aligned}
r(b,m) &= \sideset{}{'}\sum_{a \pmod m} e^{\frac{2\pi iab}{m}} \\
&= \sum_{a \pmod m} e^{\frac{2\pi iab}{m}} \sum_{d \mid (a,m)} \mu(d) && \text{Proposition A.2.1} \\
&= \sum_{d \mid m} \mu(d) \sum_{\substack{a \pmod m \\ d \mid a}} e^{\frac{2\pi iab}{m}} \\
&= \sum_{d \mid m} \mu(d) \sum_{kd \pmod m} e^{\frac{2\pi ikdb}{m}} && a \to kd \\
&= \sum_{d \mid m} \mu(d) \sum_{k \left(\bmod \frac{m}{d}\right)} e^{\frac{2\pi ikb}{\frac{m}{d}}}.
\end{aligned}$$

Now if $\frac{m}{d} \mid b$ the inner sum is $\frac{m}{d}$, and otherwise it is zero because $k \to k\overline{b}$ is a bijection on $\mathbb{Z}/\frac{m}{d}\mathbb{Z}$ and thus we are summing over all $\left(\frac{m}{d}\right)$-th roots of unity. So the double sum above reduces to

$$\sum_{\substack{\frac{m}{d} \mid b \\ d \mid m}} \frac{m}{d}\mu(d) = \sum_{\ell \mid (b,m)} \ell \mu\left(\frac{m}{\ell}\right),$$

upon performing the change of variables $\frac{m}{d} \to \ell$. □

We can also define a Ramanujan sum associated to Dirichlet characters. Let χ be a Dirichlet character modulo m. For any $b \in \mathbb{Z}$, the **Ramanujan sum** $\tau(b,\chi)$ associated to χ is given by

$$\tau(b,\chi) = \sum_{a \pmod m} \chi(a) e^{\frac{2\pi iab}{m}} = \sideset{}{'}\sum_{a \pmod m} \chi(a) e^{\frac{2\pi iab}{m}},$$

where the last equality follows because $\chi(a) = 0$ unless $(a,m) = 1$. If $b = 1$ we will write $\tau(\chi)$ instead. That is, $\tau(\chi) = \tau(1,\chi)$. We call $\tau(\chi)$ the **Gauss sum** associated to *chi*. Observe that if $m = 1$ then χ is the trivial character and $\tau(b,\chi) = 1$. So the interesting cases are when $m \geq 2$. There are some basic properties of these sums which are very useful:

Proposition 1.4.2. *Let χ and ψ be nontrivial Dirichlet characters modulo m and n respectively and let $b \in \mathbb{Z}$. Then the following hold:*

(i) $\overline{\tau(b,\overline{\chi})} = \chi(-1)\tau(b,\chi)$.

(ii) If $(b,m) = 1$, *then* $\tau(b,\chi) = \overline{\chi}(b)\tau(\chi)$.

(iii) If $(b,m) > 1$ *and* χ *is primitive, then* $\tau(b,\chi) = 0$.

(iv) If $(m,n) = 1$, *then* $\tau(b,\chi\psi) = \chi(n)\psi(m)\tau(b,\chi)\tau(b,\psi)$.

(v) Let q be the conductor of χ and let $\widetilde{\chi}$ be the primitive Dirichlet character that lifts to χ. Then

$$\tau(\chi) = \mu\left(\frac{m}{q}\right)\widetilde{\chi}\left(\frac{m}{q}\right)\tau(\widetilde{\chi}).$$

Proof. We will prove the statements separately.

(i) Observe that $a \to -a$ is an isomorphism of $\mathbb{Z}/m\mathbb{Z}$. Thus

$$\begin{aligned}\overline{\tau(b,\overline{\chi})} &= \overline{\sum_{a \pmod m} \overline{\chi}(a)e^{\frac{2\pi iab}{m}}} \\ &= \sum_{a \pmod m} \chi(a)e^{-\frac{2\pi iab}{m}} \\ &= \sum_{a \pmod m} \chi(-a)e^{\frac{2\pi iab}{m}} \\ &= \chi(-1)\sum_{a \pmod m} \chi(a)e^{\frac{2\pi iab}{m}} \\ &= \chi(-1)\tau(b,\chi),\end{aligned}$$

and (i) follows.

(ii) The map $a \to a\bar{b}$ is an isomorphism of $\mathbb{Z}/m\mathbb{Z}$ since $(b,m)=1$. Therefore

$$\tau(b,\chi) = \sum_{a \pmod m} \chi(a)e^{\frac{2\pi iab}{m}} = \sum_{a \pmod m} \chi(a\bar{b})e^{\frac{2\pi ia}{m}} = \overline{\chi}(b)\sum_{a \pmod m} \chi(a)e^{\frac{2\pi ia}{m}} = \overline{\chi}(b)\tau(\chi),$$

and (ii) is proven.

(iii) Now fix a divisor $d < m$ of m and choose an integer c such that $c \equiv 1 \pmod m$. Then necessarily $(c,m)=1$. As $d \mid m$, $c \equiv 1 \pmod d$ and $(c,d)=1$. Moreover, there is such a c with the additional property that $\chi(c) \neq 1$. For if not, χ is induced from $\chi_{d,0}$ which contradicts χ being primitive. Now set $d = \frac{m}{(b,m)} < m$ and choose c as above. Since $(c,m)=1$, $a \to a\overline{c}$ is a bijection on $\mathbb{Z}/m\mathbb{Z}$, so that

$$\chi(c)\tau(b,\chi) = \sum_{a \pmod m} \chi(ac)e^{\frac{2\pi iab}{m}} = \sum_{a \pmod m} \chi(a)e^{\frac{2\pi iab\overline{c}}{m}}.$$

As $e^{\frac{2\pi ib}{m}}$ is a d-th root of unity, and $\overline{c} \equiv 1 \pmod d$ (because c is and $d \mid m$) we have $e^{\frac{2\pi iab\overline{c}}{m}} = e^{\frac{2\pi iab}{m}}$. Thus the last sum above is $\tau(b,\chi)$. So altogether $\chi(c)\tau(b,\chi) = \tau(b,\chi)$. Since $\chi(c) \neq 1$, we conclude $\tau(b,\chi) = 0$ proving (iii).

(iv) Since $(m,n)=1$, the Chinese remainder theorem implies that $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/mn\mathbb{Z})$ via the isomorphism $(a,b) \to an + a'm$ with a taken modulo m and a' taken modulo n. Therefore

$$\begin{aligned}
\tau(b,\chi\psi) &= \sum_{an+a'm \pmod{mn}} \chi\psi(an+a'm)e^{\frac{2\pi i(an+a'm)b}{mn}} \\
&= \sum_{a \pmod{m}} \sum_{a' \pmod{n}} \chi\psi(an+a'm)e^{\frac{2\pi i(an+a'm)b}{mn}} \\
&= \sum_{a \pmod{m}} \sum_{a' \pmod{n}} \chi(an+a'm)\psi(an+a'm)e^{\frac{2\pi i(an+a'm)b}{mn}} \\
&= \sum_{a \pmod{m}} \sum_{a' \pmod{n}} \chi(an)\psi(a'm)e^{\frac{2\pi i(an+a'm)b}{mn}} \\
&= \chi(n)\psi(m) \sum_{a \pmod{m}} \sum_{a' \pmod{n}} \chi(a)\psi(a')e^{\frac{2\pi iab}{m}}e^{\frac{2\pi ia'b}{n}} \\
&= \chi(n)\psi(m) \sum_{a \pmod{m}} \chi(a)e^{\frac{2\pi iab}{m}} \sum_{a' \pmod{n}} \psi(a')e^{\frac{2\pi ia'b}{n}} \\
&= \chi(n)\psi(m)\tau(b,\chi)\tau(b,\psi).
\end{aligned}$$

This proves (iv).

(v) If $\left(\frac{m}{q},q\right) > 1$, then $\widetilde{\chi}\left(\frac{m}{q}\right) = 0$ so we need to show $\tau(\chi) = 0$. As $\left(\frac{m}{q},q\right) > 1$, there exists a prime p such that $p \mid \frac{m}{q}$ and $p \mid q$. By Euclidean division we may write any a modulo m in the form $a = a'\frac{m}{p} + a''$ with a' taken modulo p and a'' taken modulo $\frac{m}{p}$. Then

$$\tau(\chi) = \sum_{a \pmod{m}} \chi(a)e^{\frac{2\pi ia}{m}} = \sum_{\substack{a' \pmod{p} \\ a'' \left(\bmod \frac{m}{p}\right)}} \chi\left(a'\frac{m}{p}+a''\right)e^{\frac{2\pi i\left(a'\frac{m}{p}+a''\right)}{m}}. \tag{1.1}$$

Since $p \mid \left(\frac{m}{q},q\right)$, we have $p^2 \mid m$. Therefore $\left(a'\frac{m}{p}+a'',m\right) = 1$ if and only if $\left(a'\frac{m}{p}+a'',\frac{m}{p}\right) = 1$ and this latter condition is equivalent to $\left(a'',\frac{m}{p}\right) = 1$. Thus the last sum in Equation (1.1) is

$$\sum_{\substack{a' \pmod{p} \\ a'' \left(\bmod \frac{m}{p}\right) \\ \left(a'',\frac{m}{p}\right)=1}} \chi\left(a'\frac{m}{p}+a''\right)e^{\frac{2\pi i\left(a'\frac{m}{p}+a''\right)}{m}}.$$

As $p \mid \frac{m}{q}$, we know $q \mid \frac{m}{p}$ so that $a'\frac{m}{p} + a'' \equiv a'' \pmod{q}$. Then Proposition 1.3.2 implies $\chi\left(a'\frac{m}{p}+a''\right) = \widetilde{\chi}(a'')$ and this sum is further reduced to

$$\sideset{}{'}\sum_{a'' \left(\bmod \frac{m}{p}\right)} \widetilde{\chi}(a'')e^{\frac{2\pi ia''}{m}} \sum_{a' \pmod{p}} e^{\frac{2\pi ia'}{p}}. \tag{1.2}$$

The inner sum in Equation (1.2) vanishes since it is the sum over all p-th roots of unity and thus $\tau(\chi) = 0$. Now suppose $\left(\frac{m}{q},q\right) = 1$. Then (iv) implies

$$\tau(\chi) = \tau(\widetilde{\chi}\chi_{\frac{m}{q},0}) = \widetilde{\chi}\left(\frac{m}{q}\right)\chi_{\frac{m}{q},0}(q)\tau(\widetilde{\chi})\tau(\chi_{\frac{m}{q},0}) = \tau(\chi_{\frac{m}{q},0})\widetilde{\chi}\left(\frac{m}{q}\right)\tau(\widetilde{\chi}).$$

Now observe that $\tau(\chi_{\frac{m}{q},0}) = r\left(1, \frac{m}{q}\right)$. By Proposition 1.4.1 we see that $r\left(1, \frac{m}{q}\right) = \mu\left(\frac{m}{q}\right)$ and (v) follows. $\square$

Notice that Proposition 1.4.2 reduces the evaluation of the Ramanujan sum $\tau(b, \chi)$ to that of the Gauss sum $\tau(\chi)$ at least when χ is primitive. When χ is imprimitive and $(b, m) > 1$ we need to appeal to evaluating $\tau(b, \chi)$ by more direct means. Evaluating $\tau(\chi)$ for general characters χ turns out to be a very difficult problem and is still open. However, it is not difficult to determine the modulus of $\tau(\chi)$ when χ is primitive:

Theorem 1.4.1. *Let χ be a primitive Dirichlet character of conductor q. Then*

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

Proof. If χ is the trivial character this is obvious since $\tau(\chi) = 1$. So we may assume χ is nontrivial. Now this is just a computation:

$$\begin{aligned}
|\tau(\chi)|^2 &= \tau(\chi)\overline{\tau(\chi)} \\
&= \sum_{a \pmod q} \tau(\chi)\overline{\chi}(a)e^{-\frac{2\pi i a}{q}} \\
&= \sum_{a \pmod q} \tau(a, \chi)e^{-\frac{2\pi i a}{q}} && \text{Proposition 1.4.2 (i) and (ii)} \\
&= \sum_{a \pmod q} \left(\sum_{a' \pmod q} \chi(a')e^{\frac{2\pi i a a'}{q}} \right) e^{-\frac{2\pi i a}{q}} \\
&= \sum_{a, a' \pmod q} \chi(a')e^{\frac{2\pi i a(a'-1)}{q}} \\
&= \sum_{a' \pmod q} \chi(a') \left(\sum_{a \pmod q} e^{\frac{2\pi i a(a'-1)}{q}} \right).
\end{aligned}$$

Let $S(a')$ denote the inner sum. For the a' such that $a'-1 \equiv 0 \pmod q$, $S(a') = q$. Otherwise $a \to \overline{a(a'-1)}$ is a bijection on $\mathbb{Z}/q\mathbb{Z}$ ($q \neq 1$ because χ is nontrivial) so that $S(a') = 0$ because it is the sum of all q-th roots of unity. It follows that the double sum is $\chi(1)q = q$. So altogether $|\tau(\chi)|^2 = q$ and hence $|\tau(\chi)| = \sqrt{q}$. $\square$

As an almost immediate corollary to Theorem 1.4.1, we deduce a useful expression for primitive Dirichlet characters of conductor q:

Corollary 1.4.1. *Let χ be a primitive Dirichlet character of conductor q. Then*

$$\tau(n, \chi) = \overline{\chi}(n)\tau(\chi),$$

for all $n \in \mathbb{Z}$. In particular,

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a \pmod q} \overline{\chi}(a)e^{\frac{2\pi i a n}{q}},$$

for all $n \in \mathbb{Z}$.

Proof. If χ is the trivial character this is obvious since $\tau(n,\chi) = 1$. So assume χ is nontrivial. If $(n,q) = 1$, then the first identity is Proposition 1.4.2 (ii). If $(n,q) > 1$, then the first identity follows from Proposition 1.4.2 (iii) and that $\overline{\chi}(n) = 0$. This proves the first identity in full. For the second identity, first note that $\tau(\chi) \neq 0$ by Theorem 1.4.1. Replacing χ with $\overline{\chi}$, dividing the first identity by $\tau(\chi)$, and expanding the Ramanujan sum, gives the second identity. □

In light of Theorem 1.4.1 we define the **epsilon factor** ε_χ for a Dirichlet character χ modulo m by

$$\varepsilon_\chi = \frac{\tau(\chi)}{\sqrt{m}}.$$

Theorem 1.4.1 says that this value lies on the unit circle when χ is primitive and not the trivial character. In any case, the question of the evaluation of Gauss sums further boils down to determining what value the epsilon factor is. This is the real difficultly as the epsilon factor is very hard to calculate and its value is not known for general Dirichlet characters. When χ is primitive, there is a simple relationship between ε_χ and $\varepsilon_{\overline{\chi}}$:

Proposition 1.4.3. *Let χ be a primitive Dirichlet character of conductor q. Then*

$$\varepsilon_\chi \varepsilon_{\overline{\chi}} = \chi(-1).$$

Proof. If χ is trivial this is obvious since $\varepsilon_\chi = \varepsilon_{\overline{\chi}} = 1$. So assume χ is nontrivial. By Proposition 1.4.2 (iii) and that ε_χ lies on the unit circle,

$$\varepsilon_\chi = \frac{\tau(\chi)}{\sqrt{q}} = \chi(-1)\frac{\overline{\tau(\chi)}}{\sqrt{q}} = \chi(-1)\varepsilon_{\overline{\chi}}^{-1},$$

from whence the statement follows. □

Quadratic Gauss Sums

Another important sum is the quadratic Gauss sum. For any $m \geq 1$ and any $b \in \mathbb{Z}$, the **quadratic Gauss sum** $g(b,m)$ is defined by

$$g(b,m) = \sum_{a \pmod m} e^{\frac{2\pi i a^2 b}{m}}.$$

If $b = 1$ we write $g(m)$ instead. That is, $g(m) = g(1,m)$. It turns out that if χ_m is the quadratic Dirichlet character given by the Jacobi symbol, then $\tau(b,\chi_m) = g(b,m)$ provided m is square-free. This will take a little work to prove. We first reduce to the case when $(b,m) = 1$:

Proposition 1.4.4. *Let $m \geq 1$ be odd and let $b \in \mathbb{Z}$. Then*

$$g(b,m) = (b,m)g\left(\frac{b}{(b,m)}, \frac{m}{(b,m)}\right).$$

Proof. By Euclidean division write any a modulo m in the form $a = a'\frac{m}{(b,m)} + a''$ with a' take modulo (b,m) and a'' take modulo $\frac{m}{(b,m)}$. Then

$$\begin{aligned}
g(b,m) &= \sum_{a \pmod m} e^{\frac{2\pi i a^2 b}{m}} \\
&= \sum_{\substack{a' \pmod{(b,m)} \\ a'' \left(\bmod \frac{m}{(b,m)}\right)}} e^{\frac{2\pi i \left(a'\frac{m}{(b,m)}+a''\right)^2 b}{m}} \\
&= \sum_{a'' \left(\bmod \frac{m}{(b,m)}\right)} e^{\frac{2\pi i (a'')^2 b}{m}} \sum_{a' \pmod{(b,m)}} e^{\frac{2\pi i\left(2a''a'\frac{m}{(b,m)}+\left(a'\frac{m}{(b,m)}\right)^2\right)b}{m}} \\
&= \sum_{a'' \left(\bmod \frac{m}{(b,m)}\right)} e^{\frac{2\pi i (a'')^2 \frac{b}{(b,m)}}{\frac{m}{(b,m)}}} \sum_{a' \pmod{(b,m)}} e^{\frac{2\pi i\left(2a''a'\frac{m}{(b,m)}+\left(a'\frac{m}{(b,m)}\right)^2\right)\frac{b}{(b,m)}}{\frac{m}{(b,m)}}} \\
&= (b,m) \sum_{a'' \left(\bmod \frac{m}{(b,m)}\right)} e^{\frac{2\pi i (a'')^2 \frac{b}{(b,m)}}{\frac{m}{(b,m)}}},
\end{aligned}$$

where the last line follows because $\left(2a''a'\frac{m}{(b,m)} + \left(a'\frac{m}{(b,m)}\right)^2\right) \equiv 0 \left(\bmod \frac{m}{(b,m)}\right)$ and thus the second sum is (b,m). The remaining sum is $g\left(\frac{b}{(b,m)}, \frac{m}{(b,m)}\right)$ which finishes the proof. $\square$

As a consequence of Proposition 1.4.4, we may always assume $(b,m) = 1$. Now we give an equivalent formulation of the Ramanujan sum associated to quadratic Dirichlet characters given by Jacobi symbols and show that in the case $m = p$ an odd prime, the Ramanujan and quadratic Gauss sums agree:

Proposition 1.4.5. *Let $m \geq 1$ and $b \in \mathbb{Z}$ be such that $(b,m) = 1$. Also let χ_m be the quadratic Dirichlet character given by the Jacobi symbol. Then*

$$\tau(b,\chi_m) = \sum_{a \pmod m} \left(1 + \left(\frac{a}{m}\right)\right) e^{\frac{2\pi i a b}{m}}.$$

Moreover, when $m = p$ is prime,

$$\tau(b,\chi_p) = g(b,p).$$

Proof. If $m = 1$ the claim is obvious since $\tau(b,\chi_1) = 1$ so assume $m > 1$. To prove the first statement, observe

$$\sum_{a \pmod m} \left(1 + \left(\frac{a}{m}\right)\right) e^{\frac{2\pi i a b}{m}} = \sum_{a \pmod m} e^{\frac{2\pi i a b}{m}} + \sum_{a \pmod m} \left(\frac{a}{m}\right) e^{\frac{2\pi i a b}{m}}.$$

The first sum on the right-hand side is zero as it is the sum over all m-th roots of unity since $(b,m) = 1$. This proves the first claim. Now let $m = p$ be an odd prime. From the definition of the Jacobi symbol we see that $1 + \left(\frac{a}{p}\right) = 2, 0$ depending on if a is a quadratic residue modulo p or not provided $a \not\equiv 0 \pmod p$. If $a \equiv 0 \pmod p$, then $1 + \left(\frac{a}{p}\right) = 1$. Moreover, if a is a quadratic residue modulo p, then $a \equiv (a')^2 \pmod p$

for some a'. So one the one hand,

$$\tau(b,\chi_p) = \sum_{a \pmod p} \left(1+\left(\frac{a}{p}\right)\right) e^{\frac{2\pi iab}{p}} = 1 + 2 \sum_{\substack{a \pmod p \\ a\equiv (a')^2 \pmod p \\ a\not\equiv 0 \pmod p}} e^{\frac{2\pi i(a')^2 b}{p}}.$$

On the other hand,

$$g(b,p) = 1 + \sum_{\substack{a \pmod p \\ a\not\equiv 0 \pmod p}} e^{\frac{2\pi i a^2 b}{p}},$$

but this last sum counts every quadratic residue twice because $(-a)^2 = a^2$. Hence the two previous sums are equal. □

We would like to generalize the second statement in Proposition 1.4.5 to when m is square-free. In this direction, a series of reduction properties will be helpful:

Proposition 1.4.6. *Let $m,n \geq 1$, p be an odd prime, and $b \in \mathbb{Z}$. Then the following hold:*

(i) If $(b,p)=1$, then $g(b,p^r) = pg(b,p^{r-2})$ for all $r \in \mathbb{Z}$ with $r \geq 2$.

(ii) If $(m,n)=1$ and $(b,mn)=1$, then $g(b,mn) = g(bn,m)g(bm,n)$.

(iii) If m is odd and $(b,m)=1$, then $g(b,m) = \left(\frac{b}{m}\right)g(m)$ where $\left(\frac{b}{m}\right)$ is the Jacobi symbol.

Proof. We will prove the statements separately.

(i) First notice that

$$g(b,p^r) = \sum_{a \pmod{p^r}} e^{\frac{2\pi i a^2 b}{p^r}} = \sideset{}{'}\sum_{a \pmod{p^r}} e^{\frac{2\pi i a^2 b}{p^r}} + \sum_{a \pmod{p^{r-1}}} e^{\frac{2\pi i a^2 b}{p^{r-2}}},$$

since every a modulo p satisfies $(a,p)=1$ or not. By Euclidean division every element a modulo p^{r-1} is of the form $a = a'p^{r-2} + a''$ with a' taken modulo p and a'' taken modulo p^{r-2}. Since $(a'p^{r-2}+a'') \equiv a'' \pmod{p^{r-2}}$, every a'' is counted p times modulo p^{r-2}. Along with the fact that $(a'p^{r-2}+a'')^2 \equiv (a'')^2 \pmod{p^{r-2}}$, these facts give the middle equality in the following chain:

$$\sum_{a \pmod{p^{r-1}}} e^{\frac{2\pi i a^2 b}{p^{r-2}}} = \sum_{\substack{a' \pmod p \\ a'' \left(\bmod p^{r-2}\right)}} e^{\frac{2\pi i \left(a'p^{r-2}+a''\right)^2 b}{p^{r-2}}} = p \sum_{a'' \pmod p} e^{\frac{2\pi i (a'')^2 b}{p^{r-2}}} = pg(b,p^{r-2}).$$

It remains to show

$$\sideset{}{'}\sum_{a \pmod{p^r}} e^{\frac{2\pi i a^2 b}{p^r}},$$

is zero. This sum is exactly $r(b,p^r)$ so by Proposition 1.4.1, and that $(b,p)=1$, we conclude

$$\sideset{}{'}\sum_{a \pmod{p^r}} e^{\frac{2\pi i a^2 b}{p^r}} = \mu(p^r) = 0,$$

because $r \geq 2$. This proves (i).

(ii) Observe

$$g(bn,m)g(bm,n) = \left(\sum_{a \pmod m} e^{\frac{2\pi i a^2 bn}{m}}\right)\left(\sum_{a' \pmod n} e^{\frac{2\pi i (a')^2 bm}{n}}\right) = \sum_{\substack{a \pmod m \\ a' \pmod n}} e^{\frac{2\pi i\left((an)^2+(a'm)^2\right)b}{mn}}.$$

Note that $e^{\frac{2\pi i\left((an)^2+(a'm)^2\right)b}{mn}}$ only depends upon $(an)^2+(a'm)^2$ modulo mn. Clearly $(an+a'm)^2 \equiv (an)^2+(a'm)^2 \pmod{mn}$, so set $a'' = an+a'm$ taken modulo mn. Since $(m,n)=1$, the Chinese remainder theorem implies that $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/mn\mathbb{Z})$ via the isomorphism $(a,a') \to an+a'm$. Thus the last sum above is equal to

$$\sum_{a'' \pmod{mn}} e^{\frac{2\pi i (a'')^2 b}{mn}},$$

which is precisely $g(b,mn)$. So (ii) is proven.

(iii) The claim is obvious if $m=1$ because $g(b,1)=1$ so assume $m>1$. If $m=p$, then Proposition 1.4.5, Proposition 1.4.2 (ii), and that quadratic Dirichlet characters are their own conjugate altogether imply the claim. Now let $r \geq 1$ and assume by induction that the claim holds when $m = p^{r'}$ for all positive integers r' such that $r' < r$. Then by (i), we have

$$g(b,p^r) = pg(b,p^{r-2}) = \left(\frac{b}{p^{r-2}}\right)pg(p^{r-2}) = \left(\frac{b}{p^{r-2}}\right)g(p^r) = \left(\frac{b}{p^r}\right)g(p^r). \tag{1.3}$$

It now suffices to prove the claim when $m = p^r q^s$ where q is another odd prime and $s \geq 1$. Then by (ii) and Equation (1.3), we compute

$$\begin{aligned}
g(b,p^rq^s) &= g(bq^s,p^r)g(bp^r,q^s) \\
&= \left(\frac{bq^s}{p^r}\right)\left(\frac{bp^r}{q^s}\right)g(p^r)g(q^s) \\
&= \left(\frac{b}{p^r}\right)\left(\frac{q^s}{p^r}\right)\left(\frac{b}{q^s}\right)\left(\frac{p^r}{q^s}\right)g(p^r)g(q^s) \\
&= \left(\frac{b}{p^rq^s}\right)\left(\frac{q^s}{p^r}\right)\left(\frac{p^r}{q^s}\right)g(p^r)g(q^s) \\
&= \left(\frac{b}{p^rq^s}\right)g(q^s,p^r)g(p^r,q^s) \\
&= \left(\frac{b}{p^rq^s}\right)g(p^rq^s).
\end{aligned}$$

This proves (iii). □

At last we can prove that our Ramanujan and quadratic Gauss sums agree for square-free m:

Theorem 1.4.2. *Suppose $m \geq 1$ be square-free and odd and let χ_m be the quadratic Dirichlet character given by the Jacobi symbol. Let $b \in \mathbb{Z}$ such that $(b,m)=1$. Then*

$$\tau(b,\chi_m) = g(b,m).$$

Proof. The claim is obvious if $m = 1$ because $\tau(b, \chi_1) = 1$ and $g(b, 1) = 1$ so assume $m > 1$. Since χ_m is quadratic, it suffices to prove the claim when $b = 1$ by Proposition 1.4.2 (ii) and Proposition 1.4.6 (iii). Now let $m = p_1 p_2 \cdots p_k$ be the prime decomposition of m. Repeated application of Proposition 1.4.2 (iv) gives the first equality in the chain

$$\begin{aligned}\tau(\chi) &= \prod_{1\le i<j\le k} \chi_{p_i}(p_j)\chi_{p_j}(p_i)\tau(\chi_{p_i})\tau(\chi_{p_j})\\ &= \prod_{1\le i<j\le k} \chi_{p_i}(p_j)\chi_{p_j}(p_i)g(p_i)g(p_j)\\ &= \prod_{1\le i<j\le k} g(p_j, p_i)g(p_i, p_j)\\ &= g(q).\end{aligned}$$

This proves the claim. □

Now let's turn to Proposition 1.4.6 and the evaluation of the quadratic Gauss sum. Proposition 1.4.6 (ii) and (iii) reduce the evaluation of $g(b, m)$ for odd m and $(b, m) = 1$ to computing $g(p)$ for p an odd prime. As with the Gauss sum, it is not difficult to compute the modulus of the quadratic Gauss sum:

Theorem 1.4.3. *Let $m \ge 1$ be odd. Then*

$$|g(m)| = \sqrt{m}.$$

Proof. By Proposition 1.4.6 (ii), it suffices to prove this when $m = p^r$ is a power of an odd prime. By Euclidean division write $r = 2n + r'$ for some positive integer n and with $r' = 0, 1$ depending on if r is even or odd respectively. Then Proposition 1.4.6 (i) implies

$$|g(p^r)|^2 = p^{2n}|g(p^{r'})|^2.$$

If $r' = 0$, then $2n = r$ so that $p^{2n} = p^r$. Thus $|g(p^r)| = \sqrt{p^r}$. If $r' = 1$, then Theorem 1.4.1 and Proposition 1.4.5 together imply $|g(p^{r'})|^2 = p$ so that the right-hand side above is $p^{2n+1} = p^r$ and again we have $|g(p^r)| = \sqrt{p^r}$. □

Accordingly, we define the **epsilon factor** ε_m for any $m \ge 1$ by

$$\varepsilon_m = \frac{g(m)}{\sqrt{m}}.$$

Theorem 1.4.3 says that this value lies on the unit circle when m is odd. Thus the question of the evaluation of quadratic Gauss sums reduces to determining what the epsilon factor is. This was completely resolved and the original proof is due to Gauss in 1808 (see [Gau08]). He actually treated the case m is even as well. We have avoided discussing this because we will not need it in the following and many of the previous proofs need to be augmented when m is even (see [Lan94] for a treatment of the even case). As for the evaluation, one of the cleanest proofs uses analytic techniques (see [Lan94]) and the precise statement is the following:

Theorem 1.4.4. *Let $m \ge 1$. Then*

$$\varepsilon_m = \begin{cases} (1+i) & \text{if } m \equiv 0 \pmod 4,\\ 1 & \text{if } m \equiv 1 \pmod 4,\\ 0 & \text{if } m \equiv 2 \pmod 4,\\ i & \text{if } m \equiv 3 \pmod 4.\end{cases}$$

As an immediate corollary, this implies the evaluation of the epsilon factor ε_{χ_p} where χ_p is the quadratic Dirichlet character given by the Jacobi symbol for an odd prime p:

Corollary 1.4.2. *Let p be an odd prime and χ_p be the quadratic Dirichlet character given by the Jacobi symbol. Then*

$$\varepsilon_{\chi_p} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4, \\ i & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

Proof. The statement follows immediately from Theorem 1.4.4 and Proposition 1.4.5. □

Kloosterman & Salié Sums

Our last class of sums generalize both types of Ramanujan sums. For any $c \geq 1$ and $n, m \in \mathbb{Z}$, the **Kloosterman sum** $K(n, m, c)$ is defined by

$$K(n,m,c) = \sum_{\substack{a \pmod c \\ (a,c)=1}} e^{\frac{2\pi i(an+\bar{a}m)}{c}} = \sideset{}{'}\sum_{a \pmod c} e^{\frac{2\pi i(an+\bar{a}m)}{c}}.$$

Notice that if either $n = 0$ or $m = 0$ then the Kloosterman sum reduces to a Ramanujan sum. Kloosterman sums have similar properties to those of Ramanujan sums, but we will not need them. The only result we will need is a famous bound, often called the **Weil bound** for Kloosterman sums, proved by Weil (see [Wei48] for a proof):

Theorem 1.4.5 (Weil bound). *Let $c \geq 1$ and $n, m \in \mathbb{Z}$. Then*

$$|K(n,m,c)| \leq \sigma_0(c)\sqrt{(n,m,c)}\sqrt{c}.$$

Lastly, Salié sums are Kloosterman sums with Dirichlet characters. To be precise, for any $c \geq 1$, $n, m \in \mathbb{Z}$, and a Dirichlet character χ with conductor $q \mid c$, the **Salié sum** $S_\chi(n, m, c)$ is defined by

$$S_\chi(n,m,c) = \sum_{\substack{a \pmod c \\ (a,c)=1}} \chi(a) e^{\frac{2\pi i(an+\bar{a}m)}{c}} = \sideset{}{'}\sum_{a \pmod c} \chi(a) e^{\frac{2\pi i(an+\bar{a}m)}{c}}.$$

If either $n = 0$ or $m = 0$ then the Salié sum reduces to the Ramanujan sum associated to χ.

1.5 Lattices

Lattices are ubiquitous in analytic number theory because they provide a way to study discrete points by geometric methods. Let V be an n-dimensional inner product space over a characteristic zero field F with nondegenerate symmetric inner product $\langle \cdot, \cdot \rangle$. We say that a subset Λ of V is a **lattice** if Λ is a free abelian group of rank n. In particular, any lattice Λ is of the form

$$\Lambda = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n,$$

for some basis $\{v_1, \ldots, v_n\}$ of V. The **covolume** V_Λ of Λ is defined to be

$$V_\Lambda = |V/\Lambda|,$$

which is finite by Proposition C.1.1. Moreover, if $\{e_1, \ldots, e_n\}$ is an orthonormal basis for V, write

$$v_i = \sum_{1 \le i \le n} v_{i,j} e_j,$$

with $v_{i,j} \in F$ for $1 \le i, j \le n$ and define the associated **generator matrix** P by

$$P = \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{pmatrix}.$$

Then P is the base change matrix from $\{e_1, \ldots, e_n\}$ to $\{v_1, \ldots, v_n\}$ and so Proposition C.1.1 shows that

$$V_\Lambda = |\det(P)|,$$

We can now introduce the notion of dual lattices. If Λ is a lattice in V, then the **dual lattice** $\Lambda^\vee$ of Λ is defined by

$$\Lambda^\vee = \{v \in V : \langle \lambda, v \rangle \in \mathbb{Z} \text{ for all } \lambda \in \Lambda\}.$$

In other words, $\Lambda^\vee$ consists of all of the vectors in V whose inner product with elements of the lattice Λ are integers. The dual lattice is indeed a lattice as the following proposition shows:

Proposition 1.5.1. *If $\langle v_1, \ldots, v_n \rangle$ is a basis for Λ, then the dual basis $\langle v_1^\vee, \ldots, v_n^\vee \rangle$ is a basis for $\Lambda^\vee$. In particular, $\Lambda^\vee$ is a lattice.*

Proof. Let $v \in V$ and write

$$v = \sum_{1 \le j \le n} a_i v_j^\vee,$$

with $v_j^\vee \in \mathbb{R}$ for all j. Since $\langle v_i, v_j^\vee \rangle = \delta_{i,j}$ for $1 \le i, j \le n$, it follows that $v \in \Lambda^\vee$ if and only if $v_j^\vee \in \mathbb{Z}$ for all j. This means that $\langle v_1^\vee, \ldots, v_n^\vee \rangle$ is a free abelian group of rank n. Hence $\Lambda^\vee$ is a lattice. □

Since the dual of the dual basis of a vector space is the original basis, $(\Lambda^\vee)^\vee = \Lambda$. We say that Λ is **self-dual** if $\Lambda^\vee = \Lambda$. It also turns out that the covolume of the dual lattice is the inverse of the volume of the original lattice:

Proposition 1.5.2. *Let Λ be a lattice in V and let $\Lambda^\vee$ be its dual. Then*

$$V_{\Lambda^\vee} = \frac{1}{V_\Lambda}.$$

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V and let $\{v_1, \ldots, v_n\}$ be a basis for Λ. By Proposition 1.5.1, $\{v_1^\vee, \ldots, v_n^\vee\}$ is a basis for $\Lambda^\vee$. If P is the associated generator matrix for Λ, then P is the base change matrix from $\{e_1, \ldots, e_n\}$ to $\{v_1, \ldots, v_n\}$. Hence $(P^{-1})^t$ is the base change matrix from $\{e_1, \ldots, e_n\}$ to $\{v_1^\vee, \ldots, v_n^\vee\}$. The claim follows by Proposition C.1.1. □

We now turn to the case when V is an n-dimensional vector space over $\mathbb{R}$ with standard inner product $\langle \cdot, \cdot \rangle$. We let $d\lambda$ be the associated Lebesgue measure on V. If $\{v_1, \ldots, v_n\}$ is a basis of V so that any $v \in V$ has the form

$$v = t_1 v_1 + \cdots + tn v_n,$$

for some $t_i \in \mathbb{R}$ for $1 \leq i \leq n$, the Lebesgue measure $d\lambda$ is $dt_1 \cdots dt_n$. Then

$$\mathrm{Vol}(X) = \int_X d\lambda = \int_X dt_1 \cdots dt_n,$$

is the volume of any measurable subset $X \subseteq V$. Note that Λ acts on V by automorphisms given by translation. That is, we have a group action

$$\Lambda \times V \to V \qquad (\lambda, v) \to \lambda + v.$$

Moreover, Λ acts properly discontinuously on V (see Appendix D.1). To see this, let $v \in V$ and let δ_v be such that $0 < \delta_v < \min_{1 \leq i \leq n}(v - v_i)$. Then taking U_v to be the ball of radius δ_v about v, the intersection $\lambda + U_v \cap U_v$ is empty unless $\lambda = 0$. As Λ is also discrete, it follows by Proposition D.1.1 that V/Λ is also connected Hausdorff (recall that V is connected Hausdorff). In particular, V/Λ admits a fundamental domain

$$\mathcal{M} = \{t_1 v_1 + \cdots + t_n v_n \in V : 0 \leq t_i \leq 1 \text{ for } 1 \leq i \leq n\}.$$

Indeed, since Λ acts by translations, it is obvious that $\mathcal{M}$ is a fundamental domain for V/Λ. Moreover, any translation of $\mathcal{M}$ by an element of Λ is also a fundamental domain. As we might expect, the covolume of Λ is equal to the volume of $\mathcal{M}$:

Proposition 1.5.3. *Let Λ be a lattice in V and let $\mathcal{M}$ be a fundamental domain. Then*

$$V_\Lambda = \mathrm{Vol}(\mathcal{M}).$$

Proof. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V and let $\{v_1, \ldots, v_n\}$ be a basis for Λ. Making the change of variables $x_i \to t_i$ for $1 \leq i \leq n$ where

$$x_1 e_1 + \cdots + x_n e_n \to t_1 v_1 + \cdots + t_n v_n,$$

the corresponding Jacobian matrix is the generator matrix P for the base change from $\{e_1, \ldots, e_n\}$ to $\{v_1, \ldots, v_n\}$. Then

$$\mathrm{Vol}(\mathcal{M}) = \int_{\mathcal{M}} dx_1 \cdots dx_n = |\det(P)| \int_{[0,1]^n} dt_1 \cdots dt_n = |\det(P)| = V_\Lambda,$$

where the last equality follows by Proposition C.1.1. $\square$

The most important result we will require about lattices is **Minkowski's lattice point theorem** which states that, under some mild conditions, a set of sufficiently large volume in $\mathbb{R}^n$ contains a nonzero point of a lattice:

Theorem 1.5.1 (Minkowski's lattice point theorem). *Let Λ be a lattice in $\mathbb{R}^n$. Suppose $X \subset \mathbb{R}^n$ is a compact convex symmetric set. If*

$$\mathrm{Vol}(X) \geq 2^n V_\Lambda,$$

then there exists a nonzero $\lambda \in L$ with $\lambda \in X$

Proof. We will prove the claim depending on if the inequality is strict or not. First suppose $\mathrm{Vol}(X) > 2^n V_\Lambda$. Consider the linear map

$$\phi : \frac{1}{2}X \to \mathbb{R}^n/\Lambda \qquad \frac{1}{2}x \mapsto \frac{1}{2}x \pmod{\Lambda}.$$

If ϕ were injective, then

$$\mathrm{Vol}\left(\frac{1}{2}X\right) = \frac{1}{2^n}\mathrm{Vol}(X) \leq V_\Lambda,$$

so that $\mathrm{Vol}(X) \leq 2^n V_\Lambda$. This is a contradiction, so ϕ cannot be injective. Hence there exists distinct $x_1, x_2 \in \frac{1}{2}X$ such that $\phi(x_1) = \phi(x_2)$. Thus $2x_1, 2x_2 \in X$. In particular, since X is symmetric we must have $-2x_2 \in X$. But then the fact that X is convex implies that

$$\left(1 - \frac{1}{2}\right)2x_1 + \frac{1}{2}(-2x_2) = x_1 - x_2 \in X.$$

Note that $x_1 - x_2 \in \Lambda$ because $\phi(x_1) = \phi(x_2)$ and ϕ is linear. Then $\lambda = x_1 - x_2$ is a nonzero element of Λ with $\lambda \in X$. Now suppose $\mathrm{Vol}(X) = 2^n V_\Lambda$. Then for any $\varepsilon > 0$, we have

$$\mathrm{Vol}((1+\varepsilon)X) = (1+\varepsilon)^n \mathrm{Vol}(X) = (1+\varepsilon)^n 2^n V_\Lambda > 2^n V_\Lambda.$$

So what we have just proved shows that there exists a nonzero $\lambda_\varepsilon \in \Lambda$ with $\lambda_\varepsilon \in (1+\varepsilon)X$. In particular, if $\varepsilon \leq 1$ then $\lambda_\varepsilon \in 2X \cap \Lambda$. The set $2X \cap \Lambda$ is compact and discrete, because X is compact and Λ is discrete, and therefore is finite. But as this holds for all $\varepsilon \leq 1$, the sequence $(\lambda_{\frac{1}{n}})_{n \geq 1}$ belongs to the finite set $2X \cap \Lambda$ and so must converge to a point λ. Since Λ is discrete and the $\lambda_{\frac{1}{n}}$ are nonzero so too is λ. As

$$\lambda \in \bigcap_{n \geq 1}\left(1 + \frac{1}{n}\right)X,$$

and X is closed, $\lambda \in X$ as well. Thus we have found a nonzero $\lambda \in L$ with $\lambda \in X$ and we are done. □

1.6 Integration Techniques & Transforms

Integration Techniques

Complex integrals are a core backbone of many analytic number theory techniques. An extremely useful one is called **shifting the line of integration**:

Theorem 1.6.1 (Shifting the line of integration). *Suppose we are given an integral*

$$\int_{\mathrm{Re}(z)=a} f(z)\,dz \quad \text{or} \quad \int_{\mathrm{Im}(z)=a} f(z)\,dz,$$

and some real b with $b < a$ in the first case and $b > a$ in the second case. Suppose $f(z)$ is meromorphic on a strip bounded by the lines $\mathrm{Re}(z) = a, b$ or $\mathrm{Im}(z) = a, b$ and is holomorphic about the lines $\mathrm{Re}(z) = a, b$ or $\mathrm{Im}(z) = a, b$ respectively. Moreover, suppose $f(z) \to 0$ as $y \to \infty$ or $x \to \infty$ respectively. Then

$$\int_{(a)} f(z)\,dz = \int_{(b)} f(z)\,dz + 2\pi i \sum_{\rho \in P} \operatorname*{Res}_{z=\rho} f(z),$$

where P is the set of poles inside of the strip bounded by the lines $\mathrm{Re}(z) = a, b$ or $\mathrm{Im}(z) = a, b$ respectively.

Proof. To collect these cases, let (a) stand for the line $\mathrm{Re}(z) = a$ or $\mathrm{Im}(z) = a$ respectively with positive orientation. Let R_T be a rectangle, of height or width T respectively, given positive orientation, and with its edges on (a) and (b) respectively. Consider the limit

$$\lim_{T \to \infty} \int_{R_T} f(z)\,dz.$$

On the one hand, the residue theorem implies the integral is a sum of a $2\pi i$ multiple of the residues r_i in the rectangle R_T and hence the limit is a $2\pi i$ multiple of the sum of the residues in the strip bounded by (a) and (b). On the other hand, the integral can be decomposed into a sum of four integrals along the edges of R_T and by taking the limit, the edges other than (a) and (b) will tend to zero because $f(z) \to 0$ as $y \to \infty$ or $x \to \infty$ respectively. What remains in the limit is the difference between the integral along (a) and (b). So in total,

$$\int_{(a)} f(z)\,dz = \int_{(b)} f(z)\,dz + 2\pi i \sum_{\rho \in P} \operatorname*{Res}_{z=\rho} f(z).$$

□

A particular application of interest is when the integral in question is real and over the entire real line, the integrand is entire as a complex function, and one is trying to shift the line of integration of the complexified integral to $\mathrm{Im}(z) = a$. In this case, shifting the line of integration amounts to making the change of variables $x \to x - ia$ without affecting the initial line of integration. The second integral technique we will use is when we are summing integrals over a group and is called the **unfolding/folding method**:

Theorem 1.6.2 (Unfolding/folding method). *Suppose $f(z)$ is holomorphic on some region Ω. Moreover, suppose G is a countable group acting by automorphisms on Ω and let D and F be regions such that*

$$D = \bigcup_{g \in G} gF,$$

where the intersections $gF \cap hF$ are measure zero for all $g, h \in G$ with respect to some G-invariant measure $d\mu$. Then

$$\int_F \sum_{g \in G} f(gz)\,d\mu = \int_D f(z)\,d\mu,$$

provided either side is absolutely convergent.

Proof. First suppose $\int_F \sum_{g \in G} f(gz)\,d\mu$ converges absolutely. By Fubini's theorem, we may interchange the sum and integral. Upon making the change of variables $z \to g^{-1}z$, the invariance of $d\mu$ implies that the integral takes the form

$$\sum_{g \in G} \int_{gF} f(z)\,d\mu.$$

As G is countable and the intersections $gF \cap hF$ are measure zero, the overlap in $\bigcup_{g \in G} gF$ is also measure zero. As $D = \bigcup_{g \in G} gF$, the result follows. Of course, there is equality everywhere so we can also run the procedure in reverse provided $\int_D f(z)\,d\mu$ converges absolutely. □

In the unfolding/folding method, we refer to the going from the left-hand side to right-hand as **unfolding** and going from the right-hand to left-hand side as **folding**.

The Fourier Transform

The first type of integral transform we will need is the Fourier transform. Suppose $f(\mathbf{x})$ is absolutely integrable on $\mathbb{R}^n$. The **Fourier transform** $(\mathcal{F}f)(\mathbf{t})$ of $f(\mathbf{x})$ is defined by

$$(\mathcal{F}f)(\mathbf{t}) = \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\,d\mathbf{x}.$$

This integral is absolutely convergent precisely because $f(\mathbf{x})$ is absolutely integrable on $\mathbb{R}^n$. The Fourier transform is intimately related to periodic functions. If $f(\mathbf{x})$ is 1-periodic in each component and integrable on $[0,1]^n$, then we define the $\mathbf{n}$-th **Fourier coefficient** $\hat{f}(\mathbf{n})$ of $f(\mathbf{x})$ to be

$$\hat{f}(\mathbf{n}) = \int_{[0,1]^n} f(\mathbf{x}) e^{-2\pi i \langle \mathbf{n}, \mathbf{x} \rangle}\, d\mathbf{x}.$$

The **Fourier series** of $f(\mathbf{x})$ is defined by the series

$$\sum_{\mathbf{n} \in \mathbb{Z}^n} \hat{f}(\mathbf{n}) e^{2\pi i \langle \mathbf{n}, \mathbf{x} \rangle}.$$

There is the question of whether the Fourier series of $f(\mathbf{x})$ converges at all and if so does it even converge to $f(\mathbf{x})$ itself. Under reasonable conditions this is possible as the following proposition shows (see [G$^+$08] for a proof):

Proposition 1.6.1. *If $f(\mathbf{x})$ is smooth and 1-periodic in each component then it converges uniformly to its Fourier series.*

The link between the Fourier transform and Fourier series is given by the **Poisson summation formula**:

Theorem 1.6.3 (Poisson summation formula). *Suppose $f(\mathbf{x})$ is absolutely integrable on $\mathbb{R}^n$ and the function*

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

is locally absolutely uniformly convergent and smooth. Then

$$\sum_{\mathbf{n} \in \mathbb{Z}^n} f(\mathbf{x} + \mathbf{n}) = \sum_{\mathbf{t} \in \mathbb{Z}^n} (\mathcal{F} f)(\mathbf{t}) e^{2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}.$$

In particular,

$$\sum_{\mathbf{n} \in \mathbb{Z}^n} f(\mathbf{n}) = \sum_{\mathbf{t} \in \mathbb{Z}^n} (\mathcal{F} f)(\mathbf{t}).$$

Proof. By assumption $F(\mathbf{x})$ is smooth and it is clearly 1-periodic in each component. Therefore it admits a Fourier series. We compute the $\mathbf{t}$-th Fourier coefficient of $F(\mathbf{x})$ as follows:

$$\begin{aligned}
\hat{F}(\mathbf{t}) &= \int_{[0,1]^n} F(\mathbf{x}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\, d\mathbf{x} \\
&= \int_{[0,1]^n} \sum_{\mathbf{n} \in \mathbb{Z}^n} f(\mathbf{x} + \mathbf{n}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\, d\mathbf{x} \\
&= \sum_{\mathbf{n} \in \mathbb{Z}^n} \int_{[0,1]^n} f(\mathbf{x} + \mathbf{n}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\, d\mathbf{x} && \text{FT} \\
&= \sum_{\mathbf{n} \in \mathbb{Z}^n} \int_{\mathbf{n} + [0,1]^n} f(\mathbf{x}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\, d\mathbf{x} \\
&= \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-2\pi i \langle \mathbf{t}, \mathbf{x} \rangle}\, d\mathbf{x} && \text{FT} \\
&= (\mathcal{F} f)(\mathbf{t}).
\end{aligned}$$

Therefore the Fourier series of $F(\mathbf{x})$ is

$$F(\mathbf{x}) = \sum_{\mathbf{t}\in\mathbb{Z}^n} (\mathcal{F}f)(\mathbf{t})e^{2\pi i\langle \mathbf{t},\mathbf{x}\rangle}.$$

and the first statement follows by the definition of $F(\mathbf{x})$. Setting $\mathbf{x} = \mathbf{0}$ proves the second statement. $\square$

In practical settings, we need a class of functions $f(\mathbf{x})$ for which the assumptions of the Poisson summation formula hold. We say that $f(\mathbf{x})$ is of **Schwarz class** if $f \in C^\infty(\mathbb{R}^n)$ and $f(\mathbf{x})$ along with all of its partial derivatives have rapid decay. If $f(\mathbf{x})$ is of Schwarz class, the rapid decay implies that $f(\mathbf{x})$ and all of its derivatives are absolutely integrable over $\mathbb{R}^n$. Moreover, this also implies that $F(\mathbf{x})$ and all of its derivatives are locally absolutely uniformly convergent by the Weierstrass M-test. The uniform limit theorem then implies $F(\mathbf{x})$ is smooth and thus the conditions of the Poisson summation formula are satisfied. A classical example of a Schwarz class function is $e^{-||\mathbf{x}||^2}$.

The Mellin Transform

Like the Fourier transform, the Mellin transform is another type of integral transform. If $f(x)$ is a continuous function and $f(\infty) = \lim_{x\to\infty} f(x)$ exists, then the **Mellin transform** $(\mathcal{M}f)(s)$ of $f(x)$ is given by

$$(\mathcal{M}f)(s) = \int_0^\infty (f(x) - f(\infty))x^s\,\frac{dx}{x}.$$

In most cases, we assume $f(x)$ has some type of decay so that $f(\infty) = 0$. Moreover, if $f(x)$ is a sufficiently nice function then the integral will be bounded in some half-plane in s. For example, this happens if $f(x)$ exhibits rapid decay and remains bounded as $x \to 0$. In this case, the integral is locally absolutely uniformly convergent for $\sigma > 0$. There is also an inverse transform. If $g(s)$ holomorphic and tends to zero as $t \to \infty$ in a vertical strip $a < \sigma < b$, then the inverse **inverse Mellin transform** $(\mathcal{M}^{-1}g)(x)$ of $g(s)$ is given by

$$(\mathcal{M}^{-1}g)(x) = \frac{1}{2\pi i}\int_{(c)} g(s)x^{-s}\,ds,$$

for any $a < c < b$. It is not immediately clear that this integral exists and is independent of c. The following theorem makes precise what properties $f(x)$ needs to satisfy so that the inverse Mellin transform recovers $f(x)$ (see [DB15] for a proof):

Theorem 1.6.4 (Mellin inversion formula). *Let $a < b$ and suppose $g(s)$ is analytic in the strip vertical $a < \sigma < b$, tends to zero uniformly as $t \to \infty$ along any line $\sigma = c$ for $a < c < b$, and that the integral of $g(s)$ along this line is locally absolutely uniformly convergent. Then if*

$$f(x) = \frac{1}{2\pi i}\int_{(c)} g(s)x^{-s}\,ds,$$

this integral is independent of c and moreover $g(s) = (\mathcal{M}f)(s)$. Conversely, suppose $f(x)$ is piecewise continuous such that its value is halfway between the limit values at any jump discontinuity and

$$g(s) = \int_0^\infty f(x)x^s\,\frac{dx}{x},$$

is locally absolutely uniformly convergent in the vertical strip $a < \sigma < b$. Then $f(x) = (\mathcal{M}^{-1}g)(x)$.

1.7 The Gamma Function

The gamma function is ubiquitous in analytic number theory and the better one understands this function the better one will be at seeing the forest for the trees. The **gamma function** $\Gamma(s)$ is defined to be the Mellin transform of e^{-x}:

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

for $\sigma > 0$. The integral is locally absolutely uniformly convergent in this region. Indeed, let K is a compact subset in this region and set $\alpha = \min_{s\in K}(\sigma)$. Then we have to show that $\Gamma(s)$ is absolutely uniformly convergent on K. Now split the integral by writing

$$\Gamma(s) = \int_0^1 e^{-x} x^{s-1}\, dx + \int_1^\infty e^{-x} x^{s-1}\, dx.$$

The second integral is absolutely uniformly convergent on K since the integrand exhibits rapid decay. As for the first integral, we have

$$\int_0^1 e^{-x} x^{s-1}\, dx \ll \int_0^1 x^{\sigma-1}\, dx \ll_\alpha 1,$$

so that this integral is absolutely uniformly convergent on K too. Thus $\Gamma(s)$ is absolutely uniformly convergent on K. Also note that $\Gamma(s)$ is real for $s > 0$. The most basic properties of $\Gamma(s)$ are the following:

Proposition 1.7.1. $\Gamma(s)$ *satisfies the following properties:*

(i) $\Gamma(1) = 1$.

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

(iii) $\Gamma(\overline{s}) = \overline{\Gamma(s)}$.

Proof. We obtain (i) by direct computation:

$$\Gamma(1) = \int_0^\infty e^{-x}\, dx = -e^{-x}\Big|_0^\infty = 1.$$

An application of integration by parts gives (ii):

$$\Gamma(s+1) = \int_0^\infty e^{-x} x^s\, dx = -e^{-x}x^s\Big|_0^\infty + s\int_0^\infty e^{-x} x^{s-1}\, dx = s\int_0^\infty e^{-x} x^{s-1}\, dx = s\Gamma(s).$$

For (iii), since $\Gamma(s)$ is real for $s > 0$ we have $\Gamma(\overline{s}) = \overline{\Gamma(s)}$ on this half-line and then the identity theorem implies that this holds everywhere. $\square$

From Proposition 1.7.1 we see that for $s = n$ a positive integer, $\Gamma(n) = (n-1)!$. So $\Gamma(s)$ can be thought of as a holomorphic extension of the factorial function. We can use property (ii) of Proposition 1.7.1 to extended $\Gamma(s)$ to a meromorphic function on all of $\mathbb{C}$:

Theorem 1.7.1. $\Gamma(s)$ *admits meromorphic continuation to* $\mathbb{C}$ *with poles at* $s = -n$ *for* $n \geq 0$. *All of these poles are simple and with residue* $\frac{(-1)^n}{n!}$ *at* $s = -n$.

Proof. Using Proposition 1.7.1, (ii) repeatedly, for any integer $n \geq 0$ we have

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

The right-hand side defines an meromorphic function in the region $\sigma > -n$ and away from the points $0, -1, \ldots, -n$. Letting n be arbitrary, we see that $\Gamma(s)$ has meromorphic continuation to $\mathbb{C}$ with poles at $0, -1, -2, \ldots$. We now compute the residue at $s = -n$. Around this point $\Gamma(s)$ admits meromorphic continuation with representation

$$\frac{\Gamma(s+1+n)}{s(s+1)\cdots(s+n)},$$

where all of the factors except for $s+n$ are holomorphic at $s = -n$. Thus the pole is simple, and

$$\operatorname*{Res}_{s=-n} \Gamma(s) = \lim_{z \to -n} \frac{\Gamma(s+1+n)(s+n)}{s(s+1)\cdots(s+n)} = \frac{\Gamma(1)}{(-n)(1-n)\cdots(-1)} = \frac{(-1)^n}{n!}. \qquad \square$$

In particular, Theorem 1.7.1 implies $\operatorname{Res}_{s=0}\Gamma(s) = 1$ and $\operatorname{Res}_{s=1}\Gamma(s) = -1$. There are a few other properties of the gamma function that are famous and which we will use frequently. The first of which is the **Legendre duplication formula** (see [Rem98] for a proof):

Theorem 1.7.2 (Legendre duplication formula). *For any $s \in \mathbb{C} - \{0, -1, -2, \ldots\}$,*

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

As a first application, we can use this formula to compute $\Gamma\left(\frac{1}{2}\right)$. Letting $z = \frac{1}{2}$ in the Legendre duplication formula and recalling $\Gamma(1) = 1$, we see that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. There is also the important Hadamard factorization of the reciprocal of $\Gamma(s)$ (see [SSS03] for a proof):

Proposition 1.7.2. *For all $s \in \mathbb{C}$,*

$$\frac{1}{\Gamma(s)} = se^{\gamma s} \prod_{n \geq 1} \left(1 + \frac{s}{n}\right) e^{-\frac{s}{n}},$$

where γ is the Euler-Mascheroni constant.

In particular, $\frac{1}{\Gamma(s)}$ is entire so that $\Gamma(s)$ is nowhere vanishing on $\mathbb{C}$. Also, $\frac{1}{\Gamma(s)}$ is of order 1 (see Appendix B.4). In particular, this means that $\Gamma(s)$ is also order 1 for $\sigma > 0$. We call $\frac{\Gamma'}{\Gamma}(s)$ the **digamma function**. Equivalently, the digamma function is the logarithmic derivative of the gamma function. Note that upon taking the logarithmic derivative of $\Gamma(s+1) = s\Gamma(s)$, we see that the digamma function satisfies the related formula

$$\frac{\Gamma'}{\Gamma}(s+1) = \frac{\Gamma'}{\Gamma}(s) + \frac{1}{s}.$$

If we take the logarithmic derivative of the Hadamard factorization for $\frac{1}{s\Gamma(s)}$, we obtain a useful expression for the digamma function:

Corollary 1.7.1. *For all $s \in \mathbb{C}$,*

$$\frac{\Gamma'}{\Gamma}(s+1) = -\gamma + \sum_{n \geq 1} \left(\frac{1}{n} - \frac{1}{s+n}\right),$$

where γ is the Euler-Mascheroni constant. In particular, the digamma function has simple poles of residue -1 at the poles of the gamma function.

Proof. By Proposition 1.7.1 (ii), $\frac{1}{\Gamma(s+1)} = \frac{1}{s\Gamma(s)}$. Taking the logarithmic derivative using Proposition 1.7.2 we obtain

$$-\frac{\Gamma'}{\Gamma}(s+1) = \gamma + \sum_{n\geq 1}\left(\frac{1}{s+n} - \frac{1}{n}\right),$$

provided s is distance ε away from the poles of $\Gamma(s)$. This is the desired formula and the statement regarding the poles follows immediately. □

We will also require a well-known approximation for the gamma function known as **Stirling's formula** (see [Rem98] for a proof):

Theorem 1.7.3 (Stirling's formula).

$$\Gamma(s) \sim_{\varepsilon,\delta} \sqrt{2\pi}s^{s-\frac{1}{2}}e^{-s},$$

provided $|\arg(s)| < \pi - \varepsilon$ *and* $|s| > \delta$ *for some* $\varepsilon, \delta > 0$.

If σ is bounded, Stirling's formula gives a useful asymptotic showing that $\Gamma(s)$ decays as $s \to \infty$:

Corollary 1.7.2. *Let* $|\arg(s)| < \pi - \varepsilon$ *and* $|s| > \delta$ *for some* $\varepsilon, \delta > 0$. *Then if* σ *is bounded, we have*

$$\Gamma(s) \sim_{\varepsilon,\delta} \sqrt{2\pi}t^{\sigma-\frac{1}{2}}e^{-\frac{\pi}{2}|t|}.$$

Proof. Stirling's formula can be equivalently expressed as

$$\Gamma(s) \sim_{\varepsilon,\delta} \sqrt{2\pi}(\sigma+it)^{\sigma-\frac{1}{2}+it}e^{-\sigma-it}.$$

Since σ is bounded, $e^{-\sigma-it} \ll 1$ and we obtain the the simplified asymptotic

$$\Gamma(s) \sim_{\varepsilon,\delta} \sqrt{2\pi}(it)^{\sigma-\frac{1}{2}+it}.$$

Similarly, x being bounded implies $i^{\sigma-\frac{1}{2}} \ll 1$ and we compute

$$(it)^{it} = e^{i|t|\log(i|t|)} = e^{i|t|(\log(i)+\log(|t|))} = e^{-\frac{\pi}{2}|t|+i|t|\log(|t|)} \sim e^{-\frac{\pi}{2}|t|},$$

where we have used the fact that $\log(i) = i\frac{\pi}{2}$. Together, we obtain the further simplified asymptotic

$$\Gamma(s) \sim_{\varepsilon,\delta} \sqrt{2\pi}t^{\sigma-\frac{1}{2}}e^{-\frac{\pi}{2}|t|},$$

which is the desired result □

Strictly weaker than the asymptotic in Stirling's formula is the asymptotic

$$\Gamma(s) = \sqrt{2\pi}s^{s-\frac{1}{2}}e^{-s}(1+O_{\varepsilon,\delta}(1)). \tag{1.4}$$

Taking the logarithm (since $|\arg(s)| < \pi - \varepsilon$ the logarithm is defined) of this asymptotic gives

$$\log\Gamma(s) = \frac{1}{2}\log(2\pi) + \left(s - \frac{1}{2}\right)\log(s) - s + O_{\varepsilon,\delta}(1), \tag{1.5}$$

which will be useful. In fact, from Equation (1.5) we can obtain another useful estimate formula for the digamma function:

Proposition 1.7.3.

$$\frac{\Gamma'}{\Gamma}(s) = \log(s) + O_{\varepsilon,\delta}(1),$$

provided $|\arg(s)| < \pi - \varepsilon$ *and* $|s| > \delta$ *for some* $\varepsilon, \delta > 0$.

Proof. Equation (1.5) give the simplified estimate

$$\log\Gamma(s) = \frac{1}{2}\log(2\pi) + s\log(s) - s + O_{\varepsilon,\delta}(1).$$

Set $g(s) = \frac{1}{2}\log(2\pi) + s\log(s) - s$ so that $\log\Gamma(s) = g(s) + O_{\varepsilon,\delta}(1)$. Then $\log\Gamma(s) - g(s) = O_{\varepsilon,\delta}(1)$, and by Cauchy's integral formula, we have

$$\begin{aligned}\frac{\Gamma'}{\Gamma}(s) &= \frac{d}{ds}\left(g(s) + O_{\varepsilon,\delta}(1)\right) \\ &= g'(s) + \frac{d}{ds}(\log\Gamma(s) - g(s)) \\ &= \log(s) + \frac{1}{2\pi i}\oint_C \frac{\log\Gamma(u) - g(u)}{(u-s)^2}\,du,\end{aligned}$$

for some sufficiently small circle C about s of radius η depending upon ε and δ. Therefore

$$\left|\frac{\Gamma'}{\Gamma}(s) - \log(s)\right| \le \frac{1}{2\pi}\oint_C \frac{|\log\Gamma(u) - g(u)|}{\eta^2}\,|du| \ll_{\varepsilon,d} 1,$$

where the last estimate follows because $\log\Gamma(s) - g(s) = O_{\varepsilon,d}(1)$. ☐

Lastly, we introduce a useful function related to the Gamma function. We define the **beta function** $B(s,u)$ by

$$B(s,u) = \int_0^1 t^{s-1}(1-t)^{u-1}\,dt,$$

for $\sigma > 0$ and $\tau > 0$. The integral is locally absolutely uniformly convergent in this region. Indeed, let $K \times L$ be a compact subset in this region and set $\alpha = \min_{s\in K}(\sigma)$ and $\beta = \min_{u\in L}(\tau)$. Then we have to show that $B(s,u)$ is absolutely uniformly convergent on $K \times L$. Observe

$$\int_0^1 t^{s-1}(1-t)^{u-1}\,dt \ll \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \ll_{\alpha,\beta} 1,$$

so that this integral is absolutely uniformly convergent on $K \times L$. Thus $B(s,u)$ is absolutely uniformly convergent on $K \times L$. Most importantly, the beta function is related to the gamma function in this region (see [Rem98] for a proof):

Proposition 1.7.4. *For* $\mathrm{Re}(s) > 0$ *and* $\mathrm{Re}(u) > 0$,

$$B(s,u) = \frac{\Gamma(s)\Gamma(u)}{\Gamma(s+u)}.$$

In particular, Proposition 1.7.4 shows that $B(s,u)$ has meromorphic continuation to $\mathbb{C}^2$ since the gamma function does by Theorem 1.7.1.

Part II

An Introduction to Number Fields

Chapter 2

The Theory of Number Fields

Introductory analytic number theory is done over $\mathbb{Q}$. The associated set of integers $\mathbb{Z}$ is a ring inside $\mathbb{Q}$. Moreover, the fundamental theorem of arithmetic tells us that prime factorization exists in $\mathbb{Z}$. That is, every integer is uniquely a product of primes (up to reordering of the factors). The study of number fields is concerned with finite extensions of $\mathbb{Q}$ where there might no longer be prime factorization. In the following, we discuss the structure of number fields, their associated ring of integers, and the properties of prime factorization.

2.1 Numbers Fields & Algebraic Integers

An **number field** K is a finite extension of $\mathbb{Q}$. That is, K is a finite dimensional vector space over $\mathbb{Q}$. In particular, $K/\mathbb{Q}$ is a finite separable extension, so that the primitive element theorem applies, and is Galois if and only if $K/\mathbb{Q}$ is normal. We say that the **degree** of K is $[K : \mathbb{Q}]$ which is the dimension of this vector space. If K is of degree 2, 3, etc. then we say it is **quadratic**, **cubic**, etc. Any $\kappa \in K$ is called an **algebraic number**. Moreover, we say that κ is an **algebraic integer** if it is the root of a monic polynomial $f(x) \in \mathbb{Z}[x]$. If $K = \mathbb{Q}$, it is clear that any integer is an algebraic integer (n is the root of $x - n$.). Moreover, any rational root of a monic polynomial must be an integer by the rational root theorem. In other words, if $f(x) \in \mathbb{Z}[x]$ is monic and $q \in \mathbb{Q}$ is a root of $f(x)$ then $q \in \mathbb{Z}$. Therefore for the number field $\mathbb{Q}$, the algebraic integers are exactly the integers $\mathbb{Z}$. Our first goal in studying number fields is to discuss the algebraic integers. Accordingly, we define the **ring of integers** $\mathcal{O}_K$ of K by

$$\mathcal{O}_K = \{\kappa \in K : \kappa \text{ is an algebraic integer}\}.$$

From what we have just shown above, $\mathbb{Z} \subseteq \mathcal{O}_K$. For a general number field K, $\mathcal{O}_K$ can be strictly larger than $\mathbb{Z}$. The ring of integers $\mathcal{O}_K$ is the analog of $\mathbb{Z}$ in $\mathbb{Q}$ but for K. Our primarily goals will be to show that $\mathcal{O}_K$ is a ring and more precisely a free abelian group of rank equal to the degree of K. The following proposition shows that $\mathcal{O}_K$ is indeed a ring:

Proposition 2.1.1. *Let K be a number field. Then the finitely many elements $\kappa_1, \ldots, \kappa_n \in B$ are all algebraic integers if and only if $\mathbb{Z}[\kappa_1, \ldots, \kappa_n]$ is a finitely generated $\mathbb{Z}$-module. In particular, $\mathcal{O}_K$ is a ring.*

Proof. First suppose $\kappa \in K$ is an algebraic integer. Then there exists a monic polynomial $f(x) \in \mathbb{Z}[x]$, of say degree $n \geq 1$, such that $f(\kappa) = 0$. Now for any $g(x) \in \mathbb{Z}[x]$, Euclidean division implies

$$g(x) = q(x)f(x) + r(x),$$

with $q(x), r(x) \in \mathbb{Z}[x]$ and $\deg(r(x)) < n$. Letting $r(x) = a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ with $a_i \in \mathbb{Z}$ for $0 \le i \le n-1$, it follows that

$$g(\kappa) = r(\kappa) = a_{n-1}\kappa^{n-1} + \cdots + a_1\kappa + a_0.$$

As $g(x)$ was arbitrary, we see that $\{1, \kappa, \ldots, \kappa^{n-1}\}$ is a generating set for $\mathbb{Z}[\kappa]$ as a $\mathbb{Z}$-module. Now suppose $\kappa_1, \ldots, \kappa_n \in K$ are all algebraic integers. We will prove that $\mathbb{Z}[\kappa_1, \ldots, \kappa_n]$ is finitely generated as an $\mathbb{Z}$-module by induction. Our previous work implies the base case. So assume by induction that $R = \mathbb{Z}[\kappa_1, \ldots, \kappa_{n-1}]$ is a finitely generated $\mathbb{Z}$-module. Then $R[\kappa_n] = A[\kappa_1, \ldots, \kappa_n]$ is a finitely generated R-module and hence a finitely generated $\mathbb{Z}$-module as well by our induction hypothesis. Now suppose $A[\kappa_1, \ldots, \kappa_n]$ is a finitely generated $\mathbb{Z}$-module. Let $\{\omega_1, \ldots, \omega_r\}$ be a basis for $A[\kappa_1, \ldots, \kappa_n]$. Then for any $\kappa \in A[\kappa_1, \ldots, \kappa_n]$, we have

$$\kappa\omega_i = \sum_{1 \le j \le r} a_{i,j}\omega_j,$$

with $a_{i,j} \in \mathbb{Z}$ for $1 \le i, j \le r$.We can rewrite this as,

$$(\kappa - a_{i,i})\omega_i - \sum_{\substack{1 \le j \le r \\ j \ne i}} a_{i,j}\omega_j = 0,$$

for all i. These r equations are equivalent to the identity

$$\begin{pmatrix} k - a_{1,1} & a_{1,2} & \cdots & -a_{1,r} \\ -a_{2,1} & \kappa - a_{2,2} & & \\ \vdots & & \ddots & \\ -a_{r,1} & & & \kappa - a_{r,r} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_r \end{pmatrix} = \mathbf{0}.$$

Thus the determinant of the matrix on the left-hand side must be zero. This shows that κ is the root of the characteristic polynomial $\det(xI - (a_{i,j}))$ which is a monic polynomial with coefficients in $\mathbb{Z}$. Hence κ is an algebraic integer. As κ was arbitrary, this shows that the elements $\kappa_1, \ldots, \kappa_n$ are all algebraic integers and that the sum and product of algebraic integers are algebraic integers. It follows that $\mathcal{O}_K$ is a ring. $\square$

We can also show that K is the field of fractions of $\mathcal{O}_K$. Actually, the following proposition proves this and more:

Proposition 2.1.2. *Let K be a number field. Then every $\kappa \in K$ is of the form*

$$\kappa = \frac{\alpha}{a},$$

for some $\alpha \in \mathcal{O}_K$ and nonzero $a \in \mathbb{Z}$. In particular, K is the field of fractions of $\mathcal{O}_K$. Moreover, $\kappa \in K$ is an algebraic integer if and only if the minimal polynomial of κ has coefficients in $\mathbb{Z}$.

Proof. As $K/\mathbb{Q}$ is finite, it is necessarily algebraic so that any $\kappa \in K$ satisfies

$$a\kappa^n + a_{n-1}\kappa^{n-1} + \cdots + a_0 = 0,$$

with $a_i \in \mathbb{Z}$ for $0 \le i \le n-1$ and $a \ne 0$. We claim that $a\kappa$ is an algebraic integer. Indeed, multiplying the previous identity by a^{n-1} yields

$$(a\kappa)^n + a'_{n-1}(a\kappa)^{n-1} + \cdots + a'_0 = 0,$$

where $a_i' = a_i a^{n-1-i}$ for $0 \leq i \leq n-1$, and so $a\kappa$ is the root of a monic polynomial with coefficients in $\mathbb{Z}$. Then $a\kappa \in \mathcal{O}_K$ and so $a\kappa = \alpha$ for some $\alpha \in \mathcal{O}_K$ which is equivalent to $\kappa = \frac{\alpha}{a}$. As $\mathbb{Z} \subseteq \mathcal{O}_K$, this also implies that K is the field of fractions of $\mathcal{O}_K$. For the last statement, suppose $\kappa \in K$. If the minimal polynomial of κ has integer coefficients then κ is automatically an algebraic integer (since the minimal polynomial is monic). So suppose κ is an algebraic integer so that κ is a root of a monic polynomial $f(x) \in \mathbb{Z}[x]$. If $m_\kappa(x) \in \mathbb{Q}[x]$ is the minimal polynomial of κ, then $m_\kappa(x)$ divides $f(x)$ and thus all of the roots of $m_\kappa(x)$ algebraic integers too. By Vieta's formulas, the coefficients of $m_\kappa(x)$ algebraic integers as well. But then $m_\kappa(x) \in \mathbb{Z}[x]$. This completes the proof. $\square$

2.2 Traces & Norms

We will now require norms and traces of free algebras over fields. Let K be a field and let R be a free K-algebra of rank n. Then the **trace** and **norm** of R, denoted $\mathrm{Tr}_{R/K}$ and $\mathrm{N}_{R/K}$ respectively, are defined by

$$\mathrm{Tr}_{R/K}(\rho) = \mathrm{trace}(T_\rho) \quad \text{and} \quad \mathrm{N}_{R/K}(\rho) = \det(T_\rho),$$

for any $\rho \in R$, where $T_\rho : R \to R$ is the linear operator defined by

$$T_\rho(x) = \rho x,$$

for all $x \in R$. That is, T_ρ is the multiplication by ρ map. Letting $f_\rho(t)$ denote the characteristic polynomial of T_ρ, we have

$$f_\rho(t) = \det(tI - T_\rho) = t^n - \kappa_{n-1}t^{n-1} + \cdots + (-1)^n \kappa_0,$$

with $\kappa_i \in K$ for $0 \leq i \leq n-1$. Then the trace and the norm are given by

$$\mathrm{Tr}_{R/K}(\rho) = \kappa_{n-1} \quad \text{and} \quad \mathrm{N}_{R/K}(\rho) = \kappa_0, \tag{2.1}$$

and therefore take values in K. Moreover, we have

$$\mathrm{Tr}_{R/K}(\kappa\rho) = \kappa \,\mathrm{Tr}_{R/K}(\rho) \quad \text{and} \quad \mathrm{N}_{R/K}(\kappa\rho) = \kappa^m \,\mathrm{N}_{R/K}(\rho),$$

for all $\kappa \in K$ because $T_{\kappa\lambda} = \kappa T_\lambda$. As $T_{\lambda+\nu} = T_\lambda + T_\nu$ and $T_{\lambda\nu} = T_\lambda T_\nu$, we obtain homomorphisms

$$\mathrm{Tr}_{R/K} : R \to K \quad \text{and} \quad \mathrm{N}_{R/K} : R \to K.$$

In the case of a degree n extension L/K, we call $\mathrm{Tr}_{L/K}$ and $\mathrm{N}_{L/K}$ the **trace** and **norm** of L/K. Moreover, $\mathrm{N}(\lambda) = 0$ if and only if $\lambda = 0$ because otherwise T_λ has inverse $T_{\lambda^{-1}}$ and hence nonzero determinant. Therefore we obtain homomorphisms

$$\mathrm{Tr}_{L/K} : L \to K \quad \text{and} \quad \mathrm{N}_{L/K} : L^* \to K^*.$$

In the specialized setting $K/\mathbb{Q}$ for a number field K, we write $\mathrm{Tr} = \mathrm{Tr}_{K/\mathbb{Q}}$ and $\mathrm{N} = \mathrm{N}_{K/\mathbb{Q}}$. Moreover, for any $\kappa \in K$ we call $\mathrm{Tr}(\kappa)$ and $\mathrm{N}(\kappa)$ the **trace** and **norm** of κ respectively. More generally, when L/K is separable, we can derive alternative descriptions of the trace and norm of L/K. This additional assumption is weak because the only situations we will be interested in are finite extensions of $\mathbb{Q}$ and $\mathbb{F}_p$ which are always separable (because both $\mathbb{Q}$ and $\mathbb{F}_p$ are perfect). In any case, to do this we need to work in the algebraic closure $\overline{K}$ of K. As L/K is a degree n separable extension, there are exactly n distinct K-embeddings $\sigma_1, \ldots, \sigma_n$ of L into $\overline{K}$ (each given by letting θ be a primitive element for L so that $L = K[\theta]$ and sending θ to one of its conjugate roots in the minimal polynomial $m_\theta(x)$ of θ). Clearly $\sigma_1, \ldots, \sigma_n$ send $\mathcal{O}_K$ to itself and fix $\mathcal{O}_K$ pointwise. Moreover, we prove the following proposition:

Proposition 2.2.1. *Let L/K be a degree n separable extension and let σ run over all K-embeddings σ of L into $\overline{K}$. For any $\lambda \in L$, the characteristic polynomial $f_\lambda(t)$ of T_λ is a power of the minimal polynomial of λ and satisfies*

$$f_\lambda(t) = \prod_\sigma (t - \sigma(\lambda)).$$

In particular,

$$\mathrm{Tr}_{L/K}(\lambda) = \sum_\sigma \sigma(\lambda) \quad \text{and} \quad \mathrm{N}_{L/K}(\lambda) = \prod_\sigma \sigma(\lambda).$$

Proof. Let

$$m_\lambda(t) = t^m + \kappa_{m-1}t^{m-1} + \cdots + \kappa_0,$$

with $\kappa_i \in K$ for $0 \le i \le n-1$, be the minimal polynomial of λ (necessarily m is the degree of $K(\lambda)/K$). Let d be the degree of $L/K(\lambda)$. We first show that $f_\lambda(t)$ is a power of $m_\lambda(t)$. Precisely, we claim that

$$f_\lambda(t) = m_\lambda(t)^d.$$

To see this, recall that $\{1, \lambda, \dots, \lambda^{n-1}\}$ is a basis for $K(\lambda)/K$. If $\{\alpha_1, \dots, \alpha_d\}$ is a basis for $L/K(\lambda)$, then

$$\{\alpha_1, \alpha_1\lambda, \dots, \alpha_1\lambda^{m-1}, \dots, \alpha_d, \alpha_d\lambda, \dots, \alpha_d\lambda^{m-1}\},$$

is a basis for L/K. Because the minimal polynomial $m_\lambda(t)$ gives the linear relation

$$\lambda^m = -\kappa_0 - k_1\lambda - \cdots - \kappa_{m-1}\lambda^{m-1},$$

the matrix of T_λ is block diagonal with d blocks each of the form

$$\begin{pmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 \\ -\kappa_0 & -\kappa_1 & \cdots & -\kappa_{m-1} \end{pmatrix}.$$

This is the companion matrix to $m_\lambda(t)$ and hence the characteristic polynomial is $m_\lambda(t)$ as well. Our claim follows since the matrix of T_λ is block diagonal. Since λ is algebraic over K of degree m, $K(\lambda)$ is the splitting field of $m_\lambda(t)$ and there are m distinct K-embeddings of $K(\lambda)$ into $\overline{K}$. Let τ be such an Kembedding. Then the K-embeddings σ are partitioned into m many equivalence classes each of size d (because $L/K(\lambda)$ is degree d) where σ and σ' are in the same class if and only if $\sigma(\lambda) = \sigma'(\lambda)$. In particular, a complete set of representatives is given by the τ. But then

$$f_\lambda(t) = m_\lambda(t)^d = \left(\prod_\tau (t - \tau(\lambda))\right)^d = \prod_\sigma (t - \sigma(\lambda)),$$

which proves the first statement. The formulas for the trace and norm follow from Vieta's formulas and Equation (2.1). $\square$

As a corollary of Proposition 2.2.1, we can show how the trace and norm act on algebraic integers for the extension $K/\mathbb{Q}$:

Corollary 2.2.1. *Let K be a number field. If $\kappa \in K$ is an algebraic integer, then the trace and norm of κ are integers.*

Proof. By Proposition 2.1.2, if κ is an algebraic integer then its minimal polynomial $m_\kappa(t)$ has integer coefficients. By Proposition 2.2.1 the characteristic polynomial $f_\kappa(t)$ is a power of $m_\kappa(t)$. Hence $f_\kappa(t)$ has integer coefficients. From Equation (2.1) we conclude that the trace and norm of κ are integers. $\square$

We can also classify the units in $\mathcal{O}_K$ according to their norm:

Corollary 2.2.2. *Let K be a number field. Then $\alpha \in \mathcal{O}_K$ is a unit if and only if its norm is ± 1.*

Proof. Let $\alpha \in \mathcal{O}_K$. First suppose α is a unit in $\mathcal{O}_K$. Then $\frac{1}{\alpha} \in \mathcal{O}_K$ and so

$$\mathrm{N}(\alpha)\,\mathrm{N}\left(\frac{1}{\alpha}\right) = \mathrm{N}(1) = 1.$$

By Corollary 2.2.1, the norm of α and $\frac{1}{\alpha}$ are both integers. Hence they must be ± 1 and thus the norm of α is ± 1. Now suppose the norm of α is ± 1. By Proposition 2.1.2, its minimal polynomial $m_\alpha(t)$ has integer coefficients. Moreover, Equation (2.1) and Proposition 2.2.1 together imply that the constant term is ± 1. Letting the degree of $m_\alpha(t)$ be m, we have shown that

$$m_\alpha(t) = t^m + a_{m-1}t^{m-1} + \cdots \pm 1,$$

with $a_i \in \mathbb{Z}$ for $1 \leq i \leq m-1$. Dividing $m_\alpha(\alpha)$ by α^m, we find that $\frac{1}{\alpha}$ is a root of the polynomial

$$f(x) = \pm t^m + a_1 t^{m-1} + \cdots + 1.$$

Multiplying by -1 if necessary, it follows that $\frac{1}{\alpha}$ is a root of a monic polynomial with coefficients in $\mathbb{Z}$. Hence $\frac{1}{\alpha} \in \mathcal{O}_K$ and thus α is a unit in $\mathcal{O}_K$. $\square$

We can now prove a structure theorem for the ring of integers $\mathcal{O}_K$ of a number field K. We show that the ring of integers is a free abelian group with rank equal to the degree of K which clearly is a generalization of the structure of $\mathbb{Z}$ for for the number field $\mathbb{Q}$:

Theorem 2.2.1. *Let K be a number field of degree n. Then $\mathcal{O}_K$ is a free abelian group of rank n. In particular, $\mathcal{O}_K$ is a finitely generated $\mathbb{Z}$-module.*

Proof. Let $\{\kappa_1, \ldots, \kappa_n\}$ be a basis for K. By Proposition 2.1.2, we have $\kappa_i = \frac{\alpha_i}{a_i}$ with $\alpha_i \in \mathcal{O}_K$ and $a_i \in \mathbb{Z}$ for $1 \leq i \leq n$. Hence $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for K as well. In particular, any element $\alpha \in \mathcal{O}_K$ can be expressed as

$$\alpha = \sum_{1 \leq i \leq n} q_i(\alpha)\alpha_i,$$

with $q_i(\alpha) \in \mathbb{Q}$. We now show that the denominators of the $q_i(\alpha)$ are uniformly bounded for all $1 \leq i \leq n$ and all α. Assume this is not the case. Then there is a sequence $(\beta_j)_{j \geq 1}$ of nonzero elements in $\mathcal{O}_K$ where

$$\beta_j = \sum_{1 \leq i \leq n} q_i(\beta_j)\alpha_i,$$

is such that the greatest denominator of $q_i(\beta_j)$ for $1 \leq i \leq n$ tends to infinity as $j \to \infty$. In terms of the basis $\{\alpha_1, \ldots, \alpha_n\}$, $\mathrm{N}(\beta_j)$ is the determinant of an $n \times n$ matrix with coefficients in $\mathbb{Q}[q_i(\beta_j)]_{1 \leq i \leq n}$. In particular, it is a homogenous polynomial of degree n in the $q_i(\beta_j)$ for $1 \leq i \leq n$ with coefficients in $\mathbb{Q}$ determined by the basis $\{\alpha_1, \ldots, \alpha_n\}$. But $\mathrm{N}(\beta_j)$ is an integer by Corollary 2.2.1. It is also nonzero because β_j is nonzero. Hence $|\,\mathrm{N}(\beta_j)| \geq 1$ and thus, by what we have just shown, the greatest denominator of $q_i(\beta_j)$ for $1 \leq i \leq n$ must be bounded as $j \to \infty$. This gives a contradiction. Hence there is an integer $M \geq 1$ such that $Mq_i(\alpha) \in \mathbb{Z}$ for all $1 \leq i \leq n$ and $\alpha \in \mathcal{O}_K$. Therefore

$$\mathcal{O}_K \subseteq \frac{1}{M} \bigoplus_{1 \leq i \leq n} \mathbb{Z}\alpha_i.$$

As the group on the right-hand side is a free abelian group so is $\mathcal{O}_K$. Moreover, as $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for K we see that $\{\alpha_1, \ldots, \alpha_n\}$ is linearly independent over $\mathbb{Z}$ as well. This means that the rank of $\mathcal{O}_K$ must be n. The last statement is now clear. $\square$

In accordance with Theorem 2.2.1, we say that $\{\alpha_1, \ldots, \alpha_n\}$ is an **integral basis** for K if $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for K and $\mathcal{O}_K$ can be expressed as

$$\mathcal{O}_K = \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n.$$

That is, every $\alpha \in \mathcal{O}_K$ is a unique integer linear combination of the α_i. An integral basis for K always exists by Theorem 2.2.1. In the special case $\mathcal{O}_K = \mathbb{Z}[\alpha]$ for some $\alpha \in \mathcal{O}_K$, we say K is **monogenic**. By Theorem 2.2.1, we see that $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is an integral basis for K. Lastly, we can show that algebraic integers satisfy a slightly weaker condition:

Proposition 2.2.2. *Let K be a number field. Then $\kappa \in K$ is an algebraic integer if and only if κ is the root of a monic polynomial with coefficients in $\mathcal{O}_K$.*

Proof. If $\kappa \in K$ is an algebraic integer, then κ is the root of a monic polynomial with coefficients in $\mathbb{Z}$ and hence in $\mathcal{O}_K$ as well. So suppose $\kappa \in K$ is the root of a monic polynomial $f(x) \in \mathcal{O}_K$. Let $f(x)$ have degree n and write

$$f(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_0,$$

with $\alpha_i \in \mathcal{O}_K$ for $0 \leq i \leq n-1$. As $f(\kappa) = 0$, we have

$$\kappa^n = -\alpha_{n-1}\kappa^{n-1} - \cdots - \alpha_0,$$

and hence $\mathcal{O}_K[\kappa]$ is a finitely generated $\mathbb{Z}$-module because $\mathcal{O}_K$ is by Theorem 2.2.1. As $\mathbb{Z}[\kappa] \subseteq \mathcal{O}_K[\kappa]$, we see that $\mathbb{Z}[\kappa]$ must also be a finitely generated $\mathbb{Z}$-module. Hence κ is an algebraic integer by Proposition 2.1.1. $\square$

2.3 Discriminants

We will now discuss discriminants of free modules over fields. Let K be a field and R be a free K-algebra of rank n. If $\{\rho_1, \ldots, \rho_n\}$ is a basis for R, we set

$$\mathrm{disc}_{R/K}(\rho_1, \ldots, \rho_n) = \det((\mathrm{Tr}_{R/K}(\rho_i\rho_j))_{i,j}).$$

In particular, $\mathrm{disc}_{R/K}(\rho_1, \ldots, \rho_n)$ is an element of K. It is also independent of the choice of basis up to elements of $(K^*)^2$. Indeed, if $\{\rho'_1, \ldots, \rho'_n\}$ is another basis then we have

$$\rho'_i = \sum_{1 \leq j \leq m} \kappa_{i,j}\rho_j,$$

with $\kappa_{i,j} \in K$ for $1 \leq i, j \leq n$. Then $(\kappa_{i,j})_{i,j}$ is the base change matrix from $\{\rho_1, \ldots, \rho_n\}$ to $\{\rho'_1, \ldots, \rho'_n\}$ and so it has nonzero determinant. Thus $\det((\kappa_{i,j})_{i,j}) \in K^*$. Moreover, we have

$$(\mathrm{Tr}_{R/K}(\rho'_i\rho'_j))_{i,j} = (\kappa_{i,j})_{i,j}(\mathrm{Tr}_{R/K}(\rho_i\rho_j))_{i,j}(\kappa_{i,j})^t_{i,j},$$

which, upon taking the determinant, shows that

$$\operatorname{disc}_{R/K}(\rho_1',\dots,\rho_n') = \det((\kappa_{i,j})_{i,j})^2 \operatorname{disc}_{R/K}(\rho_1,\dots,\rho_n), \tag{2.2}$$

as claimed. We define the **discriminant** $\operatorname{disc}_K(R)$ of R by

$$\operatorname{disc}_K(R) = \operatorname{disc}_{R/K}(\rho_1,\dots,\rho_n) \pmod{(K^*)^2}.$$

for any basis $\{\rho_1,\dots,\rho_n\}$ of R. By what we have shown, $\operatorname{disc}_K(R)$ is well-defined. The discriminant is also multiplicative with respect to direct sums:

Proposition 2.3.1. *Let K be a field and R be a free K-algebra of rank n. Suppose we have a direct sum decomposition*

$$R = R_1 \oplus R_2,$$

for free K-algebras R_1 and R_2 of ranks n_1 and n_2 respectively. Also let $\{\eta_1,\dots,\eta_{n_1}\}$ and $\{\gamma_1,\dots,\gamma_{n_2}\}$ be bases of R_1 and R_2 respectively. Then

$$\operatorname{disc}_{R/K}(\eta_1,\dots,\eta_{n_1},\gamma_1,\dots,\gamma_{n_2}) = \operatorname{disc}_{R/K}(\eta_1,\dots,\eta_{n_1}) \operatorname{disc}_R(\gamma_1,\dots,\gamma_{n_2}).$$

In particular,

$$\operatorname{disc}_K(R) = \operatorname{disc}_K(R_1)\operatorname{disc}_K(R_2).$$

Proof. The second statement follows immediately from the first. To prove the first statement, write

$$\operatorname{disc}_{R/K}(\eta_1,\dots,\eta_{n_1}) = \det((\operatorname{Tr}_{R/K}(\eta_i\eta_j))_{i,j}) \quad \text{and} \quad \operatorname{disc}_R(\gamma_1,\dots,\gamma_{n_2}) = \det((\operatorname{Tr}_{R/K}(\gamma_k\gamma_\ell))_{k,\ell}).$$

As R is the direct sum of R_1 and R_2 as K-modules, we have $\eta_i\gamma_k = 0$ for all $1 \le i \le n_1$ and $1 \le k \le n_2$. It follows that $\operatorname{disc}_{R/K}(\eta_1,\dots,\eta_{n_1},\gamma_1,\dots,\gamma_{n_2})$ is the determinant of the block diagonal matrix

$$\begin{pmatrix} (\operatorname{Tr}_{R/K}(\eta_i\eta_j))_{i,j} & \\ & (\operatorname{Tr}_{R/K}(\gamma_k\gamma_\ell))_{k,\ell} \end{pmatrix}.$$

Moreover, we have

$$\operatorname{Tr}_{R/K}(\rho_1) = \operatorname{Tr}_{R_1/K}(\rho_1) \quad \text{and} \quad \operatorname{Tr}_{R/K}(\rho_2) = \operatorname{Tr}_{R_2/K}(\rho_2)$$

for any $\rho_1 \in R_1$ and $\rho_2 \in R_2$. Indeed, multiplication by ρ_1 and ρ_2 annihilate R_2 and R_1 respectively. But then

$$\begin{pmatrix} (\operatorname{Tr}_{R/K}(\eta_i\eta_j))_{i,j} & \\ & (\operatorname{Tr}_{R/K}(\gamma_k\gamma_\ell))_{k,\ell} \end{pmatrix} = \begin{pmatrix} (\operatorname{Tr}_{R_1/K}(\eta_i\eta_j))_{i,j} & \\ & (\operatorname{Tr}_{R_2/K}(\gamma_k\gamma_\ell))_{k,\ell} \end{pmatrix}.$$

The determinant of the matrix on right-hand side is $\operatorname{disc}_{R/K}(\eta_1,\dots,\eta_{n_1})\operatorname{disc}_R(\gamma_1,\dots,\gamma_{n_2})$. This completes the proof. ∎

We now specialize to the setting of a degree n separable extension L/K. It turns out that the discriminant is nonzero. To see this, we require a lemma:

Lemma 2.3.1. *Let L/K be a finite separable extension. Then the map*

$$\operatorname{Tr}_{L/K} : L \times L \to K \qquad (\lambda,\eta) \to \operatorname{Tr}_{L/K}(\lambda\eta),$$

is a nondegenerate symmetric bilinear form.

Proof. From the definition of the trace, it is clear that the map is a symmetric bilinear form. To see that is is nondegenerate, suppose L/K is degree n. Then for any nonzero $\lambda \in L$, Proposition 2.2.1 implies that

$$\mathrm{Tr}_{L/K}(\lambda\lambda^{-1}) = \mathrm{Tr}_{L/K}(1) = n.$$

Hence the symmetric bilinear form is nondegenerate. □

We can now show that the discriminant is never zero:

Proposition 2.3.2. *Let L/K be a degree n separable extension and let $\{\lambda_1, \dots, \lambda_n\}$ be a basis for L. Then we have that $\mathrm{disc}_K(\lambda_1, \dots, \lambda_n) \neq 0$. In particular, $\mathrm{disc}_K(L) \neq 0$.*

Proof. The second statement follows immediately from the first. To prove the first statement, suppose to the contrary that $\mathrm{disc}_K(\lambda_1, \dots, \lambda_n) = 0$. Then the matrix $(\mathrm{Tr}_{L/K}(\lambda_i\lambda_j))_{i,j}$ is not invertible. Hence there exists a nonzero column vector $(\kappa_1, \dots, \kappa_n)^t$ with $\kappa_i \in K$ for $1 \leq i \leq n$ such that

$$(\mathrm{Tr}_{L/K}(\lambda_i\lambda_j))_{i,j}(\kappa_1, \dots, \kappa_n)^t = \mathbf{0}.$$

This is equivalent to the n equations

$$\sum_{1 \leq j \leq n} \kappa_j \,\mathrm{Tr}_{L/K}(\lambda_i\lambda_j) = 0,$$

for all i. Setting

$$\lambda = \sum_{1 \leq j \leq n} \kappa_j\lambda_j,$$

linearity of the trace implies that these n equations are equivalent to the fact that $\mathrm{Tr}_{L/K}(\lambda\lambda_i) = 0$ for all i. As $\{\lambda_1, \dots, \lambda_n\}$ is a basis for L, it follows that $\lambda \in L$ is a nonzero element for which $\mathrm{Tr}_{L/K}(\lambda\eta) = 0$ for all $\eta \in L$. This is impossible by Lemma 2.3.1. Hence $\mathrm{disc}_K(\lambda_1, \dots, \lambda_n) \neq 0$. □

In addition to $\mathrm{disc}_K(\lambda_1, \dots, \lambda_n)$ never vanishing, we can also write it in an alternative form. To do this, for any basis $\{\lambda_1, \dots, \lambda_n\}$ of L we define the associated **embedding matrix** $M(\lambda_1, \dots, \lambda_n)$ by

$$M(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \sigma_1(\lambda_1) & \cdots & \sigma_1(\lambda_n) \\ \vdots & & \vdots \\ \sigma_n(\lambda_1) & \cdots & \sigma_n(\lambda_n) \end{pmatrix},$$

where $\sigma_1, \dots, \sigma_n$ are the n distinct K-embeddings of L into $\overline{K}$. Then we have the following result:

Proposition 2.3.3. *Let L/K be a degree n separable extension. Then for any basis $\{\lambda_1, \dots, \lambda_n\}$ of L, we have*

$$\mathrm{disc}_K(\lambda_1, \dots, \lambda_n) = \det(M(\lambda_1, \dots, \lambda_n))^2.$$

Proof. Recalling that the (i, j)-entry of $M(\lambda_1, \dots, \lambda_n)^tM(\lambda_1, \dots, \lambda_n)$ is the dot product of the i-th and j-th columns of $M(\lambda_1, \dots, \lambda_n)$, we have

$$\begin{aligned}
\det(M(\lambda_1, \dots, \lambda_n))^2 &= \det(M(\lambda_1, \dots, \lambda_n)^tM(\lambda 1, \dots, \lambda_n)) \\
&= \det\left(\left(\sum_\sigma \sigma(\lambda_i)\sigma(\lambda_j)\right)_{i,j}\right) \\
&= \det\left(\left(\sum_\sigma \sigma(\lambda_i\lambda_j)\right)_{i,j}\right) \\
&= \det((\mathrm{Tr}_{L/K}(\lambda_i\lambda_j))_{i,j}) \\
&= \mathrm{disc}_{L/K}(\lambda_1, \dots, \lambda_n),
\end{aligned}$$

where the sums run over all K-embeddings σ of L into $\overline{K}$ and the second to last equality follows by Proposition 2.2.1, as desired. $\square$

In the specialized case $K/\mathbb{Q}$ for a number field K, we define the **discriminant** Δ_K of K by

$$\Delta_K = \mathrm{disc}_K(\alpha_1, \ldots, \alpha_n),$$

for any integral basis $\{\alpha_1, \ldots, \alpha_n\}$. As Δ_K is not defined modulo $(K^*)^2$, we need to show that Δ_K is independent of the choice of integral basis and hence well-defined. Indeed, if $\{\alpha'_1, \ldots, \alpha'_n\}$ is another integral basis then the base change matrix, as well as its inverse, both have integer entries (because integral bases are bases for $\mathcal{O}_K$ as a $\mathbb{Z}$-module). This implies that the determinant of the base change matrix is ± 1 and so Equation (2.2) shows that Δ_K is independent of the choice of integral basis. Moreover, Δ_K is nonzero by Proposition 2.3.2 and

$$\Delta_K = \det(M(\alpha_1, \ldots, \alpha_n))^2,$$

by Proposition 2.3.3.

2.4 Integral & Fractional Ideals

For the number field $\mathbb{Q}$, its ring of integers $\mathbb{Z}$ is a unique factorization domain. Indeed, this is just a restatement of the fundamental theorem of arithmetic. Unfortunately, for a general number field K its ring of integers $\mathcal{O}_K$ need not be a unique factorization domain. However, the integral ideals of $\mathcal{O}_K$ do factor into a unique product of prime integral ideals (this is trivial for a unique factorization domain). Our main goal is to prove this. We first introduce some notation. Any nonzero ideal $\mathfrak{a}$ of $\mathcal{O}_K$ is said to be an **integral ideal** of K. As $\mathcal{O}_K$ is a free abelian group of rank n by Theorem 2.2.1, any subgroup is also free abelian. Since $\alpha\mathcal{O}_K \subseteq \mathfrak{a}$ for any nonzero $\alpha \in \mathfrak{a}$, it follows that $\mathfrak{a}$ is also a free abelian group of rank n as well. We first show that the quotient ring by an integral ideal is finite:

Proposition 2.4.1. *Let K be a number field. Then $\mathcal{O}_K/\mathfrak{a}$ is finite for any integral ideal $\mathfrak{a}$. In particular, for any $\alpha \in \mathcal{O}_K$ we have*

$$|\mathcal{O}_K/\alpha\mathcal{O}_K| = |\,\mathrm{N}(\alpha)|.$$

Proof. Since $\mathfrak{a}$ and $\mathcal{O}_K$ are both free abelian groups of rank n (recall Theorem 2.2.1) and $\mathfrak{a}$ is a subgroup, the first statement follows by Proposition C.1.1. For the second statement, let $\{\alpha_1, \ldots, \alpha_n\}$ be an integral basis for K. Writing

$$\alpha = \sum_{1 \le i \le n} a_i\alpha_i,$$

with $a_i \in \mathbb{Z}$, we see that $\{a_1\alpha_1, \ldots, a_n\alpha_n\}$ is a basis for $\alpha\mathcal{O}_K$. In particular, the base change matrix from $\{\alpha_1, \ldots, \alpha_n\}$ to this basis is the diagonal with the a_i on the diagonal. Then on the one hand, we have $|\mathcal{O}_K/\alpha\mathcal{O}_K| = |a_1 \cdots a_n|$ by Proposition C.1.1 again. On the other hand, in terms of the basis $\{a_1\alpha_1, \ldots, a_n\alpha_n\}$ the map T_α is given by

$$T_\alpha = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix},$$

and so $\mathrm{N}(\alpha) = a_1 \cdots a_n$. Hence

$$|\mathcal{O}_K/\alpha\mathcal{O}_K| = |\,\mathrm{N}(\alpha)|,$$

as desired. $\square$

For an integral ideal $\mathfrak{a}$, we define its **norm** $\mathrm{N}(\mathfrak{a})$ by

$$\mathrm{N}(\mathfrak{a}) = |\mathcal{O}_K/\mathfrak{a}|.$$

By Proposition 2.4.1, the norm is finite, necessarily a positive integer, and for every $\alpha \in \mathcal{O}_K$ we have

$$\mathrm{N}(\alpha\mathcal{O}_K) = |\,\mathrm{N}(\alpha)|.$$

We can now show that every prime integral ideal is maximal:

Proposition 2.4.2. *Let K be a number field. Every prime integral ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is maximal.*

Proof. Recall that an ideal is maximal if and only if the quotient ring is a field. Therefore it suffices to show that $\mathcal{O}_K/\mathfrak{p}$ is a field. Let $\alpha \in \mathcal{O}_K/\mathfrak{p}$ be nonzero. We will show that α is invertible in $\mathcal{O}_K/\mathfrak{p}$. Since $\mathfrak{p}$ is maximal, $\mathcal{O}_K/\mathfrak{p}$ is an integral domain. Therefore the multiplication map

$$\mathcal{O}_K/\mathfrak{p} \to \mathcal{O}_K/\mathfrak{p} \qquad x \mapsto \alpha x,$$

is injective. By Proposition 2.4.1, $\mathcal{O}_K/\mathfrak{p}$ is finite and therefore this map must be a bijection. But this means that α has an inverse in $\mathcal{O}_K/\mathfrak{p}$. Hence $\mathcal{O}_K/\mathfrak{p}$ is a field. $\square$

As prime integral ideals are maximal by Proposition 2.4.2 and distinct maximal ideals are relatively prime, we see that distinct prime integral ideals $\mathfrak{p}$ and $\mathfrak{q}$ are relatively prime. In particular, their powers are relatively prime as well (which follows by induction). We will now being working to show that every integral ideal factors uniquely into a product of prime integral ideals. First we show containment in one direction:

Lemma 2.4.1. *Let K be a number field. For every integral ideal $\mathfrak{a}$, there exist prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ such that*

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \mathfrak{a}.$$

Proof. Let $\mathcal{S}$ be the set of integral ideals which do not contain a product of prime integral. Then it suffices to show $\mathcal{S}$ is empty. Assume otherwise so that there is an integral ideal $\mathfrak{a} \in \mathcal{S}$. Then $\mathfrak{a}$ cannot be prime for otherwise $\mathfrak{a}$ contains a product of prime integral ideals (namely itself). Since $\mathfrak{a}$ is not prime, there exist $\alpha_1, \alpha_2 \in \mathcal{O}_K$ with $\alpha_1\alpha_2 \in \mathfrak{a}$ and such that $\alpha_1, \alpha_2 \notin \mathfrak{a}$. Now define integral ideals

$$\mathfrak{b}_1 = \mathfrak{a} + \alpha_1\mathcal{O}_K \quad \text{and} \quad \mathfrak{b}_2 = \mathfrak{a} + \alpha_2\mathcal{O}_K.$$

Note that $\mathfrak{b}_1$ and $\mathfrak{b}_2$ strictly contain $\mathfrak{a}$ because $\alpha_1, \alpha_2 \notin \mathfrak{a}$. Moreover, $\mathfrak{b}_1\mathfrak{b}_2 \subseteq \mathfrak{a}$ because

$$\mathfrak{b}_1\mathfrak{b}_2 = (\mathfrak{a} + \alpha_1\mathcal{O}_K)(\mathfrak{a} + \alpha_2\mathcal{O}_K) = \mathfrak{a}^2 + \alpha_1\mathcal{O}_K + \alpha_2\mathcal{O}_K + \alpha_1\alpha_2\mathcal{O}_K,$$

and $\alpha_1\alpha_2 \in \mathfrak{a}$. We now show that either $\mathfrak{b}_1$ or $\mathfrak{b}_2$ belongs to $\mathcal{S}$. Suppose otherwise. Then there exist prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ and $\mathfrak{q}_1, \ldots, \mathfrak{q}_\ell$ such that

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \mathfrak{b}_1 \quad \text{and} \quad \mathfrak{q}_1 \cdots \mathfrak{q}_\ell \subseteq \mathfrak{b}_2.$$

But then

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k\mathfrak{q}_1 \cdots \mathfrak{q}_\ell \subseteq \beta_1\beta_2 \subseteq \mathfrak{a},$$

which contradicts the fact that $\mathfrak{a}$ is in $\mathcal{S}$. Hence $\mathfrak{b}_1$ or $\mathfrak{b}_2$ belongs to $\mathcal{S}$. In total, we have shown that if $\mathfrak{a} \in \mathcal{S}$, then there exists an integral ideal $\mathfrak{a}_1 \in \mathcal{S}$ strictly larger than $\mathfrak{a}$. Thus we obtain a strictly increasing infinite sequence of integral ideals in $\mathcal{S}$:

$$\mathfrak{a} \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots .$$

Taking the norm, we obtain a strictly decreasing sequence of positive integers:

$$\mathrm{N}(\mathfrak{a}) > \mathrm{N}(\mathfrak{a}_1) > \mathrm{N}(\mathfrak{a}_2) > \cdots .$$

This is impossible since the norm of an integral ideal is a positive integer. Hence $\mathcal{S}$ is empty and the claim follows. □

In order to obtain the reverse containment in Lemma 2.4.1, we need to do more work. Precisely, we want to show that every integral ideal factors into a product of prime integral ideals. To accomplish this, we will construct a group containing the ideals. Unfortunately, ideals are not invertible and so we need to work in a slightly more general setting. First observe that an integral ideal $\mathfrak{a}$ is just an $\mathcal{O}_K$-submodule of $\mathcal{O}_K$. Moreover, it is a finitely generated $\mathcal{O}_K$-submodule of K by Theorem 2.2.1. We say $\mathfrak{f}$ is a **fractional ideal** of K if $\mathfrak{f}$ a nonzero finitely generated $\mathcal{O}_K$-submodule of K. In particular, all integral ideals are fractional ideals. Now let $\kappa_1, \ldots, \kappa_r \in K$ be generators for the fractional ideal $\mathfrak{f}$. Since K is the field of fractions of $\mathcal{O}_K$ by Proposition 2.1.2, $\kappa_i = \frac{\alpha_i}{\delta_i}$ with $\alpha_i, \delta_i \in \mathcal{O}_K$ and where δ_i is nonzero for $1 \le i \le r$. Setting $\delta = \delta_1 \cdots \delta_r$, we have that $\delta\kappa_i \in \mathcal{O}_K$ for all i and hence $\delta\mathfrak{f}$ is an integral ideal. Conversely, if there exists some nonzero $\delta \in \mathcal{O}_K$ such that $\delta\mathfrak{f}$ is an integral ideal then $\mathfrak{f}$ is a fractional ideal because $\mathfrak{a}$ is a finitely generated $\mathcal{O}_K$-submodule of K and hence $\mathfrak{f}$ is too. Thus for any fractional ideal $\mathfrak{f}$, there exists a nonzero $\delta \in \mathcal{O}_K$ and an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{f} = \frac{1}{\delta}\mathfrak{a}.$$

Every fractional ideal is of this form and integral ideals are precisely those for which $\delta = 1$. In particular, since $\mathfrak{a}$ is a free abelian group of rank n, we see that $\mathfrak{f}$ is a free abelian group of rank n as well. However, $\mathfrak{f}$ is not necessarily a subgroup of $\mathcal{O}_K$. Now let $\mathfrak{p}$ be a prime integral ideal. We define $\mathfrak{p}^{-1}$ by

$$\mathfrak{p}^{-1} = \{\kappa \in K : \kappa\mathfrak{p} \subseteq \mathcal{O}_K\}.$$

It turns out that $\mathfrak{p}^{-1}$ is a fractional ideal. Indeed, since $\mathfrak{p}$ is an integral ideal there exists a nonzero $\alpha \in \mathfrak{p}$. By definition of $\mathfrak{p}^{-1}$, we have that $\alpha\mathfrak{p}^{-1} \subseteq \mathcal{O}_K$. Hence $\alpha\mathfrak{p}^{-1}$ is an integral ideal and therefore $\mathfrak{p}^{-1}$ is a fractional ideal. Unlike integral ideals, $1 \in \mathfrak{p}^{-1}$ so that $\mathfrak{p}^{-1}$ contains units. The following proposition proves a stronger version of this and more:

Lemma 2.4.2. *Let K be a number field and $\mathfrak{p}$ be a prime integral ideal. Then the following hold:*

(i)

$$\mathcal{O}_K \subset \mathfrak{p}^{-1}.$$

(ii)

$$\mathfrak{p}^{-1}\mathfrak{p} = \mathcal{O}_K.$$

Proof. We will prove the latter two statement separately:

(i) Clearly $\mathcal{O}_K \subseteq \mathfrak{p}^{-1}$ so it suffices to show that $\mathfrak{p}^{-1}$ contains a nonzero element which is not an algebraic integer. Again, let $\alpha \in \mathfrak{p}$ be nonzero. By Lemma 2.4.1 let $k \geq 1$ be the minimal integer such that there exist prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ with

$$\mathfrak{p}_1 \cdots \mathfrak{p}_k \subseteq \alpha\mathcal{O}_K.$$

As $\alpha \in \mathfrak{p}$, we have $\alpha\mathcal{O}_K \subseteq \mathfrak{p}$. Since $\mathfrak{p}$ is prime, there must be some i with $1 \leq i \leq k$ such that $\mathfrak{p}_i \subseteq \mathfrak{p}$. Without loss of generality, we may assume $\mathfrak{p}_1 \subseteq \mathfrak{p}$. But by Proposition 2.4.2 prime integral ideals are maximal and thus $\mathfrak{p}_1 = \mathfrak{p}$. Moreover, since k is minimal we must have

$$\mathfrak{p}_2 \cdots \mathfrak{p}_k \not\subseteq \alpha\mathcal{O}_K.$$

Hence there exists a nonzero $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_k$ with $\beta \notin \alpha\mathcal{O}_K$. We will now show that $\beta\alpha^{-1}$ is a nonzero element in $\mathfrak{p}^{-1}$ that is not an algebraic integer. Clearly $\beta\alpha^{-1}$ is nonzero. Since $\mathfrak{p}_1 = \mathfrak{p}$, what we have previously shown implies $\beta\mathfrak{p} \subseteq \alpha\mathcal{O}_K$ and hence $\beta\alpha^{-1}\mathfrak{p} \in \mathcal{O}_K$ which means $\beta\alpha^{-1} \in \mathfrak{p}^{-1}$. But as $\beta \notin \alpha\mathcal{O}_K$, we also have $\beta\alpha^{-1} \notin \mathcal{O}_K$ so that $\beta\alpha^{-1}$ is not an algebraic integer. This proves (i).

(ii) By (i) and the definition of $\mathfrak{p}^{-1}$, we have $\mathfrak{p} \subseteq \mathfrak{p}^{-1}\mathfrak{p} \subseteq \mathcal{O}$. Since $\mathfrak{p}$ is maximal by Proposition 2.4.2, it follows that $\mathfrak{p}^{-1}\mathfrak{p}$ is either $\mathfrak{p}$ or $\mathcal{O}_K$. So it suffices to show that the first case cannot hold. Assume by contradiction that $\mathfrak{p}^{-1}\mathfrak{p} = \mathfrak{p}$. Let $\{\omega_1, \ldots, \omega_r\}$ be a set of generators for $\mathfrak{p}$ and let $\alpha \in \mathfrak{p}^{-1}$ be a nonzero element that is not an algebraic integer which exists by (i). Then $\alpha\omega_i \in \mathfrak{p}^{-1}\mathfrak{p}$ for $1 \leq i \leq r$ and hence $\alpha\mathfrak{p} \subseteq \mathfrak{p}^{-1}\mathfrak{p}$. By our assumption, this further implies that $\alpha\mathfrak{p} \subseteq \mathfrak{p}$. But then

$$\alpha\omega_i = \sum_{1 \leq j \leq r} \alpha_{i,j}\omega_j,$$

with $\alpha_{i,j} \in \mathcal{O}_K$ for $1 \leq i, j \leq r$. We can rewrite this as,

$$(\alpha - \alpha_{i,i})\omega_i - \sum_{\substack{1 \leq j \leq r \\ j \neq i}} \alpha_{i,j}\omega_j = 0,$$

for all i. These r equations are equivalent to the identity

$$\begin{pmatrix} \alpha - \alpha_{1,1} & \alpha_{1,2} & \cdots & -\alpha_{1,r} \\ -\alpha_{2,1} & \alpha - \alpha_{2,2} & & \\ \vdots & & \ddots & \\ -\alpha_{r,1} & & & \alpha - \alpha_{r,r} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_r \end{pmatrix} = \mathbf{0}.$$

Thus the determinant of the matrix on the left-hand side must be zero. But this means α is a root of the characteristic polynomial $\det(xI - (\alpha_{i,j}))$ which is a monic polynomial with coefficients $\mathcal{O}_K$. By Proposition 2.2.2, α is an algebraic integer which is a contradiction. Thus $\mathfrak{p}^{-1}\mathfrak{p} = \mathcal{O}_K$. $\square$

Let I_K denote the collection of fractional ideals of K. We call I_K the **ideal group** of K. The following theorem shows that I_K is indeed a group:

Theorem 2.4.1. *Let K be a number field. Then I_K is an abelian group with identity $\mathcal{O}_K$.*

Proof. Since fractional ideals are finitely generated $\mathcal{O}_K$-submodules of K, the product of fractional ideals is a fractional ideal. It is also clear that I_K is abelian if it is a group. Moreover, $\mathcal{O}_K$ is the identity since every every fractional ideal is a finitely generated $\mathcal{O}_K$-submodule of K. It now suffices to show that every fractional ideal $\mathfrak{f}$ has an inverse in I_K. By Lemma 2.4.2 (ii), the prime integral ideal $\mathfrak{p}$ has inverse $\mathfrak{p}^{-1}$. We now show that any integral ideal that is not prime has an inverse. Suppose by contradiction that $\mathfrak{a}$ is such an integral ideal with $N(\mathfrak{a})$ minimal. By Proposition 2.4.2, there exists a prime integral ideal $\mathfrak{p}$ such that $\mathfrak{a} \subset \mathfrak{p}$. This fact together with Lemma 2.4.2 (i) implies that

$$\mathfrak{a} \subseteq \mathfrak{p}^{-1}\mathfrak{a} \subseteq \mathfrak{p}^{-1}\mathfrak{p} = \mathcal{O}_K.$$

We now claim $\mathfrak{a} \subset \mathfrak{p}^{-1}\mathfrak{a}$. If not, $\mathfrak{a} = \mathfrak{p}^{-1}\mathfrak{a}$. Let $\{\omega_1, \ldots, \omega_r\}$ be a set of generators for $\mathfrak{a}$. By Lemma 2.4.2, let $\alpha \in \mathfrak{p}^{-1}$ be a nonzero element that is not an algebraic integer. Then $\alpha\omega_i \in \mathfrak{p}^{-1}\mathfrak{a}$ for $1 \le i \le r$ and hence $\alpha\mathfrak{a} \subseteq \mathfrak{p}^{-1}\mathfrak{a}$. By our assumption, this further implies that $\alpha\mathfrak{p} \subseteq \mathfrak{a}$. But then

$$\alpha\omega_i = \sum_{1 \le j \le r} \alpha_{i,j}\omega_j,$$

with $\alpha_{i,j} \in \mathcal{O}_K$ for $1 \le i, j \le r$. We can rewrite this as,

$$(\alpha - \alpha_{i,i})\omega_i - \sum_{\substack{1 \le j \le r \\ j \ne i}} \alpha_{i,j}\omega_j = 0,$$

for all i. These r equations are equivalent to the identity

$$\begin{pmatrix} \alpha - \alpha_{1,1} & \alpha_{1,2} & \cdots & -\alpha_{1,r} \\ -\alpha_{2,1} & \alpha - \alpha_{2,2} & & \\ \vdots & & \ddots & \\ -\alpha_{r,1} & & & \alpha - \alpha_{r,r} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_r \end{pmatrix} = \mathbf{0}.$$

Thus the determinant of the matrix on the left-hand side must be zero. But this means α is a root of the characteristic polynomial $\det(xI - (\alpha_{i,j}))$ which is a monic polynomial with coefficients $\mathcal{O}_K$. By Proposition 2.2.2, α is an algebraic integer which is a contradiction. Thus $\mathfrak{a} \subset \mathfrak{p}^{-1}\mathfrak{a}$. But then $\mathrm{N}(\mathfrak{p}^{-1}\mathfrak{a}) < \mathrm{N}(\mathfrak{a})$ and by the minimality of $\mathrm{N}(\mathfrak{a})$ it follows that the fractional ideal $\mathfrak{p}^{-1}\mathfrak{a}$ is invertible. Let $\mathfrak{b}$ be its inverse. Then $\mathfrak{b}\mathfrak{p}^{-1}\mathfrak{a} = \mathcal{O}_K$ and thus $\mathfrak{a}$ is invertible with inverse $\mathfrak{b}\mathfrak{p}^{-1}$. This is a contradiction, so we conclude that every integral ideal is invertible. We now show that every fractional ideal $\mathfrak{f}$ is invertible. Since $\mathfrak{f}$ is a fractional ideal, there exists a nonzero $\delta \in \mathcal{O}_K$ and an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{f} = \frac{1}{\delta}\mathfrak{a}.$$

As $\mathfrak{a}$ is invertible, $\delta\mathfrak{a}^{-1}$ is the inverse of $\mathfrak{f}$. This completes the proof. □

We can also deduce the explicit form for the inverse of any fractional ideal:

Proposition 2.4.3. *Let K be a number field and let $\mathfrak{f}$ be a fractional ideal. Then*

$$\mathfrak{f}^{-1} = \{\kappa \in K : \kappa\mathfrak{f} \subseteq \mathcal{O}_K\}.$$

In particular, $\mathcal{O}_K \subseteq \mathfrak{f}$ if and only if $\mathfrak{f}^{-1}$ is an integral ideal.

Proof. Since $\mathfrak{f}\mathfrak{f}^{-1} = \mathcal{O}_K$, we have

$$\{\kappa \in K : \kappa\mathfrak{f} \subseteq \mathcal{O}_K\} = \{\kappa \in K : \kappa\mathcal{O}_K \subseteq \mathfrak{f}^{-1}\} = \mathfrak{f}^{-1}.$$

This proves the first statement. For the second statement, if $\mathcal{O}_K \subseteq \mathfrak{f}$ then multiplying by $\mathfrak{f}^{-1}$ shows $\mathfrak{f}^{-1} \subseteq \mathcal{O}_K$ and hence $\mathfrak{f}^{-1}$ is an integral ideal. Running this argument backwards by multiplying by $\mathfrak{f}$ proves the converse. □

Now that we have proved that the ideal group I_K of K is indeed a group, we can shown that every integral ideal factors uniquely into a product of prime integral ideals (up to reordering of the factors):

Theorem 2.4.2. *Let K be a number field. Then for every integral ideal $\mathfrak{a}$ there exist prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ such that $\mathfrak{a}$ factors as*

$$\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_k.$$

Moreover, this factorization is unique up to reordering of the factors.

Proof. We first prove existence and then uniqueness. For existence, suppose to the contrary that $\mathfrak{a}$ is an integral ideal that does not factor into a product of prime integral ideals and $\mathfrak{a}$ is maximal among all such integral ideals. Necessarily $\mathfrak{a}$ is not prime and by Proposition 2.4.2 there is some prime integral idea $\mathfrak{p}_1$ for which $\mathfrak{a} \subset \mathfrak{p}_1$. Then by Lemma 2.4.2 (ii), we have $\mathfrak{p}_1^{-1}\mathfrak{a} \subset \mathcal{O}_K$ so that $\mathfrak{p}_1^{-1}\mathfrak{a}$ is also an integral ideal. Also, Lemma 2.4.2 (i) implies $\mathfrak{a} \subseteq \mathfrak{a}\mathfrak{p}_1^{-1}$. Actually, $\mathfrak{a} \subset \mathfrak{a}\mathfrak{p}_1^{-1}$ for otherwise $\mathfrak{a} = \mathfrak{a}\mathfrak{p}_1^{-1}$ and hence $\mathfrak{p}_1^{-1} = O_K$ which is impossible because $\mathfrak{p}_1$ is proper. By maximality, $\mathfrak{a}\mathfrak{p}_1^{-1}$ factors into a product of prime integral ideals. That is, there exist prime integral ideals $\mathfrak{p}_2, \ldots, \mathfrak{p}_k$ such that

$$\mathfrak{a}\mathfrak{p}_1^{-1} = \mathfrak{p}_2, \ldots, \mathfrak{p}_k.$$

Hence

$$\mathfrak{a} = \mathfrak{p}_1, \ldots, \mathfrak{p}_k,$$

so that $\mathfrak{a}$ factors into a product of prime integral ideals which is a contradiction. This proves existence of such a factorization. Now we prove uniqueness. Suppose that $\mathfrak{a}$ admits factorizations

$$\mathfrak{a} = \mathfrak{p}_1, \ldots, \mathfrak{p}_k \quad \text{and} \quad \mathfrak{a} = \mathfrak{q}_1, \ldots, \mathfrak{q}_\ell,$$

for prime integral ideals $\mathfrak{p}_i$ and $\mathfrak{q}_j$ with $1 \le i \le k$ and $1 \le j \le \ell$. Since $\mathfrak{p}_1$ is prime, there is some j for which $\mathfrak{q}_j \subseteq \mathfrak{p}_1$. Without loss of generality, we may assume $\mathfrak{q}_1 \subseteq \mathfrak{p}_1$ and Proposition 2.4.2 we have that $\mathfrak{q}_1 = \mathfrak{p}_1$. Then

$$\mathfrak{p}_2, \ldots, \mathfrak{p}_k = \mathfrak{q}_2, \ldots, \mathfrak{q}_\ell.$$

Repeating this process, we see that $k = \ell$ and $\mathfrak{q}_i = \mathfrak{p}_i$ for all i. This proves uniqueness of the factorization. □

By Theorem 2.4.2, for any integral ideal $\mathfrak{a}$ there exist distinct prime integral ideal $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{a}$ admits a factorization

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r},$$

with $e_i \ge 1$ for all i, called the **prime factorization** of $\mathfrak{a}$ with **prime factors** $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$. Just as it is common to suppress the fundamental theorem of arithmetic and just state the prime factorization of an integer, we suppress Theorem 2.4.2 and state the prime factorization of an integral ideal. Also, as a near immediate corollary of Theorem 2.4.2, all fractional ideal admits a factorization into a product of prime integral ideals and their inverses (up to reordering of the factors):

Corollary 2.4.1. *Let K be a number field. Then for every fractional ideal $\mathfrak{f}$ there exist prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_k$ and $\mathfrak{q}_1, \ldots, \mathfrak{q}_\ell$ such that $\mathfrak{f}$ factors as*

$$\mathfrak{f} = \mathfrak{p}_1 \cdots \mathfrak{p}_k \mathfrak{q}_1^{-1}, \ldots, \mathfrak{q}_\ell^{-1}.$$

Moreover, this factorization is unique up to reordering of the factors.

Proof. If $\mathfrak{f}$ is a fractional ideal, then there exists a nonzero $\delta \in \mathcal{O}_K$ and an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{f} = \frac{1}{\delta}\mathfrak{a}.$$

In particular, $\mathfrak{a}$ and $\delta\mathcal{O}_K$ are integral ideals such that $\delta\mathcal{O}_K\mathfrak{f} = \mathfrak{a}$. By Theorem 2.4.2, $\mathfrak{a}$ and $\delta\mathcal{O}_K$ admit unique factorizations

$$\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_k \quad \text{and} \quad \delta\mathcal{O}_K = \mathfrak{q}_1, \ldots, \mathfrak{q}_\ell,$$

up to reordering of the factors. Hence

$$\mathfrak{q}_1, \ldots, \mathfrak{q}_\ell\mathfrak{f} = \mathfrak{p}_1 \cdots \mathfrak{p}_k,$$

which is equivalent to the factorization for $\mathfrak{f}$. □

We will now discuss applications of the Chinese remainder theorem in the context of integral ideals. With it we can prove some interesting results. Suppose $\mathfrak{a}$ is an integral ideal with prime factorization

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}.$$

Then the integral ideals $\mathfrak{p}_1^{e_1}, \ldots, \mathfrak{p}_r^{e_r}$ are pairwise relatively prime so that the Chinese remainder theorem gives an isomorphism

$$\mathcal{O}_K/\mathfrak{a} \cong \bigoplus_{1 \le i \le r} \mathcal{O}_K/\mathfrak{p}_i^{e_i}.$$

In particular, for any $\alpha_i \in \mathcal{O}_K$ for all i, there exists a unique $\alpha \in \mathcal{O}_K$ such that

$$\alpha \equiv \alpha_i \pmod{\mathfrak{p}_i^{e_i}},$$

for all i. We can now show that any fractional ideal is generated by at most two elements:

Corollary 2.4.2. *Let K be a number field. Then any fractional ideal $\mathfrak{f}$ is generated by at most two elements.*

Proof. We first prove the claim for an integral ideal $\mathfrak{a}$. Let $\alpha \in \mathfrak{a}$ be nonzero and let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be the prime factors of $\alpha\mathcal{O}_K$. As $\alpha\mathcal{O}_K \subseteq \mathfrak{a}$, the prime factorization of $\mathfrak{a}$ is

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r},$$

with $e_i \ge 0$ for $1 \le i \le r$. By uniqueness of the prime factorization of integral ideals, $\mathfrak{p}_i^{e_i+1} \subset \mathfrak{p}_i^{e_i}$ for all i. Thus there exist nonzero $\beta_i \in \mathfrak{p}_i^{e_i} - \mathfrak{p}_i^{e_i+1}$ for all i. Since $\mathfrak{p}_1^{e_1+1}, \ldots, \mathfrak{p}_r^{e_r+1}$ are pairwise relatively prime, the Chinese remainder theorem implies that there exists $\beta \in \mathcal{O}_K$ with

$$b \equiv \beta_i \pmod{\mathfrak{p}_i^{e_i+1}},$$

for all i. As $\beta_i \in \mathfrak{p}_i^{e_i}$ for all i, we have $\beta \in \mathfrak{a}$ and hence $\beta\mathcal{O}_K \subseteq \mathfrak{a}$. But as $\beta \notin \mathfrak{p}_i^{e_i+1}$ for all i, we see that $\beta\mathcal{O}_K\mathfrak{a}^{-1}$ is necessarily an integral ideal relatively prime to $\alpha\mathcal{O}_K$. This means

$$\beta\mathcal{O}_K\mathfrak{a}^{-1} + \alpha\mathcal{O}_K = \mathcal{O}_K,$$

and hence

$$\beta\mathcal{O}_K + \alpha\mathfrak{a} = \mathfrak{a}.$$

But as $\alpha, \beta \in \mathfrak{a}$, we have $\beta\mathcal{O}_K + \alpha\mathfrak{a} \subseteq \beta\mathcal{O}_K + \alpha\mathcal{O}_K \subseteq \mathfrak{a}$ and so

$$\beta\mathcal{O}_K + \alpha\mathcal{O}_K = \mathfrak{a}.$$

This shows that $\mathfrak{a}$ is generated by at most two elements. Now suppose $\mathfrak{f}$ is a fractional ideal. Then there exists a nonzero $\delta \in \mathcal{O}_K$ and an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{f} = \frac{1}{\delta}\mathfrak{a}.$$

Since $\mathfrak{a}$ is generated by at most two elements, say α and β, we have

$$\mathfrak{f} = \frac{\alpha}{\delta}\mathcal{O}_K + \frac{\beta}{\delta}\mathcal{O}_K,$$

and so $\mathfrak{f}$ is also generated by at most two elements as well. $\square$

Corollary 2.4.2 shows that while the the ring of integers $\mathcal{O}_K$ of K may not be a principal ideal domain, it is not far off from one since we every integral ideal needs at most two generators. We will give a more refined interpretation of this when discussing quotients of the ideal group I_K. For now, we deduce some more properties of the norm of integral ideals and extend this notion to fractional ideals as well. We will need a useful proposition:

Proposition 2.4.4. *Let K be a number field. Then for any prime integral ideal $\mathfrak{p}$ and $n \geq 0$, we have an isomorphism*

$$\mathcal{O}_K/\mathfrak{p} \cong \mathfrak{p}^n/\mathfrak{p}^{n+1},$$

as $\mathcal{O}_K$-modules.

Proof. By the uniqueness of the factorization of integral ideals, there exists $\beta \in \mathfrak{p}^n - \mathfrak{p}^{n+1}$. Now consider the homomorphism

$$\phi : \mathcal{O}_K \to \mathfrak{p}^n/\mathfrak{p}^{n+1} \qquad \alpha \to \alpha\beta \pmod{\mathfrak{p}^{n+1}}.$$

By the first isomorphism theorem, it suffices to show $\ker\phi = \mathfrak{p}$ and that ϕ is surjective. Let us first show $\ker\phi = \mathfrak{p}$. As $\beta \in \mathfrak{p}^n$, it is obvious that $\mathfrak{p} \subseteq \ker\phi$. Conversely, suppose $\alpha \in \mathcal{O}_K$ is such that $\phi(\alpha) = 0$. Then $\alpha\beta \in \mathfrak{p}^{n+1}$, and as $\beta \in \mathfrak{p}^n - \mathfrak{p}^{n+1}$ we must have $\alpha \in \mathfrak{p}$. It follows that $\ker\phi = \mathfrak{p}$. We now show that ϕ is surjective. Let $\gamma \in \mathfrak{p}^n$ be a representative of a class in $\mathfrak{p}^n/\mathfrak{p}^{n+1}$. As $\beta \in \mathfrak{p}^n$, we have $\beta\mathcal{O}_K \subseteq \mathfrak{p}^n$. But since $\beta \notin \mathfrak{p}^{n+1}$, we see that $\beta\mathcal{O}_K\mathfrak{p}^{-n}$ is necessarily an integral ideal relatively prime to $\mathfrak{p}^{n+1}$. As $\mathfrak{p}^{n+1}$ and $\beta\mathcal{O}_K\mathfrak{p}^{-n}$ are relatively prime, the Chinese remainder theorem implies that we can find a unique $\alpha \in \mathcal{O}_K$ such that

$$\alpha \equiv \gamma \pmod{\mathfrak{p}^{n+1}} \quad \text{and} \quad \alpha \equiv 0 \pmod{\beta\mathcal{O}_K\mathfrak{p}^{-n}}.$$

The second condition implies $\alpha \in \beta\mathcal{O}_K\mathfrak{p}^{-n}$. As $\gamma \in \mathfrak{p}^n$ and α and γ differ by an element in $\mathfrak{p}^{n+1} \subset \mathfrak{p}^n$, we have that $\alpha \in \beta\mathcal{O}_K\mathfrak{p}^{-n} \cap \mathfrak{p}^n = \beta\mathcal{O}_K$ where the equality holds because the intersection of ideals is equal to their product if the ideals are relatively prime. Thus $\frac{\alpha}{\beta} \in \mathcal{O}_K$ and hence

$$\phi\left(\frac{\alpha}{\beta}\right) = \alpha \equiv \gamma \pmod{\mathfrak{p}^{n+1}}.$$

This shows ϕ is surjective completing the proof. $\square$

Now we can show that the norm of an integral ideal is completely multiplicative:

Proposition 2.4.5. *Let K be a number field and let $\mathfrak{a}$ and $\mathfrak{b}$ be integral ideals. Then*

$$\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b}).$$

Proof. First suppose $\mathfrak{a}$ and $\mathfrak{b}$ are relatively prime. Then the Chinese remainder theorem implies

$$\mathcal{O}_K/\mathfrak{a}\mathfrak{b} \cong \mathcal{O}_K/\mathfrak{a} \oplus \mathcal{O}_K/\mathfrak{b},$$

and hence $|\mathcal{O}_K/\mathfrak{a}\mathfrak{b}| = |\mathcal{O}_K/\mathfrak{a}||\mathcal{O}_K/\mathfrak{b}|$ so that $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$. It now suffices to show $\mathrm{N}(\mathfrak{p}^n) = \mathrm{N}(\mathfrak{p})^n$ for all prime integral ideals $\mathfrak{p}$ and $n \geq 0$. We will prove this by induction. The base case is clear so assume that the claim holds for $n-1$. By the third isomorphism theorem, we have

$$\mathcal{O}_K/\mathfrak{p}^{n-1} \cong (\mathcal{O}_K/\mathfrak{p}^n)/(\mathfrak{p}^{n-1}/\mathfrak{p}^n).$$

Using Proposition 2.4.4, it follows that

$$|\mathcal{O}_K/\mathfrak{p}^{n-1}| = \frac{|\mathcal{O}_K/\mathfrak{p}^n|}{|\mathfrak{p}^{n-1}/\mathfrak{p}^n|} = \frac{|\mathcal{O}_K/\mathfrak{p}^n|}{|\mathcal{O}_K/\mathfrak{p}|}.$$

Thus $\mathrm{N}(\mathfrak{p}^n) = \mathrm{N}(\mathfrak{p}^{n-1})\,\mathrm{N}(\mathfrak{p})$ and our induction hypothesis implies $\mathrm{N}(\mathfrak{p}^n) = \mathrm{N}(\mathfrak{p})^n$ as desired. □

Note that by Proposition 2.4.5, the norm is a homomorphism from the set of integral ideals into $\mathbb{Z}_{\geq 1}$. As last we can extend the norm to fractional ideals. Let $\mathfrak{f}$ be a fractional ideal. By Corollary 2.4.1, there exist unique integral ideals $\mathfrak{a}$ and $\mathfrak{b}$ such that

$$\mathfrak{f} = \mathfrak{a}\mathfrak{b}^{-1}.$$

For any fractional ideal $\mathfrak{f}$, we define its **norm** $\mathrm{N}(\mathfrak{f})$ by

$$\mathrm{N}(\mathfrak{f}) = \frac{\mathrm{N}(\mathfrak{a})}{\mathrm{N}(\mathfrak{b})}.$$

Then we have a homomorphism

$$\mathrm{N} : I_K \to \mathbb{Q}^* \qquad \mathfrak{f} \mapsto \mathrm{N}(\mathfrak{f}).$$

2.5 Lattices & The Different

Let K be a number field of degree n. By Lemma 2.3.1, there is a nondegenerate symmetric bilinear form on K given by

$$\mathrm{Tr} : K \times K \to \mathbb{Q} \qquad (\kappa, \lambda) \to \mathrm{Tr}(\kappa\lambda).$$

We call this bilinear form the **trace form** on K. The trace form makes K into a nondegenerate inner product space over $\mathbb{Q}$. As any fractional ideal is a free abelian group of rank n, fractional ideals are lattices in K as a vector space over $\mathbb{Q}$. In particular, integral ideals are also lattice and thus $\mathcal{O}_K$ is a lattice. For a fractional ideal $\mathfrak{f}$, note that the dual lattice $\mathfrak{f}^\vee$ is

$$\mathfrak{f}^\vee = \{\kappa \in K : \mathrm{Tr}(\kappa\alpha) \in \mathbb{Z} \text{ for all } \alpha \in \mathfrak{f}\}.$$

The following proposition shows that the dual lattice $\mathfrak{f}^\vee$ is also a fractional ideal:

Proposition 2.5.1. *Let K be a number field and $\mathfrak{f}$ be a fractional ideal. Then $\mathfrak{f}^\vee$ is a fractional ideal and*

$$\mathfrak{f}^\vee = \mathfrak{f}^{-1}\mathcal{O}_K^\vee.$$

Proof. By Proposition 1.5.1, $\mathfrak{f}$ is a finitely generated $\mathbb{Z}$-module. Therefore it is a finitely generated $\mathcal{O}_K$-submodule of K if it is preserved under multiplication by $\mathcal{O}_K$. Let $\alpha \in \mathcal{O}_K$ and $\beta \in \mathfrak{f}^\vee$. Then we must show $\alpha\beta \in \mathfrak{f}^\vee$. To see this, observe that $\mathrm{Tr}(\alpha\beta\mathfrak{f}) \subseteq \mathrm{Tr}(\beta\mathfrak{f}) \subseteq \mathbb{Z}$ by Corollary 2.2.1 since $\alpha\mathfrak{f} \subseteq \mathfrak{f}$ and $\beta \in \mathfrak{f}^\vee$. Therefore $\alpha\beta \in \mathfrak{f}^\vee$ and hence $\mathfrak{f}^\vee$ is a fractional ideal proving the first statement. To prove the second we will show containment in both directions. For the forward containment, first suppose $\alpha \in \mathfrak{f}^\vee$ and $\beta \in \mathfrak{f}$. Then $\mathrm{Tr}(\alpha\beta\mathcal{O}_K) \subseteq \mathrm{Tr}(\alpha\mathfrak{f}) \subseteq \mathbb{Z}$ by Corollary 2.2.1 since $\beta\mathcal{O}_K \subseteq \mathfrak{f}$ and $\alpha \in \mathfrak{f}^\vee$. Hence $\alpha\beta \in \mathcal{O}_K^\vee$ so that $\mathfrak{f}^\vee\mathfrak{f} \subseteq \mathcal{O}_K^\vee$ and therefore $\mathfrak{f}^\vee \subseteq \mathfrak{f}^{-1}\mathcal{O}_K^\vee$. This proves the forward containment. For the reverse containment, suppose $\alpha \in \mathfrak{f}^{-1}$ and $\beta \in \mathcal{O}_K^\vee$. Then $\mathrm{Tr}(\alpha\beta\mathfrak{f}) \subseteq \mathrm{Tr}(\beta\mathcal{O}_K) \subseteq \mathbb{Z}$ by Corollary 2.2.1 since $\alpha\mathfrak{f} \subseteq \mathcal{O}_K$ and $\beta \in \mathcal{O}_K^\vee$. This shows $\alpha\beta \in \mathfrak{f}^\vee$ and hence $\mathfrak{f}^{-1}\mathcal{O}_K^\vee \subseteq \mathfrak{f}^\vee$ proving the reverse containment and completing the proof. □

We define the **different** $\mathfrak{D}$ of K by

$$\mathfrak{D}_K = (\mathcal{O}_K^\vee)^{-1}.$$

This is an integral ideal. Indeed, first note that $\mathcal{O}_K \subseteq \mathcal{O}_K^\vee$ by Corollary 2.2.1. It follows from Proposition 2.4.3 that $\mathcal{D}_K$ is an integral ideal and

$$\mathfrak{D}_K = \{\kappa \in K : \kappa\mathcal{O}_K^\vee \subseteq \mathcal{O}_K\}.$$

It turns out that the norm of the different is the absolute value of the discriminant:

Proposition 2.5.2. *Let K be an algebraic number field of degree n. Then we have an isomorphism*

$$\mathcal{O}_K/\mathfrak{D}_K \cong \mathcal{O}_K^\vee/\mathcal{O}_K,$$

as $\mathcal{O}_K$-modules. In particular

$$\mathrm{N}(\mathfrak{D}_K) = |\Delta_K|.$$

Proof. By Proposition 2.4.3, $\mathcal{O}_K \subseteq \mathfrak{D}_K^{-1}$. Then the second isomorphism theorem implies

$$\mathcal{O}_K/\mathfrak{D}_K \cong \mathfrak{D}_K^{-1}/\mathcal{O}_K \cong \mathcal{O}_K^\vee/\mathcal{O}_K,$$

which proves the first statement. For the second, this isomorphism shows that $\mathrm{N}(\mathfrak{D}_K) = |\mathcal{O}_K^\vee/\mathcal{O}_K|$. Now let $\{\alpha_1, \ldots, \alpha_n\}$ be an integral basis for $\mathcal{O}_K$. Then $\{\alpha_1^\vee, \ldots, \alpha_n^\vee\}$ is a basis for $\mathcal{O}_K^\vee$ and by definition of the dual basis we have

$$\alpha_i^\vee = \sum_{1 \le j \le n} \mathrm{Tr}(\alpha_i\alpha_j)\alpha_j.$$

But then the base change matrix from $\{\alpha_1, \ldots, \alpha_n\}$ to $\{\alpha_1^\vee, \ldots, \alpha_n^\vee\}$ is $(\mathrm{Tr}(\alpha_i\alpha_j))_{i,j}$. The claim follows by Proposition C.1.1 and the definition of Δ_K. □

We have already remarked that the different is an integral ideal and that $\mathcal{O}_K \subseteq \mathcal{O}_K^\vee$. Therefore we have an inclusion of lattices

$$\mathcal{D}_K \subseteq \mathcal{O}_K \subseteq \mathcal{O}_K^\vee.$$

What Proposition 2.5.2 shows is that each lattice has index $|\Delta_K|$ in the next one. In particular, $\mathcal{O}_K^\vee$ is strictly larger than $\mathcal{O}_K$ if and only if $|\Delta_K| \ge 2$. So we can think of the different $\mathcal{D}_K$ as a measure of the failure of $\mathcal{O}_K$ to be self-dual since $\mathrm{N}(\mathcal{D}_K) = 1$ if and only if $\mathcal{O}_K^\vee = \mathcal{O}_K$.

2.6 Ramification

We now discuss the factorization of prime integral ideals in number fields. First, we need to introduce the concept of prime integral ideals above primes. Let K be a number field and let $\mathfrak{p}$ be a prime integral ideal. Then $\mathfrak{p} \cap \mathbb{Z}$ is a prime integral ideal of $\mathbb{Q}$. Indeed, it is clear that $\mathfrak{p} \cap \mathbb{Z}$ is an integral ideal of $\mathbb{Q}$. It is proper because $1 \notin \mathfrak{p} \cap \mathbb{Z}$ as $\mathfrak{p}$ does not contain units. It is nonzero because $\mathcal{O}_K/\mathfrak{p}$ is a finite field by Proposition 2.4.2 and thus has characteristic dividing $\mathrm{N}(\mathfrak{p})$ by definition of the norm so that $\mathrm{N}(\mathfrak{p}) \in \mathfrak{p}$. But as $\mathrm{N}(\mathfrak{p})$ is also a positive integer, $\mathrm{N}(\mathfrak{p}) \in \mathfrak{p} \cap \mathbb{Z}$. To show that $\mathfrak{p} \cap \mathbb{Z}$ is prime, suppose a and b are integers such that $ab \in \mathfrak{p} \cap \mathbb{Z}$. Then $ab \in \mathfrak{p}$ and since $\mathfrak{p}$ is prime either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. But then $a \in \mathfrak{p} \cap \mathbb{Z}$ or $b \in \mathfrak{p} \cap \mathbb{Z}$ as desired. We have now shown that $\mathfrak{p} \cap \mathbb{Z}$ is a prime integral ideal of $\mathbb{Q}$. Hence

$$\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z},$$

for some prime integer p. Accordingly, we say that $\mathfrak{p}$ is **above** p, or equivalently, p is **below** $\mathfrak{p}$. Moreover, if $\mathfrak{p}$ is above p, then $\mathfrak{p}$ must be a prime factor of $p\mathcal{O}_K$. Indeed, $p\mathbb{Z} \subseteq \mathfrak{p}$ so that $p\mathcal{O}_K \subseteq \mathfrak{p}$ and then the fact $\mathfrak{p}$ is prime implies that some prime factor of $p\mathcal{O}_K$ is contained in $\mathfrak{p}$. Since prime integral ideals are maximal by Proposition 2.4.2, this prime factor must be $\mathfrak{p}$ itself. We illustrate these relations by the extension

$$\begin{array}{c} \mathfrak{p} \subset \mathcal{O}_K \subset K \\ \Big| \\ p\mathbb{Z} \subset \mathbb{Z} \subset \mathbb{Q} \end{array}$$

Since $\mathfrak{p}$ and $p\mathbb{Z}$ are maximal in $\mathcal{O}_K$ and $\mathbb{Z}$ respectively (by Proposition 2.4.2), we have the residue fields $\mathcal{O}_K/\mathfrak{p}$ and $\mathbb{F}_p$. It turns out that $\mathcal{O}_K/\mathfrak{p}$ is a finite dimensional vector space over $\mathbb{F}_p$. To see this, consider the homomorphism

$$\phi : \mathbb{Z} \to \mathcal{O}_K/\mathfrak{p} \qquad a \to a \pmod{\mathfrak{p}}.$$

Now $\ker\phi = \mathfrak{p} \cap \mathbb{Z}$ and hence $\ker\phi = p\mathbb{Z}$ since $\mathfrak{p}$ is above p. By the first isomorphism theorem, ϕ induces an injection $\phi : \mathbb{F}_p \to \mathcal{O}_K/\mathfrak{p}$. By Proposition 2.4.1, $\mathcal{O}_K/\mathfrak{p}$ is finite and thus a finite field containing $\mathbb{F}_p$. Necessarily $\mathcal{O}_K/\mathfrak{p}$ is a finite dimensional vector space over $\mathbb{F}_p$. Accordingly, we define the **inertia degree** $f_p(\mathfrak{p})$ of $\mathfrak{p}$ by

$$f_p(\mathfrak{p}) = [\mathcal{O}_K/\mathfrak{p} : \mathbb{F}_p].$$

That is, $f_p(\mathfrak{p})$ is the dimension of the residue field $\mathcal{O}_K/\mathfrak{p}$ as a vector space over $\mathbb{F}_p$. Then we have

$$\mathrm{N}(\mathfrak{p}) = |\mathcal{O}_K/\mathfrak{p}| = |\mathbb{F}_p|^{f_p(\mathfrak{p})} = p^{f_p(\mathfrak{p})}.$$

In particular, the norm of a prime integral ideal is a power of the prime below it. As we have already noted, $\mathfrak{p}$ is a prime factor of $p\mathcal{O}_K$. The **ramification index** $e_p(\mathfrak{p})$ of $\mathfrak{p}$ is the positive integer such that $p\mathcal{O}_K\mathfrak{p}^{-e_p(\mathfrak{p})}$ is relatively prime to $\mathfrak{p}$. If $p\mathcal{O}_K$ has prime factors $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$, then the prime factorization of $p\mathcal{O}_K$ is

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_p(\mathfrak{p}_1)} \cdots \mathfrak{p}_r^{e_p(\mathfrak{p}_r)}.$$

We say that p is **ramified** if $e_p(\mathfrak{p}_i) \geq 2$ for some i and **unramified** otherwise. In particular,

$$p\mathcal{O}_K = \mathfrak{p}_1 \cdots \mathfrak{p}_r,$$

if and only if p is unramified. We also say p is **split** if $p\mathcal{O}_K$ is not prime. The degree of a number field is connected to the inertia degree and ramification index via the following proposition:

Proposition 2.6.1. *Let K be a number field of degree n and let p be a prime. Suppose $p\mathcal{O}_K$ has prime factorization*

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_p(\mathfrak{p}_1)} \cdots \mathfrak{p}_r^{e_p(\mathfrak{p}_r)}.$$

Then

$$n = \sum_{1 \le i \le r} e_p(\mathfrak{p}_i) f_p(\mathfrak{p}_i).$$

Proof. Since p is an integer and K is of degree n, on the one hand

$$\mathrm{N}(p\mathcal{O}_K) = \mathrm{N}(p) = p^n.$$

On the other hand, complete multiplicativity of the norm by Proposition 2.4.5 implies

$$\mathrm{N}(p\mathcal{O}_K) = \mathrm{N}(\mathfrak{p}_1)^{e_p(\mathfrak{p}_1)} \cdots \mathrm{N}(\mathfrak{p}_r)^{e_p(\mathfrak{p}_r)} = p^{e_p(\mathfrak{p}_1) f_p(\mathfrak{p}_1)} \cdots p^{e_p(\mathfrak{p}_r) f_p(\mathfrak{p}_r)}.$$

Thus

$$p^n = p^{e_p(\mathfrak{p}_1) f_p(\mathfrak{p}_1)} \cdots p^{e_p(\mathfrak{p}_r) f_p(\mathfrak{p}_r)},$$

and the claim follows upon comparing exponents. $\square$

We now describe some special cases of how $p\mathcal{O}_K$ may factor. If $r = n$, we say p is **totally split** and so $e_p(\mathfrak{p}) = f_p(\mathfrak{p}) = 1$ for all $\mathfrak{p}$ above p by Proposition 2.6.1. Equivalently, p is totally split if and only if the number of prime integral ideal above p is equal to the degree of K. If $r = 1$, then there is exactly one prime integral ideal $\mathfrak{p}$ above p and so

$$p\mathcal{O}_K = \mathfrak{p}^{e_p(\mathfrak{p})}.$$

If $e_p(\mathfrak{p}) = 1$, so that p does not ramify, we say p is **inert** and so $f_p(\mathfrak{p}) = n$ by Proposition 2.6.1. Thus p is inert if and only if $p\mathcal{O}_K$ is prime. If $e_p(\mathfrak{p}) \ge 2$, then p ramifies and we say p is **totally ramified** if $e_p(\mathfrak{p}) = n$ so that $f_p(\mathfrak{p}) = 1$ by Proposition 2.6.1. Equivalently, p is totally ramified if and only if it is the power of a prime integral ideal with power equal to the degree of K. The ramification of primes is intimately connected to the discriminant of a number field as the following theorem shows:

Theorem 2.6.1. *Let K be a number field. Then p is ramified if and only if p divides $|\Delta_K|$.*

Proof. Let p be a prime and suppose $p\mathcal{O}_K$ has prime factorization

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_p(\mathfrak{p}_1)} \cdots \mathfrak{p}_r^{e_p(\mathfrak{p}_r)},$$

Now let $\{\alpha_1, \ldots, \alpha_n\}$ be an integral basis for K and let $\overline{\alpha} \in \mathcal{O}_K/p\mathcal{O}_K$ denote the reduction of $\alpha \in \mathcal{O}_K$ modulo p. Then $\{\overline{\alpha_1}, \ldots, \overline{\alpha_n}\}$ is a basis for $\mathcal{O}_K/p\mathcal{O}_K$ as a vector space over $\mathbb{F}_p$. Moreover, the matrix for $T_{\overline{\alpha}}$ is obtained from T_α by reducing the coefficients modulo p. These two facts together give

$$\Delta_K = \mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, \ldots, \alpha_n) \equiv \mathrm{disc}_{(\mathcal{O}_K/p\mathbb{Z})/\mathbb{F}_p}(\overline{\alpha_1}, \ldots, \overline{\alpha_n}) \pmod{p}.$$

Recall that $\mathrm{disc}_{(\mathcal{O}_K/p\mathbb{Z})/\mathbb{F}_p}(\overline{\alpha_1}, \ldots, \overline{\alpha_n})$ is an element of $\mathbb{F}_p$. Then as $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/p\mathcal{O}_K)$ is equivalent to $\mathrm{disc}_{(\mathcal{O}_K/p\mathbb{Z})/\mathbb{F}_p}(\overline{\alpha_1}, \ldots, \overline{\alpha_n})$ up to elements of $(\mathbb{F}_p^*)^2$, it must be the case that p divides $|\Delta_K|$ if and only if $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/p\mathcal{O}_K) = 0$. By the Chinese remainder theorem,

$$\mathcal{O}_K/p\mathcal{O}_K \cong \bigoplus_{1 \le i \le r} \mathcal{O}_K/\mathfrak{p}_i^{e_p(\mathfrak{p}_i)},$$

and so Proposition 2.3.1 further implies

$$\Delta_K = \prod_{1 \le i \le r} \mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}_i^{e_p(\mathfrak{p}_i)}).$$

Hence p divides $|\Delta_K|$ if and only if $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}_i^{e_p(\mathfrak{p}_i)}) = 0$ for some i. It is now sufficient to show that $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}) = 0$ for any prime integral ideal $\mathfrak{p}$ above p if and only if $e_p(\mathfrak{p}) \ge 2$. First suppose $e_p(\mathfrak{p}) \ge 2$. We will prove $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}) = 0$. By the uniqueness of the factorization of integral ideals, there exists a nonzero $\alpha_1 \in \mathfrak{p}^{e_p(\mathfrak{p})-1} - \mathfrak{p}^{e_p(\mathfrak{p})}$. Then $\alpha_1^2 \in \mathfrak{p}^{2(e_p(\mathfrak{p})-1)} \subseteq \mathfrak{p}^{e_p(\mathfrak{p})}$ because $e_p(\mathfrak{p}) \ge 2$. By construction, $\overline{\alpha_1} \in \mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}$ is nonzero and such that $\overline{\alpha_1}^2 = 0$. Since $\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}$ is an n dimensional vector space over $\mathbb{F}_p$, there exists a basis of the form $\{\overline{\alpha_1}, \ldots, \overline{\alpha_m}\}$. Now

$$\mathrm{Tr}_{(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})})/\mathbb{F}_p}(\overline{\alpha_1\alpha_j}) = 0,$$

for $1 \le j \le n$ because $T^2_{\overline{\alpha_1\alpha_i}}$ is the zero operator (as $\overline{\alpha_1}^2 = 0$) and hence all of its eigenvalues are zero. But then the first row of $(\mathrm{Tr}_{(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})})/\mathbb{F}_p}(\overline{\alpha_i\alpha_j}))_{i,j}$ is zero and hence $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}) = 0$. Now suppose $e_p(\mathfrak{p}) = 1$. We will prove $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}) \ne 0$. Recall that $\mathcal{O}_K/\mathfrak{p}$ is a field and a vector space over $\mathbb{F}_p$ of dimension $f_p(\mathfrak{p})$. Thus $(\mathcal{O}_K/\mathfrak{p})/\mathbb{F}_p$ it is a finite separable extension. Hence $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p})$ is nonzero by Proposition 2.3.2. We have now shown that $\mathrm{disc}_{\mathbb{F}_p}(\mathcal{O}_K/\mathfrak{p}^{e_p(\mathfrak{p})}) = 0$ if and only if $e_p(\mathfrak{p}) \ge 2$ which completes the proof. □

As an immediate corollary, we see that only finitely many primes can ramify:

Corollary 2.6.1. *Let K be a number field. Then finitely many primes ramify in K.*

Proof. There are only finitely many prime divisors of $|\Delta_K|$. Hence finitely many primes ramify by Theorem 2.6.1. □

There is also another important corollary which shows that the prime factors of $\mathfrak{D}_K$ correspond exactly to the ramified primes:

Corollary 2.6.2. *Let K be a number field. If $\mathfrak{p}$ is a prime factor of $\mathfrak{D}_K$, then $\mathfrak{p}$ lies over a ramified prime p.*

Proof. By Proposition 2.5.2 and Theorem 2.6.1, p is ramified if and only if $p \mid \mathrm{N}(\mathcal{D}_K)$. By multiplicativity of the norm and the prime factorization, a ramified prime p must divide $\mathrm{N}(\mathfrak{p})$ for some prime factor $\mathfrak{p}$ of $\mathcal{D}_K$. But then $\mathfrak{p}$ is a prime integral ideal above the ramified prime p. □

There is no general way to see how a prime p factors for an arbitrary number field K. However, in the case that the ring of integers is monogenic we can describe the factorization explicitly via the **Dedekind-Kummer theorem**:

Theorem 2.6.2 (Dedekind-Kummer theorem). *Let K be a monogenic number field where $\mathcal{O}_K = \mathbb{Z}[\alpha]$ for $\alpha \in \mathcal{O}_K$ and let p be a prime. Let $m_\alpha(x)$ be the minimal polynomial for α and let $\overline{m_\alpha}(x)$ be its reduction modulo p. Also let*

$$\overline{m_\alpha}(x) = \overline{m_r}(x)^{e_r} \cdots \overline{m_r}(x)^{e_r}$$

with $\overline{m_i}(x) \in \mathbb{F}_p[x]$ and $e_i \ge 0$, be the prime factorization of $\overline{m_\alpha}(x)$ in $\mathbb{F}_p[x]$. Let $m_i(x) \in \mathbb{Z}[x]$ be any lift of $\overline{m_i}(x)$ and set

$$\mathfrak{p}_i = p\mathcal{O}_K + m_i(\alpha)\mathcal{O}_K,$$

for all i. Then $\mathfrak{p}_i$ is a prime integral ideal for all i and

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r},$$

is the prime factorization of $p\mathcal{O}_K$.

Proof. Since $m_\alpha(x)$ is the minimal polynomial for α, we have an isomorphism $\mathbb{Z}[\alpha] \cong \mathbb{Z}[x]/m_\alpha(x)\mathbb{Z}[x]$ where the inverse isomorphism is given by evaluation at α. Then we have the chain of isomorphism

$$\mathcal{O}_K/p\mathcal{O}_K \cong (\mathbb{Z}[x]/m_\alpha(x)\mathbb{Z}[x])/(p(\mathbb{Z}[x]/m_\alpha(x)\mathbb{Z}[x])) \cong \mathbb{Z}[x]/(p\mathbb{Z}[x] + m_\alpha(x)\mathbb{Z}[x]) \cong \mathbb{F}_p[x]/\overline{m_\alpha}(x)\mathbb{F}_p[x],$$

where the second and third isomorphisms follow by taking $\mathbb{Z}[x]/(p\mathbb{Z}[x] + m_\alpha(x)\mathbb{Z}[x])$ and reducing $\mathbb{Z}[x]$ modulo $m_\alpha(x)$ or p respectively. Therefore the inverse isomorphism is given by sending any representative $\overline{f}(x)$ of a class in $\mathbb{F}_p[x]/\overline{m_\alpha}(x)\mathbb{F}_p[x]$ to a lift $f(x) \in \mathbb{Z}[x]$ and then to $\overline{f(\alpha)}$ where $\overline{f(\alpha)}$ is $f(\alpha)$ modulo $p\mathcal{O}_K$. Now set $A = \mathbb{F}_p[x]/\overline{m_\alpha}(x)\mathbb{F}_p[x]$. Then the Chinese remainder theorem gives an isomorphism

$$A \cong \bigoplus_{1 \le i \le r} \mathbb{F}_p[x]/\overline{m_i}(x)^{e_i}\mathbb{F}_p[x].$$

As $\overline{m_i}(x)$ is irreducible, $\overline{m_i}(x)\mathbb{F}_p[x]$ is maximal and hence $\mathbb{F}_p[x]/\overline{m_i}(x)\mathbb{F}_p[x]$ is a field. By the third isomorphism theorem, $\overline{m_i}(x)\mathbb{F}_p[x]/\overline{m_i}(x)^{e_i}\mathbb{F}_p[x]$ is a maximal ideal of $\mathbb{F}_p[x]/\overline{m_i}(x)^{e_i}\mathbb{F}_p[x]$. It follows that the maximal ideals of A are precisely $\overline{m_i}(x)A$ and we have an isomorphism

$$A/\overline{m_i}(x)A \cong \mathbb{F}_p[x]/\overline{m_i}(x)\mathbb{F}_p[x],$$

for all i. Via the isomorphism $\mathcal{O}_K/p\mathcal{O}_K \cong A$ described above, the maximal ideals of $\mathcal{O}_K/p\mathcal{O}_K$ are exactly $\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)$. We now show that the $\mathfrak{p}_i$ are prime. To see this, consider the surjective homomorphism

$$\pi : \mathcal{O}_K \to \mathcal{O}_K/p\mathcal{O}_K \qquad \alpha \to \alpha \pmod{p\mathcal{O}_K}.$$

Then the image of $\mathfrak{p}_i$ under π is $\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)$. As this ideal is maximal and hence prime, the preimage $\mathfrak{p}_i$ is prime too. Moreover, the $\mathfrak{p}_i$ are all distinct since the $\overline{m_i(\alpha)}\mathcal{O}_K/p\mathcal{O}_K$ are which are all distinct because the $\overline{m_i}(x)A$ are (using the isomorphism $\mathcal{O}_K/p\mathcal{O}_K \cong A$). In particular, they are also relatively prime. By construction, $\mathfrak{p}_i \subseteq p\mathcal{O}_K$ so that the $\mathfrak{p}_i$ are prime factors of $p\mathcal{O}_K$. These are the only prime factors of $p\mathcal{O}_K$ because the image of any prime integral ideal under π and contained in $p\mathcal{O}_K$ must be a maximal ideal of $\mathcal{O}_K/p\mathcal{O}_K$, by Proposition 2.4.2 and the fourth isomorphism theorem, and every maximal ideal is one of the $\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)$. Together, all of this means that $p\mathcal{O}_K$ admits the prime factorization

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_p(\mathfrak{p}_1)} \cdots \mathfrak{p}_r^{e_p(\mathfrak{p}_r)},$$

for some ramification indices $e_p(\mathfrak{p}_i)$ for all i. We will be done if we can show that the ramification indices satisfy $e_p(\mathfrak{p}_i) = e_i$. To accomplish this, observe that we have an isomorphism

$$\mathcal{O}_K/\mathfrak{p}_i \cong (\mathcal{O}_K/p\mathcal{O}_K)/(\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)) \cong \mathbb{F}_p[x]/\overline{m_i}(x)\mathbb{F}_p[x],$$

where the first isomorphism follow by taking $\mathcal{O}_K/\mathfrak{p}_i$ and reducing $\mathcal{O}_K$ modulo p and the second isomorphism follows from $\mathcal{O}_K/p\mathcal{O}_K \cong A$ and that the image of the maximal ideal $\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)$ under this isomorphisms is $\overline{m_i}(x)A$. Now $\mathbb{F}_p[x]/\overline{m_i}(x)\mathbb{F}_p[x]$ is a vector space over $\mathbb{F}_p$ (as it contains $\mathbb{F}_p$) of degree $\deg(\overline{m_i}(x))$. Hence the inertia degree $f_p(\mathfrak{p}_i)$ of $\mathfrak{p}_i$ satisfies $f_p(\mathfrak{p}_i) = \deg(\overline{m_i}(x))$. The ideal $\overline{m_i}(x)^{e_i}A$ under the isomorphism $A \cong \mathcal{O}_K/p\mathcal{O}_K$ is the ideal $\overline{m_i(\alpha)}^{e_i}(\mathcal{O}_K/p\mathcal{O}_K)$. As the image of $\mathfrak{p}_i$ under π is $\overline{m_i(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K)$, we have that $\mathfrak{p}_i^{e_i}$ is contained in the preimage of $\overline{m_i(\alpha)}^{e_i}(\mathcal{O}_K/p\mathcal{O}_K)$ under π. As $\overline{m_\alpha(\alpha)}(\mathcal{O}_K/p\mathcal{O}_K) = 0$ is the zero ideal, it follows that

$$p\mathcal{O}_K = \pi^{-1}(0) \supseteq \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}.$$

Since the $\mathfrak{p}_i$ are prime, we have $e_p(\mathfrak{p}_i) \le e_i$ for all i. By Proposition 2.6.1 then gives

$$n = \sum_{1 \le i \le r} e_p(\mathfrak{p}_i)f_p(\mathfrak{p}_i) \le \sum_{1 \le i \le r} e_i f_p(\mathfrak{p}_i) \le \sum_{1 \le i \le r} e_i \deg(\overline{m_i}(x)) \le n,$$

where the last equality follows by the prime factorization of $\overline{m_\alpha}(x)$ and that $\deg(\overline{m_\alpha}(x)) = \deg(m_\alpha(x))$ because $m_\alpha(x)$ is monic. This shows that $e_p(\mathfrak{p}_i) = e_i$ for all i which completes the proof. $\square$

Lastly, we want to show that the number of integral ideals of a given norm is relatively small. Indeed, let $a_K(m)$ denote the number of integral ideals of norm m. Because the norm is multiplicative so is $a_K(m)$. Moreover, we have the following result:

Proposition 2.6.2. *Let K be a number field of degree n. Then $a_K(m) \leq \sigma_0(m)^n$.*

Proof. Let $\mathfrak{a}$ be an integral ideal of norm m. First suppose $m = p^k$ for some prime p and $k \geq 0$. As there are at most n prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ above p with inertia degrees $f_p(\mathfrak{p}_1), \ldots, f_p(\mathfrak{p}_1)$ respectively, we have

$$\mathrm{N}(\mathfrak{a}) = p^{e_1 f_p(\mathfrak{p}_1)} \cdots p^{e_n f_{\mathfrak{p}_n}},$$

for some integers $0 \leq e_i \leq k$ for $1 \leq i \leq n$. Therefore the number of possibilities is equivalent to the number of solutions

$$e_1 f_p(\mathfrak{p}_1) + \cdots + e_n f_{\mathfrak{p}_n} = k,$$

which is at most $\sigma_0(p^k)^n = (k+1)^n$. This proves the claim in the case m is a prime power. By multiplicativity of $a_K(m)$ and the divisor function, it follows that the number of integral ideals of norm m is at most $\sigma_0(m)^n$ as desired. $\square$

2.7 Minkowski Space

Let K be number field of degree n and let σ be a $\mathbb{Q}$-embedding of K into $\overline{\mathbb{Q}}$. Then either σ is real or complex and if it is complex then it has a paired $\mathbb{Q}$-embedding $\overline{\sigma}$ given by the conjugate of σ. Accordingly, let r_1 and $2r_2$ be the number of real and complex $\mathbb{Q}$-embeddings respectively. We call the pair (r_1, r_2) the **signature** of K and it satisfies the the relation

$$n = r_1 + 2r_2.$$

Now set

$$K_{\mathbb{C}} = \prod_{\sigma} \mathbb{C},$$

where the product runs over all $\mathbb{Q}$-embeddings of K into $\overline{\mathbb{Q}}$. Moreover, $K_{\mathbb{C}}$ is a complex Hilbert space with respect to the Hermitian inner product $\langle \cdot, \cdot \rangle_{K_{\mathbb{C}}}$ which is the standard inner product on $\mathbb{C}^n$. We denote the associated Lebesgue measure by $d\lambda_{K_{\mathbb{C}}}$. Now there is a $\mathbb{Q}$-embedding

$$j : K \to K_{\mathbb{C}} \qquad \kappa = (\sigma(\kappa))_{\sigma},$$

given by the n distinct $\mathbb{Q}$-embeddings of K into $\overline{\mathbb{Q}}$. Consider the complex conjugation map

$$F : \mathbb{C} \to \mathbb{C} \qquad z \to \overline{z}.$$

This defines an automorphism

$$F : K_{\mathbb{C}} \to K_{\mathbb{C}} \qquad (z_{\sigma})_{\sigma} \to (\overline{z_{\overline{\sigma}}})_{\sigma},$$

and is clearly an involution. The inner product is also F-equivariant, since any $\mathbf{z}, \mathbf{w} \in K_{\mathbb{C}}$ satisfy

$$\langle F(\mathbf{z}), F(\mathbf{z}) \rangle_{K_{\mathbb{C}}} = \langle (\overline{z_{\overline{\sigma}}})_{\sigma}, (\overline{w_{\overline{\sigma}}})_{\sigma} \rangle_{K_{\mathbb{C}}} = \sum_{\sigma} \overline{z_{\overline{\sigma}} w_{\overline{\sigma}}} = \overline{\sum_{\sigma} z_{\sigma} w_{\sigma}} = F(\langle \mathbf{z}, \mathbf{w} \rangle_{K_{\mathbb{C}}}),$$

where in the third equality we have used the fact that the complex $\mathbb{Q}$-embeddings come in conjugate pairs. On $K_{\mathbb{C}}$ we also have a linear map

$$\mathrm{Tr}_{K_{\mathbb{C}}} : K_{\mathbb{C}} \to \mathbb{C} \qquad (z_\sigma)_\sigma \to \sum_\sigma z_\sigma.$$

The composition of j with $\mathrm{Tr}_{K_{\mathbb{C}}}$ is Tr since

$$\mathrm{Tr}_{K_{\mathbb{C}}}(j(\kappa)) = \mathrm{Tr}_{K_{\mathbb{C}}}((\sigma(\kappa))_\sigma) = \sum_\sigma \sigma(\kappa) = \mathrm{Tr}(\kappa),$$

where the last equality follows by Proposition 2.2.1. We now define the **Minkowski space** $K_{\mathbb{R}}$ of K by

$$K_{\mathbb{R}} = \{(z_\sigma)_\sigma \in K_{\mathbb{C}} : F((z_\sigma)_\sigma) = (z_\sigma)_\sigma\}.$$

In other words, $K_{\mathbb{R}}$ consists of all of the F-invariant points of $K_{\mathbb{C}}$. That is, $\mathbf{z} \in K_{\mathbb{R}}$ if and only if $F(\mathbf{z}) = \mathbf{z}$ or equivalently $z_{\overline{\sigma}} = \overline{z_\sigma}$ for all σ. In particular, $j(K) \subset K_{\mathbb{R}}$ because $\overline{\sigma}(\kappa) = \overline{\sigma(\kappa)}$ by definition of $\overline{\sigma}$. We denote the restriction of the inner product $\langle \cdot, \cdot \rangle_{K_{\mathbb{C}}}$ on $K_{\mathbb{C}}$ to $K_{\mathbb{R}}$ by $\langle \cdot, \cdot \rangle_{K_{\mathbb{R}}}$. This inner product turns $K_{\mathbb{R}}$ into a real Hilbert space. Indeed, for any $\mathbf{z}, \mathbf{w} \in K_{\mathbb{R}}$ the conjugate symmetry and F-equivariance of the inner product together give

$$\overline{\langle \mathbf{z}, \mathbf{w} \rangle} = F(\langle \mathbf{w}, \mathbf{z} \rangle) = \langle F(\mathbf{z}), F(\mathbf{w}) \rangle = \langle \mathbf{z}, \mathbf{w} \rangle,$$

so that $\langle \mathbf{z}, \mathbf{w} \rangle \in \mathbb{R}$ is real. We denote the restriction of the Lebesgue measure $d\lambda_{\mathbb{C}}$ to $K_{\mathbb{R}}$ by $d\lambda_{\mathbb{R}}$ which is also the Lebesgue measure associated to $\langle \cdot, \cdot \rangle_{K_{\mathbb{R}}}$. Lastly, we denote the restriction of the trace $\mathrm{Tr}_{K_{\mathbb{C}}}$ to $K_{\mathbb{R}}$ by $\mathrm{Tr}_{K_{\mathbb{R}}}$. As $j(K) \subset K_{\mathbb{R}}$, the composition of j with $\mathrm{Tr}_{K_{\mathbb{R}}}$ is Tr. We can now give a more explicit description of $K_{\mathbb{R}}$. To do this, write the real $\mathbb{Q}$-embeddings as ρ and representatives of the pairs of complex $\mathbb{Q}$-embeddings as τ. Because $K_{\mathbb{R}}$ consists of exactly the F-invariant points of $K_{\mathbb{C}}$, we have

$$K_{\mathbb{R}} = \{(z_\sigma)_\sigma \in K_{\mathbb{C}} : x_\rho \in \mathbb{R} \text{ and } x_{\overline{\tau}} = \overline{x_\tau} \text{ for all } \rho \text{ and } \tau\}.$$

We now describe an explicit isomorphism from the Minkowski space into $\mathbb{R}^n$:

Proposition 2.7.1. *Let K be a number field of degree n and signature (r_1, r_2). Also let σ run over the $\mathbb{Q}$-embeddings of K into $\overline{\mathbb{Q}}$, ρ run over all such real $\mathbb{Q}$-embeddings, and τ run over representatives of all such pairs of complex $\mathbb{Q}$-embeddings. Then there is an isomorphism*

$$K_{\mathbb{R}} \to \prod_\sigma \mathbb{R} \qquad z_\sigma \to x_\sigma = \begin{cases} z_\sigma & \text{if } \sigma = \rho, \\ \mathrm{Re}(z_\sigma) & \text{if } \sigma = \tau, \\ \mathrm{Im}(z_\sigma) & \text{if } \sigma = \overline{\tau}. \end{cases}$$

In particular, $K_{\mathbb{R}}$ is a n-dimensional vector space over $\mathbb{R}$. Moreover, the induced inner product $\langle \cdot, \cdot \rangle$ on $\prod_\sigma \mathbb{R}$ under this map is given by

$$\langle \mathbf{x}, \mathbf{x}' \rangle = \sum_\sigma \alpha_\sigma x_\sigma x'_\sigma,$$

for any $\mathbf{x}, \mathbf{x}' \in \prod_\sigma \mathbb{R}$ and where $\alpha_\sigma = 1, 2$ corresponding to if σ is real or complex.

Proof. This map is an isomorphism since it is invertible and linear in each component. Since there are n such $\mathbb{Q}$-embeddings σ we see that $K_{\mathbb{R}}$ is an n-dimensional vector space over $\mathbb{R}$. We will now prove the statement about the inner product. Let $(z_\sigma)_\sigma$ and $(z'_\sigma)_\sigma$ be elements of $K_\mathbb{R}$ and let $(x_\sigma)_\sigma$ and $(x'_\sigma)_\sigma$ be the corresponding elements in $\prod_\sigma \mathbb{R}$. If $\sigma = \rho$, then

$$x_\rho = z_\rho \quad \text{and} \quad x'_\rho = z'_\rho,$$

and thus

$$x_\rho x'_\rho = z_\rho \overline{z'_\rho}.$$

If $\sigma = \tau$, then

$$x_\tau = \mathrm{Re}(z_\tau), \quad x_{\overline{\tau}} = \mathrm{Im}(z_\tau), \quad x'_\tau = \mathrm{Re}(z'_\tau), \text{and} \quad x'_{\overline{\tau}} = \mathrm{Im}(z'_\tau),$$

and hence

$$2(x_\tau x'_\tau + x_{\overline{\tau}} x'_{\overline{\tau}}) = 2(\mathrm{Re}(z_\tau)\mathrm{Re}(z'_\tau) + \mathrm{Im}(z_\tau)\mathrm{Im}(z'_\tau)) = 2\mathrm{Re}(z_\tau \overline{z'_\tau}) = z_\tau \overline{z'_\tau} + z_{\overline{\tau}} \overline{z'_{\overline{\tau}}}.$$

This proves the claim about the inner product. □

The **canonical embedding** σ_K of K is defined by

$$\sigma_K : K \to \mathbb{R}^n \qquad \kappa \to (\rho_1(\kappa), \ldots, \rho_{r_1}(k), \mathrm{Re}(\tau_1(\kappa)), \mathrm{Im}(\tau_1(\kappa)), \ldots, \mathrm{Re}(\tau_{r_2}(\kappa)), \mathrm{Im}(\tau_{r_2}(\kappa))),$$

where $\rho_1, \ldots, \rho_{r_1}$ are the real $\mathbb{Q}$-embeddings and $\tau_1, \ldots, \tau_{r_2}$ are representatives of pairs of complex $\mathbb{Q}$-embeddings. The canonical embedding σ_K is then a $\mathbb{Q}$-embedding of K into $\mathbb{R}^n$ since it is the composition of the $\mathbb{Q}$-embedding j (whose image under K is in $K_\mathbb{R}$ as we have noted) and the isomorphism established by Proposition 2.7.1. It is also independent of the choice of representatives $\tau_1, \ldots, \tau_{r_2}$ since the complex embeddings occur in conjugate pairs. By convention, we will always order the components of the image of σ_K by all the real $\mathbb{Q}$-embeddings and then the real and imaginary parts of the complex $\mathbb{Q}$-embeddings respectively. Since σ_K is an embedding and any fractional ideal $\mathfrak{f}$ is a lattice, $\sigma_K(\mathfrak{f})$ is a lattice in $\prod_\sigma \mathbb{R}$. We now determine the covolume of $\sigma_K(\mathfrak{f})$:

Proposition 2.7.2. *Let K be a number field with signature (r_1, r_2). Then*

$$V_{\sigma_K(\mathfrak{f})} = \mathrm{N}(\mathfrak{f}) \frac{\sqrt{|\Delta_K|}}{2^{r_2}}.$$

In particular,

$$V_{\sigma_K(\mathcal{O}_K)} = \frac{\sqrt{|\Delta_K|}}{2^{r_2}}.$$

Proof. The second statement follows from the first by taking $\mathfrak{f} = \mathcal{O}_K$ so it suffices to prove the first statement. Let $\mathfrak{f}$ be a fractional ideal with basis $\{\kappa_1, \ldots, \kappa_n\}$. Then the associated generator matrix P for $\sigma_K(\mathfrak{f})$ is

$$P = \begin{pmatrix} \rho_1(\kappa_1) & \cdots & \rho_{r_1}(\kappa_1) & \mathrm{Re}(\tau_1(\kappa_1)) & \mathrm{Im}(\tau_1(\kappa_1)) & \cdots & \mathrm{Re}(\tau_{r_2}(\kappa_1)) & \mathrm{Im}(\tau_{r_2}(\kappa_1)) \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_1(\kappa_n) & \cdots & \rho_{r_1}(\kappa_n) & \mathrm{Re}(\tau_1(\kappa_n)) & \mathrm{Im}(\tau_1(\kappa_n)) & \cdots & \mathrm{Re}(\tau_{r_2}(\kappa_n)) & \mathrm{Im}(\tau_{r_2}(\kappa_n)) \end{pmatrix}^t.$$

To compute the determinant of P, first add an i multiple of the imaginary columns to their corresponding real columns and then apply the identity $\mathrm{Im}(z) = \frac{z - \overline{z}}{2i}$ to the imaginary columns to obtain

$$P' = \begin{pmatrix} \rho_1(\kappa_1) & \cdots & \rho_{r_1}(\kappa_1) & \tau_1(\kappa_1) & \frac{\tau_1(\kappa_1) - \overline{\tau_1(\kappa_1)}}{2i} & \cdots & \tau_{r_2}(\kappa_1) & \frac{\tau_{r_2}(\kappa_1) - \overline{\tau_{r_2}(\kappa_1)}}{2i} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_1(\kappa_n) & \cdots & \rho_{r_1}(\kappa_n) & \tau_1(\kappa_n) & \frac{\tau_1(\kappa_n) - \overline{\tau_1(\kappa_n)}}{2i} & \cdots & \tau_{r_2}(\kappa_n) & \frac{\tau_{r_2}(\kappa_n) - \overline{\tau_{r_2}(\kappa_n)}}{2i} \end{pmatrix}^t.$$

Since P' differs from P by column addition, their determinants are the same. Multiplying the imaginary columns of P' by $-2i$ and then adding the corresponding columns to annihilate the negative terms results in

$$P'' = \begin{pmatrix} \rho_1(\kappa_1) & \cdots & \rho_{r_1}(\kappa_1) & \tau_1(\kappa_1) & \overline{\tau_1(\kappa_1)} & \cdots & \tau_{r_2}(\kappa_1) & \overline{\tau_{r_2}(\kappa_1)} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_1(\kappa_n) & \cdots & \rho_{r_1}(\kappa_n) & \tau_1(\kappa_n) & \overline{\tau_1(\kappa_n)} & \cdots & \tau_{r_2}(\kappa_n) & \overline{\tau_{r_2}(\kappa_n)} \end{pmatrix}^t.$$

As P'' differs from P' by column addition and column scaling of which there were r_2 many of factor $-2i$, the determinant of P'' is $(-2i)^{-r_2}$ that of P'. Altogether,

$$V_{\mathfrak{f}} = |\det(P)| = |\det(P')| = |(-2i)^{-r_2}\det(P'')| = 2^{-r_2}|\det(P'')|.$$

Since the complex $\mathbb{Q}$-embeddings occur in conjugate pairs and $\{\kappa_1, \ldots, \kappa_n\}$ is a basis for K, we see that $P'' = M(\kappa_1, \ldots, \kappa_n)$ is the embedding matrix of the basis. Hence

$$V_{\mathfrak{f}} = \frac{|\det(M(\kappa_1, \ldots, \kappa_n))|}{2^{r_2}}.$$

We will be done if we can show

$$|\det(M(\kappa_1, \ldots, \kappa_n))| = \mathrm{N}(\mathfrak{f})|\det(M(\alpha_1, \ldots, \alpha_n))|,$$

for any integral basis $\{\alpha_1, \ldots, \alpha_n\}$ since $|\det(M(\alpha_1, \ldots, \alpha_n))| = \sqrt{|\Delta_K|}$. As $\mathfrak{f}$ is a fractional ideal, there exists a nonzero $\delta \in \mathcal{O}_K$ and an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{f} = \frac{1}{\delta}\mathfrak{a}.$$

Then $\{\delta\kappa_1, \ldots, \delta\kappa_n\}$ is a basis for $\mathfrak{a}$. Now

$$|\det(M(\delta\kappa_1, \ldots, \delta\kappa_n))| = \mathrm{N}(\mathfrak{a})|\det(M(\alpha_1, \ldots, \alpha_n))|,$$

by Proposition C.1.1 since $\mathrm{N}(\mathfrak{a})$ is the absolute value of the determinant of the base change matrix from $\{\kappa_1, \ldots, \kappa_n\}$ to $\{\delta\kappa_1, \ldots, \delta\kappa_n\}$ (from the definition of the norm). Similarly,

$$|\det(M(\delta\kappa_1, \ldots, \delta\kappa_n))| = |\mathrm{N}(\delta)||\det(M(\kappa_1, \ldots, \kappa_n))|,$$

by Proposition C.1.1 because $|\mathrm{N}(\delta)|$ is the absolute value of the determinant of the base change matrix from $\{\kappa_1, \ldots, \kappa_n\}$ to $\{\delta\kappa_1, \ldots, \delta\kappa_n\}$ (since $\mathrm{N}(\delta)$ is the determinant of T_δ). As $\mathrm{N}(\mathfrak{f}) = \frac{\mathrm{N}(\mathfrak{a})}{|\mathrm{N}(\delta)|}$ (because $\mathrm{N}(\mathfrak{a}) = \mathrm{N}(\delta\mathfrak{f}) = |\mathrm{N}(\delta)|\mathrm{N}(\mathfrak{a})$ using Proposition 2.4.1), these two identities for $|\det(M(\delta\kappa_1, \ldots, \delta\kappa_n))|$ together imply the claim. □

2.8 The Ideal Class Group

Let K be a number field. Recall that the ideal group I_K is the group of fractional ideals of K. We let P_K denote the subgroup of I_K of principal ideals $\alpha\mathcal{O}_K$ for nonzero $\alpha \in K$. Since I_K is abelian, P_K is normal. The **ideal class group** $\mathrm{Cl}(K)$ of K is defined to be the quotient group

$$\mathrm{Cl}(K) = I_K/P_K,$$

of fractional ideals modulo principal ideals. We call an element of $\mathrm{Cl}(K)$ an **ideal class** of K. The **class number** h_K of K is defined by

$$h_K = |\,\mathrm{Cl}(K)|.$$

That is, the class number is the size of the ideal class group. The class number is a measure of how much the ring of integers $\mathcal{O}_K$ fails to be a principal ideal domain. Indeed, if $\mathcal{O}_K$ is a principal ideal domain then every integral ideal is principal and hence every fractional ideal is too (because every fractional ideal $\mathfrak{f}$ is of the form $\frac{1}{\delta}\mathfrak{a}$ for some integral ideal $\mathfrak{a}$ and nonzero $\delta \in \mathcal{O}_K$). But then $\mathrm{Cl}(K)$ is the trivial group an hence $h_K = 1$. Conversely, if $h_K = 1$ then every fractional ideal is principal and hence every integral ideal is too so that $\mathcal{O}_K$ is a principal ideal domain. In short, $\mathcal{O}_K$ is a principal ideal domain if and only if $h_K = 1$. Our primary goal is to show that the class number is finite:

Theorem 2.8.1. *Let K be a number field of degree n and signature (r_1, r_2). Also, let r be a realization of K and $X \subseteq \mathbb{R}^n$ be a compact convex symmetric set. Set $M = \max_{\mathbf{x}\in X}\left(\prod_{1\le i\le n}|x_i|\right)$ where $x = (x_1, \ldots, x_n)$. Then every ideal class contains an integral ideal $\mathfrak{a}$ satisfying*

$$\mathrm{N}(\mathfrak{a}) \le \frac{2^{r_1+r_2}M}{\mathrm{Vol}(X)}\sqrt{|\Delta_K|}.$$

Moreover, the ideal class group $\mathrm{Cl}(K)$ is finite so that the class number h_K is too.

Proof. Let $\mathfrak{f}$ be a fractional ideal, and set

$$\lambda^n = 2^n\frac{V_{\sigma_K(\mathfrak{f}^{-1})}}{\mathrm{Vol}(X)}.$$

Then by construction,

$$\mathrm{Vol}(\lambda X) = \lambda^n\,\mathrm{Vol}(X) = 2^n V_{\sigma_K(\mathfrak{f}^{-1})}.$$

By Minkowski's lattice point theorem, there exists a nonzero $\alpha \in \mathfrak{f}^{-1}$ such that $\sigma_K(\alpha) \in \sigma_K(\mathfrak{f}^{-1})$ and $\sigma_K(\alpha) \in \lambda X$. Since $\alpha \in \mathfrak{f}^{-1}$, $\alpha\mathfrak{f} \subseteq \mathcal{O}_K$ so that $\alpha\mathfrak{f}$ is an integral ideal in the same ideal class as $\mathfrak{f}$. Now let σ run over the n distinct $\mathbb{Q}$-embeddings of K into $\overline{K}$. Since the norm is completely multiplicative by Proposition 2.4.5, we have

$$\mathrm{N}(\alpha\mathfrak{f}) = |\,\mathrm{N}(\alpha)|\,\mathrm{N}(\mathfrak{f}) = \left|\prod_{\sigma}\sigma(\alpha)\right|\mathrm{N}(\mathfrak{f}) \le \lambda^n M\,\mathrm{N}(\mathfrak{f}),$$

where in the first equality we have applied Proposition 2.4.1, in the second we have used Proposition 2.2.1, and the inequality follows since $\sigma_K(\alpha) \in \lambda X$. This inequality, our choice of λ^n, and Proposition 2.7.2 together give

$$\mathrm{N}(\alpha\mathfrak{f}) \le \lambda^n M\,\mathrm{N}(\mathfrak{f}) = 2^n M N(\mathfrak{f})\frac{V_{\sigma_K(\mathfrak{f}^{-1})}}{\mathrm{Vol}(X)} = 2^n M\frac{\sqrt{|\Delta_K|}}{2^{r_2}\,\mathrm{Vol}(X)} = \frac{2^{r_1+r_2}M}{\mathrm{Vol}(X)}\sqrt{|\Delta_K|},$$

which proves the first statement since the fractional ideal $\mathfrak{f}$ was arbitrary. We now prove that the class group is finite. By what we have just proved, we can find a complete set of representatives of $\mathrm{Cl}(K)$ consisting of integral ideals of bounded norm. Since the norm is multiplicative by Proposition 2.4.5, the prime factors of these representatives have bounded norm as well. As we have seen, the norm of a prime integral ideal is exactly the prime p below it. Thus the norms of these prime factors are bounded primes p. As there are finitely many prime integral ideals above any prime p (because $p\mathcal{O}_K$ factors into a product of prime integral ideals and these are exactly the prime integral ideals above p), it follows that these representatives have finitely many prime factors. Altogether this means that there are finitely many representatives. Hence $\mathrm{Cl}(K)$ is finite and so the class number h_K is too. $\square$

We would like to obtain an explicit bound in Theorem 2.8.1 by making a choice for the set X. To obtain a bound that is not too large, we need to ensure that the volume of X is large while the constant M is small. The following lemma dictates our choice of X and computes its volume:

Lemma 2.8.1. *Suppose n is a positive integer and write $n = r_1 + 2r_2$ for some nonnegative integers r_1 and r_2. Let $X \subset \mathbb{R}^n$ to be the compact convex symmetric set given by*

$$X = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{1 \le i \le r_1} |x_i| + 2 \sum_{\substack{1 \le j \le r_2 \\ j \equiv 1 \pmod 2}} \sqrt{x_{r_1+j}^2 + x_{r_1+j+1}^2} \le n \right\}.$$

Then

$$\mathrm{Vol}(X) = \frac{n^n}{n!} 2^{r_1} \left(\frac{\pi}{2}\right)^{r_2}.$$

Proof. Making the change of variables $x_{r_1+j} \to u_j \sin(\theta_j)$ and $x_{r_1+j+1} \to u_j \cos(\theta_j)$ for all j gives

$$\mathrm{Vol}(X) = \int_{X'} u_1 \cdots u_{r_2} \, dx_1 \cdots dx_{r_1} \, du_1 \theta_1 \cdots du_{r_2} \theta_{r_2},$$

where X' is the region

$$X' = \left\{ (x_1, \ldots, x_{r_1}, u_1, \theta_1, \ldots, u_{r_2}, \theta_{r_2}) : \sum_{1 \le i \le r_1} |x_i| + 2 \sum_{1 \le j \le r_2} u_j \le n \right\}.$$

Since the integrand is independent of the θ_j, we have

$$\mathrm{Vol}(X) = (2\pi)^{r_2} \int_{X'} u_1 \cdots u_{r_2} \, dx_1 \cdots dx_{r_1} \, du_1 \cdots du_{r_2}.$$

Making the change of variables $u_j \to \frac{u_j}{2}$ for all j and using the fact that the integrand is symmetric in the x_i for all i gives

$$\mathrm{Vol}(X) = 2^{r_1} \left(\frac{\pi}{2}\right)^{r_2} \int_{X''} u_1 \cdots u_{r_2} \, dx_1 \cdots dx_{r_1} \, du_1 \cdots du_{r_2}, \tag{2.3}$$

where X'' is the region

$$X'' = \left\{ (x_1, \ldots, x_{r_1}, u_1, \ldots, u_{r_2}) : \sum_{1 \le i \le r_1} x_i + \sum_{1 \le j \le r_2} u_j \le n \right\}.$$

To compute the remaining integral, for nonnegative integers ℓ and k and $t \ge 0$, we let

$$X''_{\ell,k}(t) = \left\{ (x_1, \ldots, x_\ell, u_1, \ldots, u_k) : \sum_{1 \le i \le \ell} x_i + \sum_{1 \le j \le k} u_j \le t \right\},$$

and set

$$I_{\ell,k}(t) = \int_{X''_{\ell,k}(t)} u_1 \cdots u_\ell \, dx_1 \cdots dx_n \, du_1 \cdots du_k.$$

Then we have to compute $I_{r_1,r_2}(n)$. To this end, the change of variables $x_i \to tx_i$ and $u_j \to tu_j$ for all i and j gives

$$I_{\ell,k}(t) = t^{\ell+2k} I_{\ell,k}(1). \tag{2.4}$$

Now note that the condition
$$\sum_{1\le i\le \ell} x_i + \sum_{1\le j\le k} u_j \le t,$$
is equivalent to
$$\sum_{1\le i\le \ell-1} x_i + \sum_{1\le j\le k} u_j \le t - x_i.$$
This fact together with Fubini's theorem and Equation (2.4) give
$$I_{\ell,k}(1) = \int_0^1 I_{\ell-1,k}(1-x_\ell)\,dx_\ell = \int_0^1 (1-x_\ell)^{\ell-1+2k} I_{\ell-1,k}(1)\,dx_\ell = \frac{1}{\ell+2k} I_{\ell-1,k}(1).$$
Repeating this procedure $\ell-1$ times results in
$$I_{\ell,k}(1) = \frac{1}{(\ell+2k)\cdots(2k+1)} I_{0,k}(1). \tag{2.5}$$
Similarly, the condition
$$\sum_{1\le j\le k} u_j \le t,$$
is equivalent to
$$\sum_{1\le j\le k-1} u_j \le t - u_k.$$
This fact together with Fubini's theorem, Equation (2.4), and Proposition 1.7.4 give
$$I_{0,k}(1) = \int_0^1 u_k I_{0,k-1}(1-u_k)\,du_k = \int_0^1 u_k(1-u_k)^{2k-2} I_{0,k-1}(1)\,du_k = B(1,2k-1) I_{0,k-1}(1) = \frac{1}{2k} I_{0,k-1}(1).$$
Repeating this procedure $k-1$ times results in
$$I_{0,k}(1) = \frac{1}{k!}, \tag{2.6}$$
since $I_{0,0}(1) = 1$. Combining Equations (2.4) to (2.6) we find that
$$I_{\ell,k}(t) = t^{\ell+2k}\frac{1}{(\ell+2k)!}.$$
In particular, $I_{r_1,r_2}(n) = \frac{n^n}{n!}$ and from Equation (2.3) we obtain
$$\mathrm{Vol}(X) = \frac{n^n}{n!} 2^{r_1}\left(\frac{\pi}{2}\right)^{r_2},$$
as desired. $\square$

Observe that the set X in Lemma 2.8.1 just consists of those points in $\mathbb{R}^n$ whose induced norm corresponding to the induced inner product in Proposition 2.7.1 is at most n. We can now obtain an explicit bound in Theorem 2.8.1 known as the **Minkowski bound**:

Theorem 2.8.2 (Minkowski bound). *Let K be a number field of degree n and signature (r_1, r_2). Then every ideal class contains an integral ideal $\mathfrak{a}$ satisfying*
$$\mathrm{N}(\mathfrak{a}) \le \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n}\sqrt{|\Delta_K|}.$$

Proof. Let X be given by

$$X = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{1 \le i \le r_1} |x_i| + 2 \sum_{\substack{1 \le j \le r_2 \\ j \equiv 1 \pmod 2}} \sqrt{x_{r_1+j}^2 + x_{r_1+j+1}^2} \le n \right\}.$$

Then Theorem 2.8.1 and Lemma 2.8.1 together give

$$\mathrm{N}(\mathfrak{a}) \le M \left(\frac{4}{\pi}\right)^{r_2} \frac{n!}{n^n} \sqrt{|\Delta_K|},$$

where $M = \max_{\mathbf{x} \in X} \left(\prod_{1 \le \ell \le n} |x_\ell|\right)$. But for all $\mathbf{x} \in X$, the arithmetic-geometric mean inequality gives

$$\left(\prod_{1 \le \ell \le n} |x_\ell|\right)^{\frac{1}{n}} \le \frac{1}{n} \sum_{1 \le \ell \le n} |x_\ell| \le 1,$$

where the second inequality holds by the definition of X. Hence $M \le 1$ and this completes the proof. □

As a corollary we can obtain a lower bound for the discriminant of a number field and show that every number field other than $\mathbb{Q}$ has at least one ramified prime:

Corollary 2.8.1. *Let K be a number field of degree n. Then*

$$|\Delta_K| \ge \left(\frac{\pi}{4}\right)^{\frac{n}{2}} \frac{n^n}{n!}.$$

In particular, every number field of degree at least 2 contains at least one ramified prime.

Proof. Since the norm of every integral ideal is at least 1, $\pi < 4$, and r_2 is at most n, the desired inequality follows immediately from Minkowski's bound. Now suppose $n \ge 2$. In the case $n = 2$, the lower bound is larger than 1 so that $|\Delta_K|$ is at least 2 for every quadratic number field. As $n^n \ge n!$ for all $n \ge 1$ (which easily follows by induction), $\left(\frac{\pi}{4}\right)^{\frac{n}{2}} \frac{n^n}{n!}$ is an increasing function in n. Therefore $|\Delta_K| \ge 2$ for all $n \ge 2$ so that $|\Delta_K|$ has a prime divisor. Then Theorem 2.6.1 implies that at least one prime is ramified in K. □

2.9 Todo: [Dirichlet's Unit Theorem]

2.10 Quadratic Number Fields

We will now classify and discuss the structure of quadratic number fields. We first show that quadratic number fields are exactly those where we adjoint the square-root of a fundamental discriminant:

Proposition 2.10.1. *Every quadratic number field K is of the form $K = \mathbb{Q}(\sqrt{d})$ for some square-free integer d other than 0 or 1.*

Proof. Suppose K is a quadratic number field. In particular, $K/\mathbb{Q}$ is separable so by the primitive element theorem there exists $\theta \in K$ such that $K = \mathbb{Q}(\theta)$. The minimal polynomial $m_\theta(x)$ of θ is of the form

$$m_\theta(x) = x^2 + ax + b,$$

for $a, b \in \mathbb{Q}$. Then the quadratic formula gives

$$\theta = -\frac{a}{2} \pm \frac{\sqrt{q}}{2},$$

where $q = a^2 - 4b \in \mathbb{Q}$. Clearly $q \neq 0$ and $q \neq 1$ for otherwise $\theta \in \mathbb{Q}$. It follows that $K = \mathbb{Q}(\sqrt{q})$. Write $q = \frac{n}{m}$ for relatively prime $n, m \in \mathbb{Z}$ and set $d = m^2 q = nm \in \mathbb{Z}$. Then d is square-free, $d \neq 0$, and $d \neq 1$. Moreover, $\sqrt{d} = m\sqrt{q}$ so that $K = \mathbb{Q}(\sqrt{d})$. □

From Proposition 2.10.1, we see that the d for a quadratic number field $\mathbb{Q}(\sqrt{d})$ satisfies $d \equiv 1, 2, 3 \pmod 4$ (otherwise d is not square-free). Moreover, any element of a quadratic number field is of the form $a + b\sqrt{d}$ with $a, b \in \mathbb{Q}$ and for some square-free d other than 0 or 1. We say that a quadratic number field $\mathbb{Q}(\sqrt{d})$ is **real** if $d > 0$ and **imaginary** if $d < 0$. Now $\mathbb{Q}(\sqrt{d})$ is real or imaginary according to if $\sqrt{d}$ is real or purely imaginary so that the two $\mathbb{Q}$-embeddings σ_1 and σ_2 of $\mathbb{Q}(\sqrt{d})$ into $\mathbb{C}$ are

$$\sigma_1(a + b\sqrt{d}) = a + b\sqrt{d} \quad \text{and} \quad \sigma_2(a + b\sqrt{d}) = a - b\sqrt{d},$$

because the roots of the minimal polynomial for $\sqrt{d}$ are $\pm\sqrt{d}$. In particular, the signature is $(2, 0)$ or $(0, 1)$ according to if $\mathbb{Q}(\sqrt{d})$ is real or imaginary. In either case, Proposition 2.2.1 shows that the norm and trace of $\kappa = a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ are given by

$$\mathrm{Tr}(\kappa) = 2a \quad \text{and} \quad \mathrm{N}(\kappa) = a^2 - b^2 d.$$

We will now begin describing the ring of integers, discriminant, and the factorization of primes in $\mathbb{Q}(\sqrt{d})$. For simplicity, we write $\mathcal{O}_d = \mathcal{O}_{\mathbb{Q}(\sqrt{d})}$ and $\Delta_d = \Delta_{\mathbb{Q}(\sqrt{d})}$. The ring of integers has a particularly simple description since it is monogenic as the following proposition shows:

Proposition 2.10.2. *Let $\mathbb{Q}(\sqrt{d})$ be a quadratic number field. Then $\mathbb{Q}(\sqrt{d})$ is monogenic and*

$$\mathcal{O}_d = \begin{cases} \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] & \text{if } d \equiv 1 \pmod 4, \\ \mathbb{Z}[\sqrt{d}] & \text{if } d \equiv 2, 3 \pmod 4. \end{cases}$$

Proof. Let $\alpha = a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ be an algebraic integer. If $b = 0$, then $\alpha \in \mathbb{Q}$ and since the only elements of $\mathbb{Q}$ that are algebraic integers are the integers themselves we must have that α is an integer. Now suppose $b \neq 0$. Then the minimal polynomial of α is

$$m_\alpha(x) = x^2 + 2ax + (a^2 - b^2 d) = (x - (a + b\sqrt{d}))(x - (a - b\sqrt{d})).$$

As α is an algebraic integer, $2a \in \mathbb{Z}$ and $a^2 - b^2 d \in \mathbb{Z}$ (note that these are the trace and norm of α respectively). In particular, $(2a)^2 + (2b)^2 d \in \mathbb{Z}$ and hence $(2b)^2 \in \mathbb{Z}$ is as well. But as $b \in \mathbb{Q}$, it must be the case that $2b \in \mathbb{Z}$. If $2a = n + 1$ is odd then n is even. We compute

$$a^2 - b^2 d = \left(\frac{n+1}{2}\right)^2 - b^2 d = \frac{n^2 + 2n + 1 + 4b^2 d}{4},$$

and since the right-hand side must be an integer $b \notin \mathbb{Z}$. For if $b \in \mathbb{Z}$, the numerator of the right-hand side is equivalent to 1 modulo 4 because n is even. As $2b \in \mathbb{Z}$ it follows that $2b$ must be odd so set $2b = m + 1$ with m even. Again, we compute

$$a^2 - b^2 d = \left(\frac{n+1}{2}\right)^2 - \left(\frac{m+1}{2}\right)^2 d = \frac{n^2 + 2n + 1 - d(m^2 + 2m + 1)}{4},$$

and since the right-hand side must be an integer the numerator must be divisible by 4. As n and m are even, this is equivalent to $d \equiv 1 \pmod 4$. So we have shown $2a$ or $2b$ is odd if and only if $d \equiv 1 \pmod 4$. Thus if $d \equiv 1 \pmod 4$, we have $a = \frac{a'}{2}$ and $b = \frac{b'}{2}$ for some $a', b' \in \mathbb{Z}$ and hence $\alpha \in \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$. Otherwise, $d \equiv 2, 3 \pmod 4$ (because d is square-free) so that $2a$ and $2b$ are both even, $a, b \in \mathbb{Z}$, and therefore $\alpha \in \mathbb{Z}[\sqrt{d}]$. We have now shown that $\mathcal{O}_d \subseteq \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ and $\mathcal{O}_d \subseteq \mathbb{Z}[\sqrt{d}]$ according to if $d \equiv 1 \pmod 4$ or $d \equiv 2, 3 \pmod 4$ respectively. For the reverse containment, just note that $\sqrt{d}$ is an algebraic integer since its minimal polynomial $m_{\sqrt{d}}(x)$ is

$$m_{\sqrt{d}}(x) = x^2 \pm d,$$

according to if $d < 0$ or $d > 0$. The reverse containment now follows by Proposition 2.1.1 and that all integers are algebraic integers. $\square$

It follows from Proposition 2.10.2 that

$$\left\{1, \frac{1+\sqrt{d}}{2}\right\} \quad \text{and} \quad \{1, \sqrt{d}\},$$

are integral bases for $\mathcal{O}_d$ according to if $d \equiv 1 \pmod 4$ or $d \equiv 2, 3 \pmod 4$ respectively. Let us now show that the discriminants quadratic number fields are exactly the fundamental discriminants D other than 1:

Proposition 2.10.3. *Let $\mathbb{Q}(\sqrt{d})$ be a quadratic number field. Then*

$$\Delta_d = \begin{cases} d & \text{if } d \equiv 1 \pmod 4, \\ 4d & \text{if } d \equiv 2, 3 \pmod 4. \end{cases}$$

In particular, the discriminants quadratic number fields are exactly the fundamental discriminants other than 1.

Proof. Let σ_1 and σ_2 be the two $\mathbb{Q}$-embeddings of $\mathbb{Q}(\sqrt{d})$ into $\mathbb{C}$ where σ_1 is the identity and σ_2 is given by sending $\sqrt{d}$ to its conjugate. If $d \equiv 1 \pmod 4$, an integral basis for $\mathcal{O}_d$ is $\left\{1, \frac{1+\sqrt{d}}{2}\right\}$. In this case, the embedding matrix is

$$M\left(1, \frac{1+\sqrt{d}}{2}\right) = \begin{pmatrix} 1 & \frac{1+\sqrt{d}}{2} \\ 1 & \frac{1-\sqrt{d}}{2} \end{pmatrix},$$

and thus $\Delta_d = d$. If $d \equiv 2, 3 \pmod 4$, an integral basis for $\mathcal{O}_d$ is $\left\{1, \sqrt{d}\right\}$. In this case, the embedding matrix is

$$M(1, \sqrt{d}) = \begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{pmatrix},$$

and hence $\Delta_d = 4d$. This proves the first statement and the second statement is clear since d is square-free and not 0 or 1. $\square$

We will now discuss the factorization of a prime p in a quadratic number field $\mathbb{Q}(\sqrt{d})$. Since $\mathbb{Q}(\sqrt{d})$ is a degree 2 extension, Proposition 2.6.1 implies that p is ramified if and only if it is totally ramified and if p is split but not ramified then it is totally split. In other words, there are three possible cases for how $p\mathcal{O}_d$ factors:

$$p\mathcal{O}_d = \mathfrak{p}, \quad p\mathcal{O}_d = \mathfrak{p}^2, \quad \text{and} \quad p\mathcal{O}_d = \mathfrak{p}\mathfrak{q},$$

according to if p is inert, totally ramified, or totally split. Since $\mathbb{Q}(\sqrt{d})$ is monogenic by Proposition 2.10.2, we can describe the factorization using the Dedekind-Kummer theorem and connect it to the quadratic character χ_{Δ_d} given by the fundamental discriminant Δ_d:

Proposition 2.10.4. *Let $\mathbb{Q}(\sqrt{d})$ be a quadratic number field and let χ_{Δ_d} be the quadratic character given by the fundamental discriminant Δ_d. Then for any prime p, we have*

$$\chi_{\Delta_d}(p) = \begin{cases} 1 & \text{if } p \text{ is split,} \\ -1 & \text{if } p \text{ is inert,} \\ 0 & \text{if } p \text{ is ramified.} \end{cases}$$

Proof. By Theorem 2.6.1, p is ramified if and only if p divides $|\Delta_d|$ but this is exactly when $\chi_{\Delta_d}(p) = 0$. Therefore it suffices to prove the cases when p is split and inert. First suppose $d \equiv 1 \pmod 4$ so that $\mathcal{O}_d = \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ and $\Delta_d = d$ by Propositions 2.10.2 and 2.10.3. The minimal polynomial $m_{\frac{1+\sqrt{d}}{2}}(x)$ for $\frac{1+\sqrt{d}}{2}$ is

$$m_{\frac{1+\sqrt{d}}{2}}(x) = x^2 - x + \frac{1-d}{4},$$

where $\frac{1-d}{4} \in \mathbb{Z}$ because $d \equiv 1 \pmod 4$. The reduction of $m_{\frac{1+\sqrt{d}}{2}}(x)$ modulo p is either irreducible, factors into two distinct linear factors, or is a square, and Dedekind-Kummer theorem implies that this is equivalent to p being inert, split, or ramified accordingly because the prime factorization is unique. First suppose $p \neq 2$. Then from the quadratic formula, $m_{\frac{1+\sqrt{d}}{2}}(x)$ reduces modulo p as

$$m_{\frac{1+\sqrt{d}}{2}}(x) \equiv \left(x - \frac{1+\sqrt{d}}{2}\right)\left(x - \frac{1-\sqrt{d}}{2}\right) \pmod p,$$

if and only if the roots $\frac{1\pm\sqrt{d}}{2}$ are elements of $\mathbb{F}_p$ and is otherwise irreducible. As $p \neq 2$, these factors are distinct. Moreover, $\frac{1\pm\sqrt{d}}{2}$ is an element of $\mathbb{F}_p$ if and only if d is a square modulo p and hence p is split or inert according to if $\chi_d(p) = \pm 1$. Now suppose $p = 2$. Since $m_{\frac{1+\sqrt{d}}{2}}(x)$ has a nonzero linear term with an odd coefficient, it reduces modulo 2 as

$$m_{\frac{1+\sqrt{d}}{2}}(x) \equiv x(x-1) \pmod 2,$$

if and only if $\frac{1-d}{4} \equiv 0 \pmod 2$ and is otherwise irreducible. Clearly these factors are distinct. Now observe $\frac{1-d}{4} \equiv 0 \pmod 2$ is equivalent to $d \equiv 1 \pmod 8$ provided $d > 0$ and $d \equiv 7 \pmod 8$ provided $d < 0$ and thus p is split or inert according to if $\chi_d(2) = \pm 1$. This completes the argument in the case $d \equiv 1 \pmod 4$. Now suppose $d \equiv 2, 3 \pmod 4$ so that $\mathcal{O}_d = \mathbb{Z}[\sqrt{d}]$ and $\Delta_d = 4d$ by Propositions 2.10.2 and 2.10.3. The minimal polynomial $m_{\sqrt{d}}(x)$ for $\sqrt{d}$ is

$$m_{\sqrt{d}}(x) = x^2 - d.$$

As $\Delta_d = 4d$, we see that 2 is ramified and therefore we may assume $p \neq 2$. Similarly, the reduction of $m_{\sqrt{d}}(x)$ modulo p is either irreducible, factors into two distinct linear factors, or is a square, and Dedekind-Kummer theorem implies that this is equivalent to p being inert, split, or ramified accordingly because the prime factorization is unique. As $p \neq 2$, the quadratic formula implies that $m_{\sqrt{d}}(x)$ reduces modulo p as

$$m_{\sqrt{d}}(x) \equiv (x - \sqrt{d})(x + \sqrt{d}) \pmod p,$$

if and only if the roots $\pm\sqrt{d}$ are elements of $\mathbb{F}_p$. As $p \neq 2$, these factors are distinct. Moreover, $\sqrt{d}$ is an element of $\mathbb{F}_p$ if and only if d and hence $4d$ are squares modulo p so that p is split or inert according to if $\chi_{4d}(p) = \pm 1$. This completes the verification in the case $d \equiv 2, 3 \pmod 4$. $\square$

From Proposition 2.10.4, we see that the factorization of primes in $\mathbb{Q}(\sqrt{d})$ is controlled by the quadratic character χ_{Δ_d} attached to the fundamental discriminant Δ_d. In other words, the factorization of p depends completely upon if Δ_d is a square modulo p.

Part III

An Introduction to Holomorphic, Automorphic & Maass Forms

Chapter 3

Congruence Subgroups & Modular Curves

Every holomorphic or Maass form is a special type of function depending on certain subgroups of $\mathrm{PSL}_2(\mathbb{Z})$. These are the congruence subgroups. The associated modular curve is the quotient of the upper half-space $\mathbb{H}$ by an action of this subgroup. We introduce these topics first as they are the foundation for discussing holomorphic and Maass forms in complete generality.

3.1 Congruence Subgroups

The **modular group** is $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$. That is, the modular group is the set of matrices with integer entries and of determinant 1 determined up to sign. The reason we are only interested in these matrices up to sign is because the modular group has a natural action on the upper half-space $\mathbb{H}$ and this action will be invariant under a change in sign. The first result usually proved about the modular group is that it is generated by two matrices:

Proposition 3.1.1.

$$\mathrm{PSL}_2(\mathbb{Z}) = \left\langle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\rangle.$$

Proof. Set S and T to be the first and second generators respectively. Clearly they belong to $\mathrm{PSL}_2(\mathbb{Z})$. Also, S and T^n for $n \in \mathbb{Z}$ acts on $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{Z})$ by

$$S\gamma = S\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix} \quad \text{and} \quad T^n\gamma = T^n\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+nc & b+nd \\ c & d \end{pmatrix}.$$

In particular, S interchanges the upper-left and lower-left entries of γ up to sign and T^n adds an n multiple of the lower-left entry to the upper-left entry. We have to show $\gamma \in \langle S, T \rangle$ and we will accomplish this by showing that the inverse is in $\langle S, T \rangle$. If $|c| = 0$ then γ is the identity since $\det(\gamma) = 1$ so suppose $|c| \neq 0$. By Euclidean division we can write $a = qc + r$ for some $q \in \mathbb{Z}$ and $|r| < |c|$. Then

$$T^{-q}\gamma = \begin{pmatrix} a-qc & b-qd \\ c & d \end{pmatrix} = \begin{pmatrix} r & b-qd \\ c & d \end{pmatrix}.$$

Multiplying by S yields

$$ST^{-q}\gamma = S\begin{pmatrix} r & b-qd \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ r & b-qd \end{pmatrix},$$

and this matrix has the upper-left entry at least as large as the lower-left entry in norm. Actually the upper-left entry is strictly larger since $|c| > |r|$ by Euclidean division. Therefore if we repeatedly apply this procedure, it must terminate with the lower-left entry vanishing. But then we have reached the identity matrix. Therefore we have show γ has an inverse in $\langle S, T\rangle$. $\square$

We will also be interested in special subgroups of the modular group defined by congruence conditions on their entries. For $N \geq 1$, set

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

Then $\Gamma(N)$ is the kernel of the homomorphism $\mathrm{PSL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})$ given by reducing the coefficients modulo N so it is a normal subgroup with finite index. We call $\Gamma(N)$ the **principal congruence subgroup** of level N. For $\Gamma \leq \mathrm{PSL}_2(\mathbb{Z})$, we say Γ is a **congruence subgroup** if $\Gamma(N) \leq \Gamma$ for some N and the minimal such N is called the **level** of Γ. Note that if $M \mid N$, then $\Gamma(N) \leq \Gamma(M)$. Thus if Γ is a congruence subgroup of level N, then $\Gamma(kN) \leq \Gamma$ for all $k \geq 1$. This implies that congruence subgroups are closed under intersection. Also, it turns out that the aforementioned homomorphism is surjective:

Proposition 3.1.2. *The homomorphism* $\mathrm{PSL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})$ *given by reducing the coefficients modulo* N *is surjective.*

Proof. Suppose $\left(\begin{smallmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Then $\overline{a}\overline{d} - \overline{b}\overline{c} \equiv 1 \pmod N$ so by Bézout's identity (generalized to three integers) $(\overline{c}, \overline{d}, N) = 1$. We claim that there exists s and t such that $c = \overline{c} + sN$, $d = \overline{d} + tN$ with $(c, d) = 1$. Set $g = (\overline{c}, \overline{d})$. Then $(g, N) = 1$ because $(\overline{c}, \overline{d}, N) = 1$. If $\overline{c} = 0$ then set $s = 0$ so $c = 0$ and choose t such that $t \equiv 1 \pmod p$ for any prime $p \mid g$ and $t \equiv 0 \pmod p$ for any prime $p \nmid g$ and $p \mid \overline{d}$. Such a t exists by the Chinese remainder theorem. Now if $p \mid (c, d)$, then either $p \mid g$ or $p \nmid g$. If $p \mid g$, then $p \mid d - \overline{d} = tN$ which is absurd since $t \equiv 1 \pmod p$ and $(t, N) = 1$. If $p \nmid g$, then $p \nmid d - \overline{d} = tN$ but this is also absurd since $t \equiv 0 \pmod p$. Therefore $(c, d) = 1$ as claimed. If $\overline{c} = 0$ then $\overline{d} \neq 0$, and we can proceed similarly. Since $(c, d) = 1$ there exists a and b such that $ad - bc = 1$. Then $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{Z})$ and maps onto $\left(\begin{smallmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{smallmatrix}\right)$. This proves surjectivity. $\square$

By Proposition 3.1.2, $[\mathrm{PSL}_2(Z) : \Gamma(N)] = |\mathrm{PSL}_2(\mathbb{Z}/N\mathbb{Z})|$. Since $\Gamma(N) \leq \Gamma$ and $\Gamma(N)$ has finite index in $\mathrm{PSL}_2(\mathbb{Z})$ so does Γ. The subgroups

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

and

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod N \right\},$$

are particularly important and are congruence subgroups of level N. The latter subgroup is called the **Hecke congruence subgroup** of level N. Note that $\Gamma(N) \leq \Gamma_1(N) \leq \Gamma_0(N)$. If Γ is a general congruence subgroup, it is useful to find a generating set for Γ in order to reduce results about Γ to that of the generators. This is usually achieved by performing some sort of Euclidean division argument on the entries of a matrix $\gamma \in \Gamma$ using the supposed generating set to construct the inverse for γ. We will also require a useful lemma which says that congruence subgroups are preserved under conjugation by elements of $\mathrm{GL}_2^+(\mathbb{Q})$ provided we restrict to those elements in $\mathrm{PSL}_2(\mathbb{Z})$:

Lemma 3.1.1. *Let* Γ *be a congruence subgroup and let* $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. *Then* $\alpha^{-1}\Gamma\alpha \cap \mathrm{PSL}_2(\mathbb{Z})$ *is a congruence subgroup.*

Proof. Recall that if Γ is of level M, then $\Gamma(kM) \leq \Gamma$ for every $k \geq 1$. Thus there is an integer $\widetilde{N}$ such that $\Gamma(\widetilde{N}) \leq \Gamma$, $\widetilde{N}\alpha \in \mathrm{GL}_2^+(\mathbb{Z})$, and $\widetilde{N}\alpha \in \mathrm{GL}_2^+(\mathbb{Z})$. Now let $N = \widetilde{N}^3$ and notice that any $\gamma \in \Gamma(N)$ is of the form

$$\gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + N \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix},$$

for $k_1, \dots, k_4 \in \mathbb{Z}$. Therefore $\Gamma(N) \subseteq I + N\mathrm{Mat}_2(\mathbb{Z})$. Thus

$$\alpha\Gamma(N)\alpha^{-1} \leq \alpha(I + N\mathrm{Mat}_2(\mathbb{Z}))\alpha^{-1} = I + \widetilde{N}\mathrm{Mat}_2(\mathbb{Z}).$$

As every matrix in $\alpha\Gamma(N)\alpha^{-1}$ has determinant 1 and $\Gamma(\widetilde{N}) \subseteq I + \widetilde{N}\mathrm{Mat}_2(\mathbb{Z})$, it follows that $\alpha\Gamma(N)\alpha^{-1} \leq \Gamma(\widetilde{N})$. As $\Gamma(\widetilde{N}) \leq \Gamma$, we conclude

$$\Gamma(N) \leq \alpha^{-1}\Gamma\alpha,$$

and intersecting with $\mathrm{PSL}_2(\mathbb{Z})$ completes the proof. $\square$

Note that by Lemma 3.1.1, if $\alpha^{-1}\Gamma\alpha \subset \mathrm{PSL}_2(\mathbb{Z})$ then $\alpha^{-1}\Gamma\alpha$ is a congruence subgroup if Γ is. Moreover, since congruence subgroups are closed under intersection, Lemma 3.1.1 further implies that $\alpha^{-1}\Gamma_1\alpha \cap \Gamma_2$ is a congruence subgroup for and two congruence subgroups Γ_1 and Γ_2 any any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$.

3.2 Modular Curves

Recall that $\mathrm{GL}_2^+(\mathbb{Q})$ naturally acts on the Riemann sphere $\hat{\mathbb{C}}$ by Möbius transformations. Explicitly, any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ acts on $z \in \hat{\mathbb{C}}$ by

$$\gamma z = \frac{az+b}{cz+d},$$

where $\gamma\infty = \frac{a}{c}$ and $\gamma\left(-\frac{d}{c}\right) = \infty$. Moreover, recall that this action is a group action, is invariant under scalar multiplication, and acts as automorphisms of $\hat{\mathbb{C}}$. Now observe

$$\mathrm{Im}(\gamma z) = \mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \mathrm{Im}\left(\frac{az+b}{cz+d}\frac{c\overline{z}+d}{c\overline{z}+d}\right) = \mathrm{Im}\left(\frac{ac|z|^2 + adz + bc\overline{z} + bd}{|cz+d|^2}\right) = \det(\gamma)\frac{\mathrm{Im}(z)}{|cz+d|^2},$$

where the last equality follows because $\mathrm{Im}(\overline{z}) = -\mathrm{Im}(z)$ and $\det(\gamma) = ad - bc$. Since $\deg(\gamma) > 0$ and $|cz+d|^2 > 0$, γ preserves the sign of the imaginary part of z. So γ preserves the upper half-space $\mathbb{H}$, the lower half-space $\overline{\mathbb{H}}$, and the extended real line $\hat{\mathbb{R}}$ respectively. Moreover, γ restricts to an automorphism on these subspaces since Möbius transformations are automorphisms. In particular, $\mathrm{PSL}_2(\mathbb{Z})$ naturally acts on $\hat{\mathbb{C}}$ by Möbius transformations and preserves the upper half-space. Certain actions of subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ also play important roles. A **Fuchsian group** is any discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ that acts properly discontinuously on $\mathbb{H}$ (see Appendix D.1). It turns out that the modular group is a Fuchsian group (see [DS05] for a proof):

Proposition 3.2.1. *The modular group is a Fuchsian group.*

Note that Proposition 3.2.1 immediately implies that any subgroup of the modular group is also Fuchsian. In particular, all congruence subgroups are Fuchsian. A **modular curve** is a quotient $\Gamma\backslash\mathbb{H}$ of the upper half-space $\mathbb{H}$ by the action of a congruence subgroup Γ. Since Γ is Fuchsian and acts on $\mathbb{H}$ by automorphisms Proposition D.1.1 implies that $\Gamma\backslash\mathbb{H}$ is also connected Hausdorff (recall that $\mathbb{H}$ is connected Hausdorff). In particular, $\Gamma\backslash\mathbb{H}$ admits the fundamental domain

$$\mathcal{F} = \left\{ z \in \mathbb{H} : |\mathrm{Re}(z)| \leq \frac{1}{2} \text{ and } |z| \geq 1 \right\},$$

as the following proposition shows:

Proposition 3.2.2. *$\mathcal{F}$ is a fundamental domain for* $\mathrm{PSL}_2(\mathbb{Z})$.

Proof. Set $\mathrm{PSL}_2(\mathbb{Z}) = \langle S, T\rangle$ where S and T are as in Proposition 3.1.1. We first show any point in $\mathbb{H}$ is $\mathrm{PSL}_2(\mathbb{Z})$-equivalent to a point in $\mathcal{F}$. Then for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{Z})$, we have

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

Since $\det(\gamma) = 1$ we cannot have $c = d = 0$. Then as $y \neq 0$, $|cz+d|^2$ is bounded away from zero and moreover there are finitely many pairs (c, d) such that $|cz+d|^2$ is less than any given upper bound. Therefore there exists $\gamma_0 \in \mathrm{PSL}_2(\mathbb{Z})$ that minimizes $|cz+d|^2$ and hence maximizes $\mathrm{Im}(\gamma_0 z)$. In particular,

$$\mathrm{Im}(S\gamma_0 z) = \frac{\mathrm{Im}(\gamma_0 z)}{|\gamma_0 z|^2} \leq \mathrm{Im}(\gamma_0 z).$$

The inequality above implies $|\gamma_0 z| \geq 1$. Since $\mathrm{Im}(T^n\gamma_0 z) = \mathrm{Im}(\gamma_0 z)$ for all $n \in \mathbb{Z}$, repeating the argument above with $T^n\gamma_0$ in place of γ_0, we see that $|T^n\gamma_0 z| \geq 1$. But T shifts the real part by 1 so we can choose n such that $|\mathrm{Re}(T^n\gamma_0 z)| \leq \frac{1}{2}$. Therefore $T^n\gamma_0 \in \mathrm{PSL}_2(\mathbb{Z})$ sends z into $\mathcal{F}$ as desired. We will now show that if two points in $\mathcal{F}$ are $\mathrm{PSL}_2(\mathbb{Z})$-equivalent via a non-identity element, then they lie on the boundary of $\mathcal{F}$. Since $\mathrm{PSL}_2(\mathbb{Z})$ acts on $\mathbb{H}$ by automorphisms, by Proposition 3.1.1 it suffices to show that S and T map $\mathcal{F}$ outside of $\mathcal{F}$ except for possibly the boundary. This is clear for T since it maps the left boundary line $\left\{z \in \mathbb{H} : \mathrm{Re}(z) = -\frac{1}{2} \text{ and } |z| \geq 1\right\}$ to the right boundary line $\left\{z \in \mathbb{H} : \mathrm{Re}(z) = \frac{1}{2} \text{ and } |z| \geq 1\right\}$ and every other point of $\mathcal{F}$ is mapped to the right of this line. For S, note that it maps the semicircle $\{z \in \mathbb{H} : |z| = 1\}$ to itself (although not identically) and maps ∞ to zero. Since Möbius transformations send circles to circles and lines to lines it follows that every other point of $\mathcal{F}$ is taken to a point enclosed by the semicircle $\{z \in \mathbb{H} : |z| = 1\}$. Lastly, the interior of $\mathcal{F}$ is a domain since it is open and path-connected. This finishes the proof. $\square$

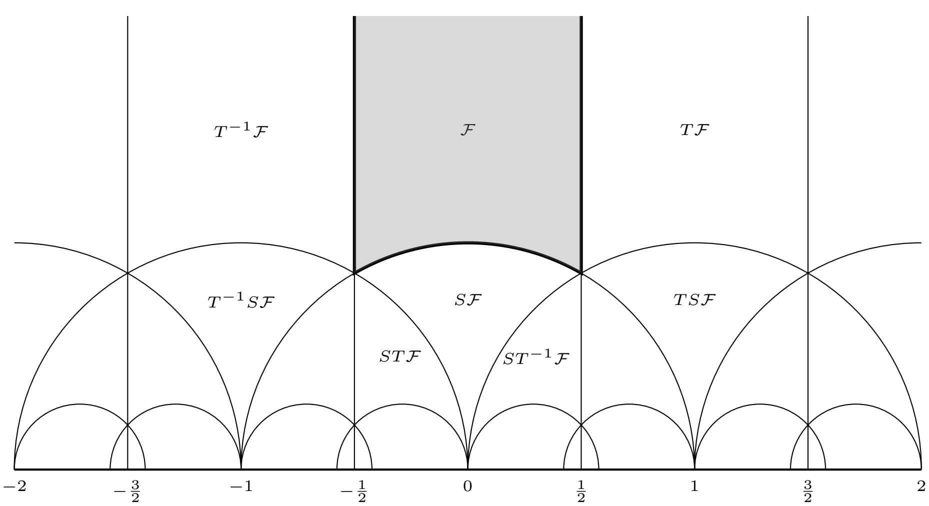

Figure 3.1: The standard fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$.

The region $\mathcal{F}$, shaded in Figure 3.1, is called the **standard fundamental domain**. Figure 3.1 also displays how this fundamental domain changes under the actions of the generators of $\mathrm{PSL}_2(\mathbb{Z})$ as in Proposition 3.1.1. A fundamental domain for any other modular curve can be built from the standard fundamental domain as the following proposition shows (see [Kil15] for a proof):

Proposition 3.2.3. *Let Γ be any congruence subgroup. Then*

$$\mathcal{F}_\Gamma = \bigcup_{\gamma \in \Gamma \backslash \mathrm{PSL}_2(\mathbb{Z})} \gamma \mathcal{F},$$

is a fundamental domain for $\Gamma\backslash\mathbb{H}$.

We might notice that $\mathcal{F}$ in Figure 3.1 is unbounded as it doesn't contain the point ∞. However, if we consider $\mathcal{F} \cup \{\infty\}$ then it would appear that this space is compact. The point ∞ is an example of a cusp and we now make this idea precise. Since any $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ preserves $\hat{\mathbb{R}}$ and γ has integer entries, γ also preserves $\mathbb{Q} \cup \{\infty\}$. A **cusp** of $\Gamma\backslash\mathbb{H}$ is an element of of $\Gamma\backslash(\mathbb{Q} \cup \{\infty\})$. As Γ has finite index in the modular group, there can only be finitely many cusps and the number of cusps is at most the index of Γ. In particular, the Γ-orbit of ∞ is a cusp of $\Gamma\backslash\mathbb{H}$. We denote cusps by gothic characters $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \ldots$ or by representatives of their equivalence classes. For example, we let ∞ denote the cusp $\Gamma\infty$.

Remark 3.2.1. *It turns out that the cusps can be represented as the points needed to make a fundamental domain $\mathcal{F}_\Gamma$ compact as a subset of $\hat{\mathbb{C}}$. To see this, suppose $\mathfrak{a}$ is a limit point of $\mathcal{F}_\Gamma$ that does not belong to $\mathcal{F}_\Gamma$. Then $\mathfrak{a} \in \hat{\mathbb{R}}$. In the case of the standard fundamental domain $\mathcal{F}$, $\mathfrak{a} = \infty$ which is a cusp. Otherwise, $\mathcal{F}_\Gamma$ is a union of images of $\mathcal{F}$ by Proposition 3.2.3 and since $\mathrm{PSL}_2(\mathbb{Z})\infty = \mathbb{Q} \cup \{\infty\}$, we find that $\mathfrak{a} \in \mathbb{Q} \cup \{\infty\}$.*

Let $\Gamma_\mathfrak{a} \leq \Gamma$ denote the stabilizer subgroup of the cusp $\mathfrak{a}$. For the ∞ cusp, we can describe Γ_∞ explicitly. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ stabilizes ∞, then necessarily $c = 0$ and since $\det(\gamma) = 1$ we must have $a = d = 1$. Therefore $\gamma = \left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right)$ for some $b \in \mathbb{Z}$ and γ acts on $\mathbb{H}$ by translation by b. Of course, not every translation is guaranteed to belong to Γ. Letting t be the smallest positive integer such that $\left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$, we have $\Gamma_\infty = \langle \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right) \rangle$. In particular, Γ_∞ is an infinite cyclic group. We say that Γ is **reduced at infinity** if $t = 1$ so that $\Gamma_\infty = \langle \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \rangle$. In particular, $\Gamma_1(N)$ and $\Gamma_0(N)$ are reduced at infinity.

Remark 3.2.2. *If Γ is of level N, then N is the smallest positive integer such that $\Gamma(N) \leq \Gamma$ so that $\left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$ is the minimal translation guaranteed to belong to Γ. However, there may be smaller translations so in general $t \leq N$.*

Moreover, for any cusp $\mathfrak{a}$ we have that $\Gamma_\mathfrak{a}$ is also an infinite cyclic group and we denote its generator by $\gamma_\mathfrak{a}$. To see this, if $\mathfrak{a} = \frac{a}{c}$ with $(a, c) = 1$ is a cusp of $\Gamma\backslash\mathbb{H}$ not equivalent to ∞, then there exists an $\sigma_\mathfrak{a} \in \mathrm{PSL}_2(\mathbb{Z})$ such that $\sigma_\mathfrak{a}\infty = \mathfrak{a}$. Indeed, there exists integers d and b such that $ad - bc = 1$ by Bézout's identity and then $\sigma_\mathfrak{a} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is such a matrix. It follows that $\Gamma_\mathfrak{a} = \sigma_\mathfrak{a}\Gamma_\infty\sigma_\mathfrak{a}^{-1}$ and since Γ_∞ is infinite cyclic so is $\Gamma_\mathfrak{a}$. We call any matrix $\sigma_\mathfrak{a} \in \mathrm{PSL}_2(\mathbb{Z})$ satisfying

$$\sigma_\mathfrak{a}\infty = \mathfrak{a} \quad \text{and} \quad \sigma_\mathfrak{a}^{-1}\gamma_\mathfrak{a}\sigma_\mathfrak{a} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix},$$

a **scaling matrix** for the cusp $\mathfrak{a}$. Note that $\sigma_\mathfrak{a}$ is determined up to composition on the right by an element of Γ_∞. Scaling matrices are useful because they allow us to transfer information at the cusp $\mathfrak{a}$ to the cusp at ∞. Let $\mathfrak{a}$ and $\mathfrak{b}$ be cusps of $\Gamma\backslash\mathbb{H}$ with scaling matrices $\sigma_\mathfrak{a}$ and $\sigma_\mathfrak{b}$ respectively. When investigating holomorphic forms, it will be useful to have a double coset decomposition for sets of the form $\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{b}$. This is referred to as the **Bruhat decomposition** for Γ:

Theorem 3.2.1 (Bruhat decomposition). *Let Γ be any congruence subgroup and let $\mathfrak{a}$ and $\mathfrak{b}$ be cusps of $\Gamma\backslash\mathbb{H}$ with scaling matrices $\sigma_\mathfrak{a}$ and $\sigma_\mathfrak{b}$ respectively. Then we have the disjoint decomposition*

$$\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{b} = \delta_{\mathfrak{a},\mathfrak{b}}\Omega_\infty \bigcup_{\substack{c \geq 1 \\ d \pmod{c}}} \Omega_{d/c},$$

where
$$\Omega_\infty = \Gamma_\infty\omega_\infty = \omega_\infty\Gamma_\infty = \Gamma_\infty\omega_\infty\Gamma_\infty \quad \textit{and} \quad \Omega_{d/c} = \Gamma_\infty\omega_{d/c}\Gamma_\infty,$$
for some $\omega_\infty = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ *and* $\omega_{d/c} = \left(\begin{smallmatrix} a & * \\ c & d \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ *with* $c \geq 1$ *if such a matrix exists otherwise* $\Omega_{d/c}$ *is empty. Moreover, the entries* a *and* d *of* $\omega_{d/c}$ *are determined modulo* c*.*

Proof. We first show that Ω_∞ is nonempty if and only if $\mathfrak{a} = \mathfrak{b}$. Indeed, if $\omega \in \Omega_\infty$ then $\omega = \sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{b}}$ for some $\gamma \in \Gamma$. Then
$$\gamma\mathfrak{b} = \sigma_{\mathfrak{a}}\omega\sigma_{\mathfrak{b}}^{-1}\mathfrak{b} = \sigma_{\mathfrak{a}}\omega\infty = \sigma_{\mathfrak{a}}\infty = \mathfrak{a}.$$
This shows that $\mathfrak{a} = \mathfrak{b}$. Conversely, suppose $\mathfrak{a} = \mathfrak{b}$. Then $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ contains $\sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \Gamma_\infty$ so that Ω_∞ is nonempty. So Ω_∞ is nonempty if and only if $\mathfrak{a} = \mathfrak{b}$. In this case, for any two elements $\omega = \sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{a}}$ and $\omega' = \sigma_{\mathfrak{a}}^{-1}\gamma'\sigma_{\mathfrak{a}}$ of Ω_∞, we have
$$\gamma'\gamma^{-1}\mathfrak{a} = \sigma_{\mathfrak{a}}\omega'\omega^{-1}\sigma_{\mathfrak{a}}^{-1}\mathfrak{a} = \sigma_{\mathfrak{a}}\omega'\omega^{-1}\infty = \sigma_{\mathfrak{a}}\mathfrak{a}.$$
Hence $\gamma'\gamma^{-1} \in \Gamma_{\mathfrak{a}}$ which implies $\omega'\omega^{-1} = \sigma_{\mathfrak{a}}^{-1}\gamma'\gamma^{-1}\sigma_{\mathfrak{a}} \in \sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \Gamma_\infty$. Therefore
$$\Omega_\infty = \Gamma_\infty\omega = \omega\Gamma_\infty = \Gamma_\infty\omega\Gamma_\infty,$$
where the latter two equalities hold because ω is a translation and translations commute. Every other element of $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ belongs to one of the double cosets $\Omega_{d/c}$ with $c \geq 1$ (since we are working in $\mathrm{PSL}_2(\mathbb{Z})$). The relation
$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & * \\ c & d \end{pmatrix}\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a+cn & * \\ c & d+cm \end{pmatrix},$$
shows that $\Omega_{d/c}$ is determined uniquely by c and $d \pmod{c}$. Moreover, this relation shows that a and d are determined modulo c. This completes the proof of the theorem. □

Notice that the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ implies
$$\Gamma_\infty\backslash\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}} = \delta_{\mathfrak{a},\mathfrak{b}}\omega_\infty \bigcup_{\substack{c\geq 1 \\ d \pmod{c}}} \omega_{d/c}\Gamma_\infty,$$
where it is understood that the coset $\omega_{d/c}\Gamma_\infty$ is empty if the double coset $\Omega_{d/c}$ is too. This shows that every element of $\Gamma_\infty\backslash\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ corresponds to a unique $(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}$ with $c \geq 1$, $d \in \mathbb{Z}$, and $(c,d) = 1$, and additionally the pair $(0,1)$ if and only if $\mathfrak{a} = \mathfrak{b}$ (this pair corresponds to ω_∞). Of course, this correspondence need not be surjective since many of the double cosets $\Omega_{d/c}$ may be empty. To track such c and d for which $\Omega_{d/c}$ is nonempty, let $\mathcal{C}_{\mathfrak{a},\mathfrak{b}}$ and $\mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ be the sets given by
$$\mathcal{C}_{\mathfrak{a},\mathfrak{b}} = \left\{c \geq 1 : \begin{pmatrix} * & * \\ c & * \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\right\} \quad \text{and} \quad \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c) = \left\{d \pmod{c} : \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\right\}.$$
Then $\mathcal{C}_{\mathfrak{a},\mathfrak{b}}$ and $\mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ are precisely the sets of c and d take modulo c such that $\Omega_{d/c}$ is nonempty.

Remark 3.2.3. *The Bruhat decomposition for* $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ *implies*
$$\Gamma_\infty\backslash\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}} = \delta_{\mathfrak{a},\mathfrak{b}}\omega_\infty \bigcup_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}} \\ d\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}} \omega_{d/c}\Gamma_\infty,$$
where none of the cosets $\omega_{d/c}\Gamma_\infty$ *are empty. In particular,* $\gamma = \left(\begin{smallmatrix} * & * \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty\backslash\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ *if and only if* (c,d) *is a pair with* $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$*,* $d \in \mathbb{Z}$*, and* $d \pmod{c} \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$*, or additionally* $(0,1)$ *if* $\mathfrak{a} = \mathfrak{b}$*.*

We will now introduce Kloosterman & Salié sums associated to cusps. We being with the Kloosterman sums. Let $\sigma_{\mathfrak{a}}$ and $\sigma_{\mathfrak{b}}$ be scaling matrices for the cusps $\mathfrak{a}$ and $\mathfrak{b}$ respectively. Then for any $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$ and $n, m \in \mathbb{Z}$, the **generalized Kloosterman sum** $K_{\mathfrak{a},\mathfrak{b}}(n, m, c)$ relative to $\mathfrak{a}$ and $\mathfrak{b}$ is defined by

$$K_{\mathfrak{a},\mathfrak{b}}(n, m, c) = \sum_{d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)} e^{\frac{2\pi i(an+\bar{a}m)}{c}},$$

where a has been determined by $ad - bc = 1$. This sum is well-defined by the Bruhat decomposition because a is determined modulo c. In general, $K_{\mathfrak{a},\mathfrak{b}}(n, m, c)$ is not independent of the scaling matrices $\sigma_{\mathfrak{a}}$ and $\sigma_{\mathfrak{b}}$. However, if $n = m = 0$ we trivially see by the Bruhat decomposition for Γ that

$$K_{\mathfrak{a},\mathfrak{b}}(0, 0, c) = \left|\left\{d \pmod{c} : \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}\right\}\right|,$$

which is independent of the scaling matrices $\sigma_{\mathfrak{a}}$ and $\sigma_{\mathfrak{b}}$. Moreover, for $\Gamma = \Gamma_1(1)$ we have $\mathfrak{a} = \mathfrak{b} = \infty$ and the Bruhat decomposition for $\Gamma_1(1)$ implies

$$K_{\infty,\infty}(n, m, c) = K(n, m, c),$$

is the usual Kloosterman sum. Therefore if $\mathfrak{a} = \mathfrak{b} = \infty$, we will suppress these dependencies accordingly. The Salié sums are defined in a similar manner. Let $\sigma_{\mathfrak{a}}$ and $\sigma_{\mathfrak{b}}$ be scaling matrices for the cusps $\mathfrak{a}$ and $\mathfrak{b}$ respectively. Then for any $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$, and $n, m \in \mathbb{Z}$, and Dirichlet character χ with conductor $q \mid c$, the **generalized Salié sum** $S_{\chi,\mathfrak{a},\mathfrak{b}}(n, m, c)$ relative to $\mathfrak{a}$ and $\mathfrak{b}$ is defined by

$$S_{\chi,\mathfrak{a},\mathfrak{b}}(n, m, c) = \sum_{d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)} \chi(a) e^{\frac{2\pi i(an+\bar{a}m)}{c}},$$

where a has been determined by $ad - bc = 1$. This sum is well-defined by the Bruhat decomposition because a is determined modulo c. Like the generalized Kloosterman sum, $S_{\mathfrak{a},\mathfrak{b}}(n, m, c)$ need not independent of the scaling matrices $\sigma_{\mathfrak{a}}$ and $\sigma_{\mathfrak{b}}$. Moreover, for $\Gamma = \Gamma_1(1)$ we have $\mathfrak{a} = \mathfrak{b} = \infty$ and the Bruhat decomposition for $\Gamma_1(1)$ implies

$$S_{\chi,\infty,\infty}(n, m, c) = S_{\chi}(n, m, c),$$

is the usual Salié sum. Therefore if $\mathfrak{a} = \mathfrak{b} = \infty$, we will suppress these dependencies accordingly.

3.3 The Hyperbolic Measure

We will also need to integrate over $\Gamma\backslash\mathbb{H}$. In order to do this, we require a measure on $\mathbb{H}$. Our choice of measure will be the **hyperbolic measure** $d\mu$ given by

$$d\mu = d\mu(z) = \frac{dx\,dy}{y^2}.$$

The most important property about the hyperbolic measure is that it is $\mathrm{GL}_2^+(\mathbb{R})$-invariant (see [DS05] for a proof):

Proposition 3.3.1. *The hyperbolic measure $d\mu$ is $\mathrm{GL}_2^+(\mathbb{R})$-invariant.*

As this fact will be used so frequently any time we integrate, we will not mention it explicitly. A particularly important fact is that if Γ is a congruence subgroup then $d\mu$ is Γ-invariant. One of the reasons this is useful is because we can apply the unfolding/folding method to many integrals. The most common instance is when we are integrating the sum $\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} f(\gamma z)$ of some holomorphic function $f(z)$ over a fundamental domain $\mathcal{F}_\Gamma$ for $\Gamma\backslash\mathbb{H}$. Indeed, $\mathbb{H} = \bigcup_{\gamma\in\Gamma}\gamma\mathcal{F}_\Gamma$ and so $\Gamma_\infty\backslash\mathbb{H} = \bigcup_{\gamma\in\Gamma_\infty\backslash\Gamma}\gamma\mathcal{F}_\Gamma$. Since $\mathcal{F}_\Gamma$ is a fundamental domain, the conditions of the unfolding/folding method are satisfied and it follows that

$$\int_{\mathcal{F}_\Gamma}\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} f(\gamma z)\,d\mu = \int_{\Gamma_\infty\backslash\mathbb{H}} f(z)\,d\mu,$$

provided either side is absolutely convergent. It is also worth highlighting another fact. Any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ acts as an automorphism of $\mathbb{H}$ which implies that it induces a bijection between $\alpha^{-1}\Gamma\alpha\backslash\mathbb{H}$ and $\Gamma\backslash\mathbb{H}$ and hence between the fundamental domains $\mathcal{F}_{\alpha^{-1}\Gamma\alpha}$ and $\mathcal{F}_\Gamma$. Thus the change of variables $z \to \alpha z$ transforms the fundamental domain $\mathcal{F}_\Gamma$ into $\mathcal{F}_{\alpha^{-1}\Gamma\alpha}$. Therefore

$$\int_{\mathcal{F}_\Gamma} f(z)\,d\mu = \int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}} f(\alpha z)\,d\mu,$$

provided either side is bounded. Now let us discuss the volume of $\Gamma\backslash\mathbb{H}$. We define the **volume** V_Γ of $\Gamma\backslash\mathbb{H}$ by

$$V_\Gamma = \int_{\mathcal{F}_\Gamma} d\mu.$$

In other words, V_Γ is the volume of the fundamental domain $\mathcal{F}_\Gamma$ with respect to the hyperbolic measure. Also, if $\mathcal{F}_\Gamma = \mathcal{F}$ we write $V_\Gamma = V$. Since the integrand is Γ-invariant, V_Γ is independent of the choice of fundamental domain. Using Proposition 3.2.2, we have

$$V = \int_{\mathcal{F}} d\mu = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{\sqrt{1-x^2}}^{\infty}\frac{dy\,dx}{y^2} = \int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{1}{\sqrt{1-x^2}}\,dx = \arcsin(x)\Big|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{\pi}{3}.$$

Therefore V is finite. There is also a simple relation between V_Γ and the index of Γ in $\mathrm{PSL}_2(\mathbb{Z})$:

$$V_\Gamma = [\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]V, \tag{3.1}$$

which follows immediately from Proposition 3.2.3. Moreover, V_Γ is finite for every congruence subgroup Γ by Equation (3.1) and that congruence subgroups have finite index in the modular group. A particularly nice application of this fact is that any integral of the form

$$\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\,d\mu,$$

is absolutely convergent provided $f(z)$ is bounded. That is, bounded functions are absolutely convergent over $\mathcal{F}_\Gamma$ with respect to $d\mu$. Moreover, we have a useful lemma:

Lemma 3.3.1. *Let Γ be a congruence subgroup and $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. If $\alpha^{-1}\Gamma\alpha \subseteq \mathrm{PSL}_2(\mathbb{Z})$, then $V_{\alpha^{-1}\Gamma\alpha} = V_\Gamma$ and $[\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\Gamma\alpha] = [\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]$.*

Proof. The first statement follows from the chain

$$V_{\alpha^{-1}\Gamma\alpha} = \int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}} d\mu = \int_{\mathcal{F}_\Gamma} d\mu = V_\Gamma,$$

where the middle equality is justified by making the change of variables $z \to \alpha^{-1}z$. The second statement is now immediate from Equation (3.1). $\square$

Chapter 4

The Theory of Holomorphic Forms

Holomorphic forms are special classes of functions on the upper half-space $\mathbb{H}$ of the complex plane. They are holomorphic, have a transformation law with respect to a congruence subgroup, and satisfy a growth condition. We will introduce these forms in a general context. Throughout we assume that all of our congruence subgroups are reduced at infinity.

4.1 Holomorphic Forms

Define $j(\gamma, z)$ by

$$j(\gamma, z) = (cz + d),$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ and $z \in \mathbb{H}$. There is a very useful property that $j(\gamma, z)$ satisfies. To state it, let $\gamma' = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$. Then

$$\gamma\gamma' = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a'a + b'c & a'b + b'b \\ c'a + d'c & c'b + d'd \end{pmatrix},$$

and we have

$$\begin{aligned} j(\gamma', \gamma z) j(\gamma, z) &= \left(c' \frac{az + b}{cz + d} + d' \right)(cz + d) \\ &= (c'(az + b) + d'(cz + d)) \\ &= (c'a + d'c)z + c'b + d'd) \\ &= j(\gamma'\gamma, z). \end{aligned}$$

In short,

$$j(\gamma'\gamma, z) = j(\gamma', \gamma z) j(\gamma, z),$$

and this is called the **cocycle condition** for $j(\gamma, z)$. For any integer $k \geq 1$ and any $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ we define the **slash operator** $|_{j,k}\gamma : C(\mathbb{H}) \to C(\mathbb{H})$ to be the linear operator given by

$$(f|_{j,k}\gamma)(z) = \deg(\gamma)^{k-1} j(\gamma, z)^{-k} f(\gamma z).$$

If j is clear from content we will suppress this dependence accordingly. Note that if $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, the slash operator takes the simpler form

$$(f|_{j,k}\gamma)(z) = j(\gamma, z)^{-k} f(\gamma z).$$

The cocycle condition implies that the slash operator is multiplicative. Indeed, if $\gamma, \gamma' \in \mathrm{GL}_2^+(\mathbb{Q})$, then

$$\begin{aligned}((f|_{j,k}\gamma')|_{j,k}\gamma)(z) &= \deg(\gamma)^{k-1}j(\gamma,z)^{-k}(f|_{j,k}\gamma')(\gamma z)\\ &= \deg(\gamma'\gamma)^{k-1}j(\gamma',\gamma z)^{-k}j(\gamma,z)^{-k}f(\gamma'\gamma z)\\ &= \deg(\gamma'\gamma)^{k-1}j(\gamma'\gamma,z)^{-k}f(\gamma'\gamma z) && \text{cocycle condition}\\ &= (f|_{j,k}\gamma'\gamma)(z).\end{aligned}$$

If an operator commutes with the slash operators $|_{j,k}\gamma$ for every $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, we say that it is **invariant**. We will now introduce holomorphic forms. Let Γ be a congruence subgroup of level N and let χ be a Dirichlet character of conductor $q \mid N$, and set $\chi(\gamma) = \chi(d)$ for all for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. We say that a function $f : \mathbb{H} \to \mathbb{C}$ is **holomorphic form** (or **modular form**) on $\Gamma\backslash\mathbb{H}$ of **weight** k, **level** N, and **character** χ if the following properties are satisfied:

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

(ii) $(f|_{j,k}\gamma)(z) = \chi(\gamma)f(z)$ for all $\gamma \in \Gamma$.

(iii) $(f|_{j,k}\alpha)(z) = O(1)$ for all $\alpha \in \mathrm{PSL}_2(\mathbb{Z})$ (or equivalently $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$).

We say f is a **(holomorphic) cusp form** if the additional property is satisfied:

(iv) For all cusps $\mathfrak{a}$ and any $y > 0$, we have

$$\int_0^1 (f|_k\sigma_{\mathfrak{a}})(x+iy)\,dx = 0.$$

Property (ii) is called the **modularity condition** and we say f is **modular**. In particular, f is a function on $\mathcal{F}_\Gamma$. The modularity condition can equivalently be expressed as

$$f(\gamma z) = \chi(\gamma)j(\gamma,z)^k f(z).$$

Property (iii) is called the **growth condition** for holomorphic forms and we say f is **holomorphic at the cusps**. Clearly we only need to verify the growth condition on a set of scaling matrices for the cusps. To see the equivalence in the growth condition, we claim that every $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ is of the form $\alpha = \gamma\eta$ for some $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ and $\eta \in \mathrm{GL}_2^+(\mathbb{Q})$ of the form $\eta = \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$. To see this, if $c = 0$ the claim is obvious. For $c \neq 0$, let $r \geq 1$ be such that $r\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Z})$ and set $a' = \frac{a}{(a,c)}$ and $c' = \frac{c'}{(a,c)}$ so that $a', c' \in \mathbb{Z}$ with $(a', c') = 1$. Then there exists $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ with $\gamma^{-1} = \left(\begin{smallmatrix} * & * \\ -c' & a' \end{smallmatrix}\right)$. Moreover, $\gamma^{-1}r\alpha = \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Z})$. Upon setting $\eta = \gamma^{-1}\alpha$, the claim is complete. From the decomposition $\alpha = \gamma\eta$, the cocycle condition gives

$$j(\alpha, z) = j(\gamma, \eta z),$$

and it follows that $(f|_{j,k}\alpha)(z) = O(1)$ for all $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ which proves the forward implication. The reverse implication is trivial since $\mathrm{PSL}_2(\mathbb{Z}) \subset \mathrm{GL}_2^+(\mathbb{Q})$. Holomorphic forms also admit Fourier series. Indeed, modularity implies

$$f(z+1) = f\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} z\right) = f(z),$$

so that f is 1-periodic. Let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the $\mathfrak{a}$ cusp. As Lemma 3.1.1 implies $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$ is a congruence subgroup, it follows by the cocycle condition that $f|_k\sigma_{\mathfrak{a}}$ is a holomorphic form on $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}\backslash\mathbb{H}$

of the same weight and character as f. In particular, $f|_k\sigma_{\mathfrak{a}}$ is 1-periodic. Note that this means we only need to verify the growth condition as $y \to \infty$. As $f|_k\sigma_{\mathfrak{a}}$ is 1-periodic it admits a Fourier series given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = \sum_{n \ge 0} a_{\mathfrak{a}}(n,y)e^{2\pi inx},$$

where the sum is only over $n \ge 0$ because holomorphy at the cusps implies that $f|_k\sigma_{\mathfrak{a}}$ is bounded. As f is smooth (since it is holomorphic), it converges uniformly to its Fourier series everywhere. We can simplify the Fourier coefficients $a_{\mathfrak{a}}(n,y)$. To see this, since $f|_k\sigma_{\mathfrak{a}}$ is holomorphic it satisfies the first order Cauchy-Riemann equations so that

$$\frac{1}{2}\left(\frac{\partial f|_k\sigma_{\mathfrak{a}}}{\partial x} + i\frac{\partial f|_k\sigma_{\mathfrak{a}}}{\partial y}\right) = 0.$$

Substituting in the Fourier series and equating coefficients we obtain the ODE

$$2\pi n a_{\mathfrak{a}}(n,y) + a_{\mathfrak{a},y}(n,y) = 0,$$

Solving this ODE by separation of variables, we see that there exists an $a_{\mathfrak{a}}(n)$ such that

$$a_{\mathfrak{a}}(n,y) = a_{\mathfrak{a}}(n)e^{-2\pi ny}.$$

The coefficients $a_{\mathfrak{a}}(n)$ are the only part of the Fourier series depending on the implicit congruence subgroup Γ. Using these coefficients instead, f admits a **Fourier series** at the $\mathfrak{a}$ cusp given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = \sum_{n \ge 0} a_{\mathfrak{a}}(n)e^{2\pi inz},$$

with **Fourier coefficients** $a_{\mathfrak{a}}(n)$. If $\mathfrak{a} = \infty$, we will drop this dependence and in this case $f|_k\sigma_{\mathfrak{a}} = f$. Moreover, property (iv) implies that f is a cusp form if and only if $a_{\mathfrak{a}}(n) = 0$ for every cusp $\mathfrak{a}$. We can also easily derive a bound for the size of the Fourier coefficients of cusp forms. To see this, note that $\left|(f|_k\sigma_{\mathfrak{a}})(z)\mathrm{Im}(z)^{\frac{k}{2}}\right|$ is $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$-invariant by the modularity of $f|_k\sigma_{\mathfrak{a}}$, the cocycle condition, the identity $\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{a}}z)^{\frac{k}{2}} = \frac{\mathrm{Im}(z)^{\frac{k}{2}}}{|j(\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{a}},z)|^k}$, and that $|\chi(\gamma)| = 1$. Moreover, this function is bounded on $\mathcal{F}_{\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}}$ because f is a cusp form. Then $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$-invariance implies $\left|(f|_k\sigma_{\mathfrak{a}})(z)\mathrm{Im}(z)^{\frac{k}{2}}\right|$ is bounded on $\mathbb{H}$. From the definition of Fourier series, it follows that

$$a_{\mathfrak{a}}(n)\mathrm{Im}(z)^{\frac{k}{2}} = \int_0^1 (f|_k\sigma_{\mathfrak{a}})(z)\mathrm{Im}(z)^{\frac{k}{2}}e^{-2\pi inz}\,dx \ll \int_0^1 e^{2\pi ny}\,dx \ll e^{2\pi ny}.$$

Upon setting $y = \frac{1}{n}$, the last expression is constant and we obtain

$$a_{\mathfrak{a}}(n) \ll n^{\frac{k}{2}}.$$

This bound is known as the **Hecke bound** for holomorphic forms. It follows from the Hecke bound and the Taylor series of the k-th derivative of $\frac{e^y}{1-e^y}$ that

$$(f|_k\sigma_{\mathfrak{a}})(z) = O\left(\sum_{n \ge 1} n^{\frac{k}{2}}e^{-2\pi ny}\right) = O\left(\sum_{n \ge 1} n^k e^{-2\pi ny}\right) = O(e^{-2\pi y}).$$

This implies $(f|_k\sigma_{\mathfrak{a}})(z)$ exhibits rapid decay. Accordingly, we say that f exhibits **rapid decay at the cusps**. Observe that $(f|_k\sigma_{\mathfrak{a}})$ is then bounded on $\mathbb{H}$ and, in particular, f is bounded on $\mathbb{H}$.

4.2 Poincaré & Eisenstein Series

Let Γ be a congruence subgroup of level N. We will introduce two important classes of holomorphic forms on $\Gamma\backslash\mathbb{H}$ namely the Poincaré and Eisenstein series. Let $m \ge 0$, $k \ge 4$, χ be a Dirichlet character with conductor $q \mid N$, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. We define the m-th **(holomorphic) Poincaré series** $P_{m,k,\chi,\mathfrak{a}}(z)$ of weight k with character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp by

$$P_{m,k,\chi,\mathfrak{a}}(z) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)j(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}.$$

We call m the **index** of $P_{m,k,\chi,\mathfrak{a}}(z)$. If χ is the trivial character or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly.

Remark 4.2.1. *The reason why we restrict to $k \ge 4$ is because for $k = 0, 2$ the Poincaré series need not converge (see Proposition B.8.1).*

We first verify that $P_{m,k,\chi,\mathfrak{a}}(z)$ is well-defined. It suffices to show that the summands are independent of the representatives γ and $\sigma_{\mathfrak{a}}$. To see that $\overline{\chi}(\gamma)$ is independent of γ, recall that $\Gamma_{\mathfrak{a}} = \sigma_{\mathfrak{a}}\Gamma_\infty\sigma_{\mathfrak{a}}^{-1}$ and let $\gamma' = \sigma_{\mathfrak{a}}\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma$ with $\eta_\infty = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_\infty$. Then

$$\overline{\chi}(\gamma') = \overline{\chi}(\sigma_{\mathfrak{a}}\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma) = \overline{\chi}(\sigma_{\mathfrak{a}})\chi(\eta_\infty)\overline{\chi}(\sigma_{\mathfrak{a}})^{-1}\overline{\chi}(\gamma) = \overline{\chi}(\gamma),$$

verifying that $\overline{\chi}(\gamma)$ is independent of the representative γ. As the set of representatives for the scaling matrix $\sigma_{\mathfrak{a}}$ is $\sigma_{\mathfrak{a}}\Gamma_\infty$ and the set of representatives for γ is $\Gamma_{\mathfrak{a}}\gamma$, the set of representatives for $\sigma_{\mathfrak{a}}^{-1}\gamma$ is $\Gamma_\infty\sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\gamma$. But as $\Gamma_{\mathfrak{a}} = \sigma_{\mathfrak{a}}\Gamma_\infty\sigma_{\mathfrak{a}}^{-1}$, this set of representatives is $\Gamma_\infty\sigma_{\mathfrak{a}}^{-1}\gamma$ and therefore it remains to verify independence from multiplication on the left by an element of Γ_∞ namely η_∞. The cocycle relation gives

$$j(\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma, z) = j(\eta_\infty, \sigma_{\mathfrak{a}}^{-1}\gamma z)j(\sigma_{\mathfrak{a}}^{-1}\gamma, z) = j(\sigma_{\mathfrak{a}}^{-1}\gamma, z),$$

where the last equality follows because $j(\eta_\infty, \sigma_{\mathfrak{a}}^{-1}\gamma z) = 1$. This verifies that $j(\sigma_{\mathfrak{a}}^{-1}\gamma, z)$ is independent of the representatives γ and $\sigma_{\mathfrak{a}}$. Moreover, we have

$$e^{2\pi i m\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma z} = e^{2\pi i m(\sigma_{\mathfrak{a}}^{-1}\gamma z+n)} = e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma}e^{2\pi i mnz} = e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma},$$

which verifies that $e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}$ is independent of the representatives γ and $\sigma_{\mathfrak{a}}$. Therefore $P_{m,k,\chi,\mathfrak{a}}(z)$ is well-defined. To see that $P_{m,k,\chi,\mathfrak{a}}(z)$ is holomorphic on $\mathbb{H}$, first note that $|e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}| = e^{-2\pi m\mathrm{Im}(\mathfrak{a}^{-1}\gamma z)} < 1$. Then the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma$ gives

$$P_{m,k,\chi,\mathfrak{a}}(z) \ll \sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{1}{|cz+d|^k}.$$

As $k \ge 4$, this latter series is locally absolutely uniformly convergent for $z \in \mathbb{H}$ by Proposition B.8.1. Hence $P_{m,k,\chi,\mathfrak{a}}(z)$ does too and so it is holomorphic on $\mathbb{H}$. We now verify modularity for $P_{m,k,\chi,\mathfrak{a}}(z)$. This is just a computation:

$$
\begin{aligned}
P_{m,k,\chi,\mathfrak{a}}(\gamma z) &= \sum_{\gamma' \in \Gamma_{\mathfrak{a}} \backslash \Gamma} \overline{\chi}(\gamma') j(\sigma_{\mathfrak{a}}^{-1}\gamma', \gamma z)^{-k} j(\gamma, z)^{-k} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \sum_{\gamma' \in \Gamma_{\mathfrak{a}} \backslash \Gamma} \overline{\chi}(\gamma') \left(\frac{j(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)}{j(\gamma, z)} \right)^{-k} j(\gamma, z)^{-k} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= j(\gamma, z)^{k} \sum_{\gamma' \in \Gamma_{\mathfrak{a}} \backslash \Gamma} \overline{\chi}(\gamma') j(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)^{-k} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma) j(\gamma, z)^{k} \sum_{\gamma' \in \Gamma_{\mathfrak{a}} \backslash \Gamma} \overline{\chi}(\gamma'\gamma) j(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)^{-k} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma) j(\gamma, z)^{k} \sum_{\gamma' \in \Gamma_{\mathfrak{a}} \backslash \Gamma} \overline{\chi}(\gamma') j(\sigma_{\mathfrak{a}}^{-1}\gamma', z)^{-k} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma' z} \\
&= \chi(\gamma) P_{m,k,\chi,\mathfrak{a}}(z),
\end{aligned}
$$

where in the second line we have used the cocycle condition and in the second to last line we have used that $\gamma' \to \gamma'\gamma^{-1}$ is a bijection on Γ. To verify the growth condition, we will need a technical lemma:

Lemma 4.2.1. *Let $a, b > 0$ be reals and consider the half-strip*

$$
S_{a,b} = \{z \in \mathbb{H} : |x| \le a \text{ and } y \ge b\}.
$$

Then there is a $\delta \in (0,1)$ such that

$$
|nz + m| \ge \delta |ni + m|,
$$

for all $n, m \in \mathbb{Z}$ and all $z \in S_{a,b}$.

Proof. If $n = 0$ then any δ is sufficient and this δ is independent of z. If $n \neq 0$, then the desired inequality is equivalent to

$$
\left| \frac{z + \frac{m}{n}}{i + \frac{n}{m}} \right| \ge \delta.
$$

So consider the function

$$
f(z, r) = \left| \frac{z + r}{i + r} \right|,
$$

for $z \in S_{a,b}$ and $r \in \mathbb{R}$. It suffices to show $f(z,r) \ge \delta$. As $z \in \mathbb{H}$, $z - r \neq 0$ so that $f(z,r)$ is continuous and positive on $S_{a,b} \times \mathbb{R}$. Now let $Y > b$ and consider the region

$$
S_{a,b}^{Y} = \{z \in \mathbb{H} : |x| \le a \text{ and } b \le y \le Y\}.
$$

We claim that there exists a Y such that if $y > Y$ and $|x| > Y$ then $f(z,r)^2 > \frac{1}{4}$. Indeed, we compute

$$
f(z,r)^2 = \frac{(z + r)(\overline{z} + r)}{(i + r)(-i + r)} = \frac{|z|^2 + 2xr + r^2}{1 + r^2} \ge \frac{y + r^2}{1 + r^2},
$$

where in the inequality we have used the bound $|z|^2 \ge y$ and that x is bounded. Now $\frac{r^2}{1+r^2} \to 1$ as $r \to \pm\infty$ so there exists a Y such that $|r| > Y$ implies $\frac{r^2}{1+r^2} \ge \frac{1}{4}$. Then

$$
\frac{y + r^2}{1 + r^2} \ge \frac{y}{1 + r^2} + \frac{r^2}{1 + r^2} \ge \frac{y}{1 + r^2} + \frac{1}{4} > \frac{1}{4}.
$$

It follows that $f(z,r) > \frac{1}{2}$ outside of $S_{a,b}^{Y} \times [-Y, Y]$. But this latter region is compact and so $f(z,r)$ obtains a minimum δ' on it. Setting $\delta = \min\left(\frac{1}{2}, \delta'\right)$ completes the proof. $\square$

We can now verify the growth condition for $P_{m,k,\chi,\mathfrak{a}}(z)$. Let $\sigma_{\mathfrak{b}}$ be a scaling matrix for the cusp $\mathfrak{b}$. Then the bound $|e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{b}}z}| = e^{-2\pi m\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{b}}z)} < 1$, cocycle condition, and Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ together give

$$j(\sigma_{\mathfrak{b}}, z)^{-k}P_{m,k,\chi,\mathfrak{a}}(\sigma_{\mathfrak{b}}z) \ll \sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{1}{|cz+d|^k}.$$

Now decompose this last sum as

$$\sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{1}{|cz+d|^k} = \sum_{d\neq 0}\frac{1}{d^k} + \sum_{c\neq 0}\sum_{d\in\mathbb{Z}}\frac{1}{|cz+d|^k} = 2\sum_{d\geq 1}\frac{1}{d^k} + 2\sum_{c\geq 1}\sum_{d\in\mathbb{Z}}\frac{1}{|cz+d|^k}.$$

Since the first sum is absolutely uniformly bounded, it suffices to show that the double sum is too. To see this, let $y \geq 1$ and δ be as in Lemma 4.2.1. Then for any integer $N \geq 1$ we can write

$$\begin{aligned}
\sum_{c\geq 1}\sum_{d\in\mathbb{Z}}\frac{1}{|cz+d|^k} &= \sum_{c+|d|\leq N}\frac{1}{|cz+d|^k} + \sum_{c+|d|>N}\frac{1}{|cz+d|^k} \\
&\leq \sum_{c+|d|\leq N}\frac{1}{|cz+d|^k} + \sum_{c+|d|>N}\frac{1}{(\delta|ci+d|)^k} \\
&\leq \sum_{c+|d|\leq N}\frac{1}{|cz+d|^k} + \frac{1}{\delta^k}\sum_{c+|d|>N}\frac{1}{|ci+d|^k}.
\end{aligned}$$

As $\sum_{c\geq 1}\sum_{d\in\mathbb{Z}}\frac{1}{|ci+d|^k}$ converges by Proposition B.8.1, the second sum tends to zero as $N\to\infty$. As for the first sum, it is finite and each term is bounded. Thus the double sum is absolutely uniformly bounded. This verifies the growth condition. We collect this work as a theorem:

Theorem 4.2.1. *Let $m \geq 0$, $k \geq 4$, χ be a Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. The Poincaré series*

$$P_{m,k,\chi,\mathfrak{a}}(z) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma} j(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z},$$

is a weight k holomorphic form with character χ on $\Gamma\backslash\mathbb{H}$.

For $m = 0$, we write $E_{k,\chi,\mathfrak{a}}(z) = P_{0,k,\chi,\mathfrak{a}}(z)$ and call $E_{k,\chi,\mathfrak{a}}(z)$ the **(holomorphic) Eisenstein series** of weight k and character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp. It is defined by

$$E_{k,\chi,\mathfrak{a}}(z) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma}\overline{\chi}(\gamma)j(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}.$$

If χ is the trivial character or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly. In particular, we have already verified the following theorem:

Theorem 4.2.2. *Let $k \geq 4$, χ be Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. The Eisenstein series*

$$E_{k,\chi,\mathfrak{a}}(z) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma}\overline{\chi}(\gamma)j(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k},$$

is a weight k holomorphic form with character χ on $\Gamma\backslash\mathbb{H}$.

We will now compute the Fourier series of the Poincaré series with positive index:

Proposition 4.2.1. *Let $m \geq 1$, $k \geq 4$, χ be Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ and $\mathfrak{b}$ be cusps of $\Gamma\backslash\mathbb{H}$. The Fourier series of $P_{m,k,\chi,\mathfrak{a}}(z)$ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{b}$ cusp is given by*

$$(P_{m,k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \sum_{t \geq 1}\left(\delta_{\mathfrak{a},\mathfrak{b}}\delta_{m,t} + \left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1}\sum_{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(m,t,c)\right)e^{2\pi itz}.$$

Proof. From the cocycle condition, the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$, and Remark 3.2.3, we have

$$(P_{m,k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \delta_{\mathfrak{a},\mathfrak{b}}e^{2\pi imz} + \sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}, d \in \mathbb{Z} \\ d \pmod{c} \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}}\overline{\chi}(d)\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+cd}\right)}}{(cz+d)^k},$$

where a has been determined modulo c by $ad-bc=1$ and we have used the fact that

$$\frac{a}{c}-\frac{1}{c^2z+cd} = \frac{az+b}{cz+d}.$$

Summing over all pairs (c,d) with $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$, $d \in \mathbb{Z}$, and $d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ is the same as summing over all triples (c,ℓ,r) with $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$, $\ell \in \mathbb{Z}$, and r taken modulo c with $r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$. Indeed, this is seen by writing $d = c\ell+r$. Moreover, since $ad-bc=1$ we have $a(c\ell+r)-bc=1$ which further implies that $ar \equiv 1 \pmod{c}$. So we may take a to be the inverse for r modulo c. Then

$$\begin{aligned}
\sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}, d \in \mathbb{Z} \\ d \pmod{c} \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}}\overline{\chi}(d)\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+cd}\right)}}{(cz+d)^k} &= \sum_{(c,\ell,r)}\overline{\chi}(c\ell+r)\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+c^2\ell+cr}\right)}}{(cz+c\ell+r)^k} \\
&= \sum_{(c,\ell,r)}\overline{\chi}(r)\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+c^2\ell+cr}\right)}}{(cz+c\ell+r)^k} \\
&= \sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}} \\ r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}}\sum_{\ell \in \mathbb{Z}}\overline{\chi}(r)\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+c^2\ell+cr}\right)}}{(cz+c\ell+r)^k} \\
&= \sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}} \\ r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}}\overline{\chi}(r)\sum_{\ell \in \mathbb{Z}}\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+c^2\ell+cr}\right)}}{(cz+c\ell+r)^k},
\end{aligned}$$

where on the right-hand side it is understood that we are summing over all triples (c,ℓ,r) with the prescribed properties and the second line holds since χ has conductor diving the level and $d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ is determined modulo c. Now let

$$I_{c,r}(z) = \sum_{\ell \in \mathbb{Z}}\frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+c^2\ell+cr}\right)}}{(cz+c\ell+r)^k}.$$

We will apply the Poisson summation formula to $I_{c,r}(z)$. This is possible since the summands are absolutely integrable because they exhibit polynomial decay of order $k>1$ and $I_{c,r}(z)$ is holomorphic (because

$(P_{m,k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z)$ is). By the identity theorem it suffices to apply the Poisson summation formula for $z = iy$ with $y > 0$. Accordingly, let $f(x)$ be given by

$$f(x) = \frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2x+cr+ic^2y}\right)}}{(cx+r+icy)^k}.$$

As we have just noted, $f(x)$ is absolutely integrable on $\mathbb{R}$. We compute the Fourier transform:

$$(\mathcal{F}f)(t) = \int_{\mathbb{R}} f(x)e^{-2\pi itx}\,dx = \int_{\mathbb{R}} \frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2x+cr+ic^2y}\right)}}{(cx+r+icy)^k} e^{-2\pi itx}\,dx.$$

Complexify the integral to get

$$\int_{\mathrm{Im}(z)=0} \frac{e^{2\pi im\left(\frac{a}{c}-\frac{1}{c^2z+cr+ic^2y}\right)}}{(cz+r+icy)^k} e^{-2\pi itz}\,dz.$$

Now make the change of variables $z \to z - \frac{r}{c} - icy$ to obtain

$$e^{2\pi im\frac{a}{c}+2\pi it\frac{r}{c}-2\pi ty} \int_{\mathrm{Im}(z)=y} \frac{e^{-\frac{2\pi im}{c^2z}}}{(cz)^k} e^{-2\pi itz}\,dz.$$

The integrand is meromorphic with a pole at $z = 0$. Moreover, we have

$$\frac{1}{(cz)^k} \ll \frac{1}{|cz|^k}, \quad e^{-\frac{2\pi im}{c^2z}} \ll e^{-\frac{2\pi m\mathrm{Im}(z)}{|cz|^2}}, \quad \text{and} \quad e^{-2\pi itz} \ll e^{2\pi ty}.$$

The first expression has polynomial decay, the second expression is bounded, and the third expression exhibits rapid decay if and only if $t < 0$ and when $t = 0$ it is bounded. So when $t \le 0$ we may take the limit as $\mathrm{Im}(z) \to \infty$ by shifting the line of integration and conclude that the integral vanishes. It remains to compute the integral for $t \ge 1$. To do this, make the change of variables $z \to -\frac{z}{2\pi it}$ to the last integral to rewrite it as

$$\begin{aligned}
-\frac{1}{2\pi it}\int_{(2\pi ty)} \frac{e^{-\frac{4\pi^2mt}{c^2z}}}{\left(-\frac{cz}{2\pi it}\right)^k} e^z\,dz &= -\frac{1}{2\pi it}\int_{(2\pi ty)} \left(-\frac{2\pi it}{cz}\right)^k e^{z-\frac{4\pi^2mt}{c^2z}}\,dz \\
&= \frac{(-2\pi it)^{k-1}}{c^k}\int_{(2\pi ty)} z^{-k}e^{z-\frac{4\pi^2mt}{c^2z}}\,dz \\
&= \frac{(-2\pi it)^{k-1}}{c^k}\int_{-\infty}^{(0^+)} z^{-k}e^{z-\frac{4\pi^2mt}{c^2z}}\,dz \\
&= \frac{2\pi i^{-k}}{c}\left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1} J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right),
\end{aligned}$$

where in the second to last line we have homotoped the line of integration to a Hankel contour about the negative real axis and in the last line we have used the Schläflin integral representation for the J-Bessel function (see Appendix B.6). So in total we obtain

$$(\mathcal{F}f)(t) = \begin{cases} \left(\frac{2\pi i^{-k}}{c}\left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1} J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right) e^{2\pi im\frac{a}{c}+2\pi it\frac{r}{c}}\right) e^{-2\pi ty} & \text{if } t \ge 1, \\ 0 & \text{if } t \le 0. \end{cases}$$

By the Poisson summation formula and the identity theorem, we have

$$I_{c,r}(z) = \sum_{t\geq 1}\left(\frac{2\pi i^{-k}}{c}\left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1} J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)e^{2\pi i m\frac{a}{c}+2\pi i t\frac{r}{c}}\right)e^{2\pi i t z},$$

for all $z \in \mathbb{H}$. Plugging this back into the Poincaré series gives a form of the Fourier series:

$$\begin{aligned}(P_{m,k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) &= \delta_{\mathfrak{a},\mathfrak{b}}e^{2\pi i m z} + \sum_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}\\ r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}}\overline{\chi}(r)\sum_{t\geq 1}\left(\frac{2\pi i^{-k}}{c}\left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1} J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)e^{2\pi i m\frac{a}{c}+2\pi i t\frac{r}{c}}\right)e^{2\pi i m z}\\
&= \sum_{t\geq 1}\left(\delta_{\mathfrak{a},\mathfrak{b}}\delta_{m,t} + \left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1}\sum_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}\\ r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}}\overline{\chi}(r)\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)e^{2\pi i m\frac{a}{c}+2\pi i t\frac{r}{c}}\right)e^{2\pi i t z}\\
&= \sum_{t\geq 1}\left(\delta_{\mathfrak{a},\mathfrak{b}}\delta_{m,t} + \left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1}\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)\sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(r)e^{2\pi i m\frac{a}{c}+2\pi i t\frac{r}{c}}\right)e^{2\pi i t z}.\end{aligned}$$

We will simplify the innermost sum. Since a is the inverse for r modulo c, the innermost sum above becomes

$$\sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(r)e^{2\pi i m\frac{a}{c}+2\pi i t\frac{r}{c}} = \sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(\overline{a})e^{2\pi i m\frac{a}{c}+2\pi i t\frac{\overline{a}}{c}} = \sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\chi(a)e^{\frac{2\pi i(am+\overline{a}t)}{c}} = S_{\chi,\mathfrak{a},\mathfrak{b}}(m,t,c).$$

So at last, we obtain our desired Fourier series:

$$(P_{m,k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \sum_{t\geq 1}\left(\delta_{\mathfrak{a},\mathfrak{b}}\delta_{m,t} + \left(\frac{\sqrt{t}}{\sqrt{m}}\right)^{k-1}\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{mt}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(m,t,c)\right)e^{2\pi i t z}.$$

□

An immediate consequence of Proposition 4.2.1 is that the Poincaré series $P_{m,k,\chi,\mathfrak{a}}(z)$ with positive index are cusp forms. In a similar manner, we can obtain the Fourier series of the Eisenstein series too:

Proposition 4.2.2. *Let $k \geq 4$, χ be Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ and $\mathfrak{b}$ be cusps of $\Gamma\backslash\mathbb{H}$. The Fourier series of $E_{k,\chi,\mathfrak{a}}(z)$ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{b}$ cusp is given by*

$$(E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \sum_{t\geq 0}\left(\delta_{\mathfrak{a},\mathfrak{b}} + \sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{(-2\pi i t)^k}{(k-1)!c^k}S_{\chi,\mathfrak{a},\mathfrak{b}}(0,t,c)\right)e^{2\pi i t z}.$$

Proof. From the cocycle condition, the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$, and Remark 3.2.3, we have

$$(E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \delta_{\mathfrak{a},\mathfrak{b}} + \sum_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}},d\in\mathbb{Z}\\ d \pmod{c}\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}}\overline{\chi}(d)\frac{1}{(cz+d)^k},$$

where a has been determined modulo c by $ad-bc=1$ and we have used the fact that

$$\frac{a}{c} - \frac{1}{c^2z+cd} = \frac{az+b}{cz+d}.$$

Summing over all pairs (c,d) with $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$, $d \in \mathbb{Z}$, and $d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ is the same as summing over all triples (c,ℓ,r) with $c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}$, $\ell \in \mathbb{Z}$, and r taken modulo c with $r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$. Indeed, this is seen by writing $d = c\ell + r$. Moreover, since $ad - bc = 1$ we have $a(c\ell + r) - bc = 1$ which further implies that $ar \equiv 1 \pmod{c}$. So we may take a to be the inverse for r modulo c. Then

$$\begin{aligned}
\sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}}, d \in \mathbb{Z} \\ d \pmod{c} \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}} \overline{\chi}(d) \frac{1}{(cz+d)^k} &= \sum_{(c,\ell,r)} \overline{\chi}(c\ell + r) \frac{1}{(cz + c\ell + r)^k} \\
&= \sum_{(c,\ell,r)} \overline{\chi}(r) \frac{1}{(cz + c\ell + r)^k} \\
&= \sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}} \\ r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}} \sum_{\ell \in \mathbb{Z}} \overline{\chi}(r) \frac{1}{(cz + c\ell + r)^k} \\
&= \sum_{\substack{c \in \mathcal{C}_{\mathfrak{a},\mathfrak{b}} \\ r \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)}} \overline{\chi}(r) \sum_{\ell \in \mathbb{Z}} \frac{1}{(cz + c\ell + r)^k},
\end{aligned}$$

where on the right-hand side it is understood that we are summing over all triples (c,ℓ,r) with the prescribed properties and the second line holds since χ has conductor diving the level and $d \in \mathcal{D}_{\mathfrak{a},\mathfrak{b}}(c)$ is determined modulo c. Now let

$$I_{c,r}(z) = \sum_{\ell \in \mathbb{Z}} \frac{1}{(cz + c\ell + r)^k}.$$

We apply the Poisson summation formula to $I_{c,r}(z)$. This is allowed since the summands are absolutely integrable as they exhibit polynomial decay of order $k > 1$, and $I_{c,r}(z)$ is holomorphic (because $(E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z)$ is). By the identity theorem it suffices to apply the Poisson summation formula for $z = iy$ with $y > 0$. So let $f(x)$ be given by

$$f(x) = \frac{1}{(cx + r + icy)^k}.$$

As we have just noted, $f(x)$ is absolutely integrable on $\mathbb{R}$. We compute the Fourier transform:

$$(\mathcal{F}f)(t) = \int_{\mathbb{R}} f(x) e^{-2\pi i t x}\, dx = \int_{\mathbb{R}} \frac{e^{-2\pi i t x}}{(cx + r + icy)^k}\, dx.$$

Complexify the integral to get

$$\int_{\mathrm{Im}(z)=0} \frac{e^{-2\pi i t z}}{(cz + r + icy)^k}\, dz.$$

Now make the change of variables $z \to z - \frac{r}{c} - icy$ to obtain

$$e^{2\pi i t \frac{r}{c} - 2\pi t y} \int_{\mathrm{Im}(z)=y} \frac{e^{-2\pi i t z}}{(cz)^k}\, dz.$$

The integrand is meromorphic with a pole at $z = 0$. Moreover, we have

$$\frac{1}{(cz)^k} \ll \frac{1}{|cz|^k} \quad \text{and} \quad e^{-2\pi i t z} \ll e^{2\pi t y}.$$

The first expression has polynomial decay while the expression exhibits rapid decay if and only if $t < 0$ and when $t = 0$ it is bounded. So when $t \leq 0$ we may take the limit as $\mathrm{Im}(z) \to \infty$ by shifting the line of integration and conclude that the integral vanishes. It remains to compute the integral for $t \geq 1$. To do this, make the change of variables $z \to -\frac{z}{2\pi it}$ to the last integral to rewrite it as

$$-\frac{1}{2\pi it}\int_{(2\pi ty)} \frac{e^z}{\left(-\frac{cz}{2\pi it}\right)^k}\,dz = -\frac{1}{2\pi it}\int_{(2\pi ty)} \left(-\frac{2\pi it}{cz}\right)^k e^z\,dz = \frac{(-2\pi it)^{k-1}}{c^k}\int_{(2\pi ty)} \frac{e^z}{z^k}\,dz.$$

The integrand of the last integral has a pole of order k at $z = 0$. To find the residue, the Laurent series of $\frac{e^z}{z^k}$ is

$$\frac{e^z}{z^k} = \sum_{n\geq 0}\frac{z^{n-k}}{n!},$$

and thus the residue of the integrand is $\frac{1}{(k-1)!}$. In shifting the line of integration to $(-x)$, for some $x > 0$, we pass by this pole and obtain

$$\frac{(-2\pi it)^k}{(k-1)!c^k} + \int_{(-x)} \frac{e^z}{z^k}\,dz.$$

Moreover, we have

$$\frac{1}{z^k} \ll \frac{1}{|z|^k} \quad \text{and} \quad e^z \ll e^x.$$

The first expression has polynomial decay while the second expression exhibits rapid decay provided $x < 0$. Therefore we make take the limit as $x \to \infty$ by shifting the line of integration again and conclude that the latter integral vanishes. Altogether, we have show that

$$(\mathcal{F}f)(t) = \begin{cases} \left(\frac{(-2\pi it)^k}{(k-1)!c^k}e^{2\pi it\frac{r}{c}}\right)e^{-2\pi ty} & \text{if } t \geq 1, \\ 0 & \text{if } t \leq 0. \end{cases}$$

By the Poisson summation formula and the identity theorem, we have

$$I_{c,r}(z) = \sum_{t\geq 1}\left(\frac{(-2\pi it)^k}{(k-1)!c^k}e^{2\pi it\frac{r}{c}}\right)e^{2\pi itz},$$

for all $z \in \mathbb{H}$. Substituting this back into the Eisenstein series gives a form of the Fourier series:

$$\begin{aligned}(E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) &= \delta_{\mathfrak{a},\mathfrak{b}} + \sum_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}\\ r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}}\overline{\chi}(r)\sum_{t\geq 1}\left(\frac{(-2\pi it)^k}{(k-1)!c^k}e^{2\pi it\frac{r}{c}}\right)e^{2\pi imz} \\ &= \sum_{t\geq 0}\left(\delta_{\mathfrak{a},\mathfrak{b}} + \sum_{\substack{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}\\ r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}}\overline{\chi}(r)\frac{(-2\pi it)^k}{(k-1)!c^k}e^{2\pi it\frac{r}{c}}\right)e^{2\pi itz} \\ &= \sum_{t\geq 0}\left(\delta_{\mathfrak{a},\mathfrak{b}} + \sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{(-2\pi it)^k}{(k-1)!c^k}\sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(r)e^{2\pi it\frac{r}{c}}\right)e^{2\pi itz}.\end{aligned}$$

We will simplify the innermost sum. Since a is the inverse for r modulo c, the innermost sum above becomes

$$\sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(r)e^{2\pi it\frac{r}{c}} = \sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\overline{\chi}(\overline{a})e^{2\pi it\frac{\overline{a}}{c}} = \sum_{r\in\mathcal{D}_{\mathfrak{a},\mathfrak{b}}}\chi(a)e^{\frac{2\pi i\overline{a}t}{c}} = S_{\chi,\mathfrak{a},\mathfrak{b}}(0,t,c).$$

So at last, we obtain our desired Fourier series:

$$(E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}})(z) = \sum_{t\geq 0}\left(\delta_{\mathfrak{a},\mathfrak{b}} + \sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{(-2\pi it)^k}{(k-1)!c^k}S_{\chi,\mathfrak{a},\mathfrak{b}}(0,t,c)\right)e^{2\pi itz}.$$

□

An interesting observation from Proposition 4.2.2 is that $E_{k,\chi,\mathfrak{a}}|\sigma_{\mathfrak{b}}$ is necessarily a cusp form unless $\mathfrak{a} = \mathfrak{b}$.

4.3 Inner Product Spaces of Holomorphic Forms

Let $\mathcal{H}_k(\Gamma,\chi)$ denote the space of all weight k holomorphic forms with character χ on $\Gamma\backslash\mathbb{H}$ and $\mathcal{S}_k(\Gamma,\chi)$ denote the associated subspace of cusp forms. Moreover, if χ is the trivial character, we will suppress the dependence upon χ. Note that if Γ_1 and Γ_2 are two congruence subgroups such that $\Gamma_1 \leq \Gamma_2$, then we have the inclusion

$$\mathcal{H}_k(\Gamma_2,\chi) \subseteq \mathcal{H}_k(\Gamma_1,\chi),$$

and this respects the subspace of cusp forms. So in general, the smaller the congruence subgroup the more holomorphic forms there are. Our goal is to turn $\mathcal{S}_k(\Gamma,\chi)$ into a complex Hilbert space to which we can apply a linear theory. To this end, for $f,g \in \mathcal{S}_k(\Gamma,\chi)$ define their **Petersson inner product** by

$$\langle f,g\rangle_\Gamma = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu.$$

If the congruence subgroup is clear from context we will suppress the dependence upon Γ. Since f and g have rapid decay at the cusps, the integrand is bounded and thus is locally absolutely uniformly convergent because we are integrating over a region of finite volume. The integrand is also Γ-invariant so that the integral is independent of the choice of fundamental domain. These two facts together imply that the Petersson inner product is well-defined. We will continue to use this notation even if f and g do not belong to $\mathcal{S}_k(\Gamma,\chi)$ provided the integral is locally absolutely uniformly convergent. A basic property of the Petersson inner product is that it is invariant with respect to the slash operator:

Proposition 4.3.1. *For any $f,g \in \mathcal{S}_k(\Gamma,\chi)$ and $\alpha \in \mathrm{PSL}_2(\mathbb{Z})$, we have*

$$\langle f|_k\alpha, g|_k\alpha\rangle_{\alpha^{-1}\Gamma\alpha} = \langle f,g\rangle_\Gamma.$$

Proof. This is just a computation:

$$\begin{aligned}
\langle f|_k\alpha, g|_k\alpha\rangle_{\alpha^{-1}\Gamma\alpha} &= \frac{1}{V_{\alpha^{-1}\Gamma\alpha}}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}} (f|_k\alpha)(z)\overline{(g|_k\alpha)(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}} (f|_k\alpha)(z)\overline{(g|_k\alpha)(z)}\mathrm{Im}(z)^k\,d\mu && \text{Lemma 3.3.1} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}} |j(\alpha,z)|^{-2k}f(\alpha z)\overline{g(\alpha z)}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |j(\alpha,z)|^{-2k}f(z)\overline{g(z)}\mathrm{Im}(\alpha z)^k\,d\mu && z\to\alpha^{-1}z \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \langle f,g\rangle_\Gamma.
\end{aligned}$$

□

To show that $\mathcal{S}_k(\Gamma,\chi)$ is a complex Hilbert space, we will need a dimensionality result (see [DS05] for a proof in the case χ is trivial):

Theorem 4.3.1. $\mathcal{H}_k(\Gamma,\chi)$ *is finite dimensional.*

We can now show that the Petersson inner product turns $\mathcal{S}_k(\Gamma,\chi)$ into a complex Hilbert space:

Proposition 4.3.2. $\mathcal{S}_k(\Gamma,\chi)$ *is a complex Hilbert space with respect to the Petersson inner product.*

Proof. Let $f,g \in \mathcal{S}_k(\Gamma,\chi)$. Linearity of the integral immediately implies that the Petersson inner product is linear on $\mathcal{S}_k(\Gamma,\chi)$. It is also positive definite since

$$\langle f,f\rangle = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{f(z)}\mathrm{Im}(z)^k\,d\mu = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z)|^2\mathrm{Im}(z)^k\,d\mu \geq 0,$$

with equality if and only if f is identically zero. To see that it is conjugate symmetric, observe

$$\begin{aligned}
\overline{\langle g,f\rangle} &= \overline{\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} g(z)\overline{f(z)}\mathrm{Im}(z)^k\,d\mu} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \overline{g(z)}f(z)\mathrm{Im}(z)^k\,\overline{d\mu} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \overline{g(z)}f(z)\mathrm{Im}(z)^k\,d\mu && d\mu = \frac{dx\,dy}{y^2} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \langle f,g\rangle.
\end{aligned}$$

So the Petersson inner product is a Hermitian inner product on $\mathcal{S}_k(\Gamma,\chi)$. Since $\mathcal{S}_k(\Gamma,\chi)$ is finite dimensional by Theorem 4.3.1, it follows immediately that $\mathcal{S}_k(\Gamma,\chi)$ is a complex Hilbert space. $\square$

Now suppose $f \in \mathcal{S}_k(\Gamma,\chi)$ has Fourier coefficients $a_\mathfrak{a}(n)$ at the $\mathfrak{a}$ cusp. Define linear functionals $\phi_{m,k,\chi,\mathfrak{a}} : \mathcal{S}_k(\Gamma,\chi) \to \mathbb{C}$ by

$$\phi_{m,k,\chi,\mathfrak{a}}(f) = a_\mathfrak{a}(m).$$

Since $\mathcal{S}_k(\Gamma,\chi)$ is a finite dimensional complex Hilbert space, the Riesz representation theorem implies that there exists unique $v_{m,k,\chi,\mathfrak{a}} \in \mathcal{S}_k(\Gamma,\chi)$ such that

$$\langle f, v_{m,k,\chi,\mathfrak{a}}\rangle = \phi_{m,k,\chi,\mathfrak{a}}(f) = a_\mathfrak{a}(m).$$

We would like to know what these cusp forms are. It turns out that $v_{m,k,\chi,\mathfrak{a}}(z)$ will be the Poincaré series $P_{m,k,\chi,\mathfrak{a}}(z)$ of positive index up to a constant. To deduce this, we will need the very useful identity

$$\int_0^1 e^{2\pi i(n-m)x}\,dx = \delta_{n-m,0}, \tag{4.1}$$

where $n,m \in \mathbb{Z}$. We will prove the following theorem:

Theorem 4.3.2. *Let* $f \in \mathcal{S}_k(\Gamma,\chi)$ *have Fourier coefficients* $a_\mathfrak{a}(n)$ *at the* $\mathfrak{a}$ *cusp of* $\Gamma\backslash\mathbb{H}$*. Then*

$$\langle f, P_{m,k,\chi,\mathfrak{a}}\rangle = \frac{\Gamma(k-1)}{V_\Gamma(4\pi m)^{k-1}}a_\mathfrak{a}(m),$$

for all $m \geq 1$*.*

Proof. We compute the inner product as follows:

$$
\begin{aligned}
\langle f, P_{m,k,\chi,\mathfrak{a}}\rangle &= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{P_{m,k,\chi,\mathfrak{a}}(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \chi(\gamma)\overline{j(\sigma_\mathfrak{a}^{-1}\gamma,z)^{-k}}f(z)e^{-2\pi i m\overline{\sigma_\mathfrak{a}^{-1}\gamma z}}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \chi(\gamma)j(\sigma_\mathfrak{a}^{-1}\gamma,z)^{k}f(z)e^{-2\pi i m\overline{\sigma_\mathfrak{a}^{-1}\gamma z}}\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \left(\frac{j(\sigma_\mathfrak{a}^{-1}\gamma,z)}{j(\gamma,z)}\right)^k f(\gamma z)e^{-2\pi i m\overline{\sigma_\mathfrak{a}^{-1}\gamma z}}\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z)^k\,d\mu && \text{modularity} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} j(\sigma_\mathfrak{a},\sigma_\mathfrak{a}^{-1}\gamma z)^{-k} f(\gamma z)e^{-2\pi i m\overline{\sigma_\mathfrak{a}^{-1}\gamma z}}\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z)^k\,d\mu && \text{cocycle condition} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} j(\sigma_\mathfrak{a},\sigma_\mathfrak{a}^{-1}\gamma\sigma_\mathfrak{a} z)^{-k} f(\gamma\sigma_\mathfrak{a} z)e^{-2\pi i m\overline{\sigma_\mathfrak{a}^{-1}\gamma\sigma_\mathfrak{a} z}}\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma\sigma_\mathfrak{a} z)^k\,d\mu && z\to\sigma_\mathfrak{a} z \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\infty\backslash\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}} j(\sigma_\mathfrak{a},\gamma z)^{-k} f(\sigma_\mathfrak{a}\gamma z)e^{-2\pi i m\overline{\gamma z}}\mathrm{Im}(\gamma z)^k\,d\mu && \gamma\to\sigma_\mathfrak{a}\gamma\sigma_\mathfrak{a}^{-1} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\infty\backslash\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}} (f|_k\sigma_\mathfrak{a})(\gamma z)e^{-2\pi i m\overline{\gamma z}}\mathrm{Im}(\gamma z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\Gamma_\infty\backslash\mathbb{H}} (f|_k\sigma_\mathfrak{a})(z)e^{-2\pi i m\overline{z}}\mathrm{Im}(z)^k\,d\mu && \text{unfolding.}
\end{aligned}
$$

Substituting in the Fourier series of f at the $\mathfrak{a}$ cusp, we can finish the computation:

$$
\begin{aligned}
\frac{1}{V_\Gamma}\int_{\Gamma_\infty\backslash\mathbb{H}} (f|_k\sigma_\mathfrak{a})(z)e^{-2\pi i m\overline{z}}\mathrm{Im}(z)^k\,d\mu &= \frac{1}{V_\Gamma}\int_0^\infty\int_0^1\sum_{n\ge1}a_\mathfrak{a}(n)e^{2\pi i(n-m)x}e^{-2\pi(n+m)y}y^k\,\frac{dx\,dy}{y^2} \\
&= \frac{1}{V_\Gamma}\int_0^\infty\sum_{n\ge1}\int_0^1 a_\mathfrak{a}(n)e^{2\pi i(n-m)x}e^{-2\pi(n+m)y}y^k\,\frac{dx\,dy}{y^2} && \text{FT} \\
&= \frac{1}{V_\Gamma}\int_0^\infty a_\mathfrak{a}(m)e^{-4\pi m y}y^k\,\frac{dy}{y^2},
\end{aligned}
$$

where the last line follows because Equation (4.1) implies that the inner integral cuts off all the terms except the diagonal $n=m$. Then

$$
\begin{aligned}
\frac{1}{V_\Gamma}\int_0^\infty a_\mathfrak{a}(m)e^{-4\pi m y}y^k\,\frac{dy}{y^2} &= \frac{a_\mathfrak{a}(m)}{V_\Gamma}\int_0^\infty e^{-4\pi m y}y^{k-1}\,\frac{dy}{y} \\
&= \frac{a_\mathfrak{a}(m)}{V_\Gamma}\int_0^\infty e^{-4\pi m y}y^{k-1}\,\frac{dy}{y} && y\to\frac{y}{4\pi m} \\
&= \frac{a_\mathfrak{a}(m)}{V_\Gamma(4\pi m)^{k-1}}\int_0^\infty e^{-y}y^{k-1}\,\frac{dy}{y} \\
&= \frac{\Gamma(k-1)}{V_\Gamma(4\pi m)^{k-1}}a_\mathfrak{a}(m).
\end{aligned}
$$

This completes the proof. □

It follows immediately from Theorem 4.3.2 that

$$v_{m,k,\chi,\mathfrak{a}}(z) = \frac{V_\Gamma(4\pi m)^{k-1}}{\Gamma(k-1)} P_{m,k,\chi,\mathfrak{a}}(z).$$

Thus the Poincaré series $P_{m,k,\chi,\mathfrak{a}}(z)$ of positive index extract the Fourier coefficients of f at different cusps up to a constant. With Theorem 4.3.2 in hand we can prove the following result:

Theorem 4.3.3. *The Poincaré series of positive index span $\mathcal{S}_k(\Gamma,\chi)$.*

Proof. Let $f \in \mathcal{S}_k(\Gamma,\chi)$ have Fourier coefficients $a_{\mathfrak{a}}(n)$ at the $\mathfrak{a}$ cusp. Since $\Gamma(k-1) \neq 0$, Theorem 4.3.2 implies $\langle f, P_{m,k,\chi,\mathfrak{a}}\rangle = 0$ if and only if $a_{\mathfrak{a}}(m) = 0$. So f is orthogonal to all the Poincaré series $P_{m,k,\chi,\mathfrak{a}}$ of positive index if and only if every Fourier coefficient $a_{\mathfrak{a}}(m)$ is zero. But this happens if and only if f is identically zero. □

4.4 Double Coset Operators

We are ready to introduce a class of general operators, depending upon double cosets, on a congruence subgroup Γ of level N. We will use these operators to define the diamond and Hecke operators. For $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, consider the double coset

$$\Gamma_1\alpha\Gamma_2 = \{\gamma_1\alpha\gamma_2 : \gamma_1 \in \Gamma_1 \text{ and } \gamma_2 \in \Gamma_2\}.$$

Then Γ_1 acts on the set $\Gamma_1\alpha\Gamma_2$ by left multiplication so that it decomposes into a disjoin union of orbit spaces. Thus

$$\Gamma_1\alpha\Gamma_2 = \bigcup_{\beta \in \Gamma_1\backslash\Gamma_1\alpha\Gamma_2} \Gamma_1\beta,$$

where the sum is over the orbit representatives β. However, in order for these operators to be well-defined it is necessary that the orbit decomposition above is a finite union. This is indeed the case and we will require a lemma which gives a way to describe the orbit representatives for $\Gamma_1\alpha\Gamma_2$ in terms of coset representatives:

Lemma 4.4.1. *Let Γ_1 and Γ_2 be congruence subgroups and let let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Set $\Gamma_3 = \alpha^{-1}\Gamma_1\alpha \cap \Gamma_2$. Then left multiplication map*

$$\Gamma_2 \to \Gamma_1\alpha\Gamma_2 \qquad \gamma_2 \mapsto \alpha\gamma_2,$$

induces a bijection from the coset space $\Gamma_3\backslash\Gamma_2$ to the orbit space $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2$.

Proof. We will show that the induced map is both surjective and injective. For surjectivity, the orbit representatives β of $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2$ are of the form $\beta = \gamma_1\alpha\gamma_2$ for some $\gamma_1 \in \Gamma_1$ and $\gamma_2 \in \Gamma_2$. Since Γ_1 is acting on $\Gamma_1\alpha\Gamma_2$ by left multiplication, β can be written as $\beta = \alpha\gamma_2'$ for some $\gamma_2' \in \Gamma_2$. This shows that the induced map is a surjection. To prove injectivity, let $\gamma_2, \gamma_2' \in \Gamma_2$ be such that the orbit space representatives $\alpha\gamma_2$ and $\alpha\gamma_2'$ are equivalent. That is,

$$\Gamma_1\alpha\gamma_2 = \Gamma_1\alpha\gamma_2'.$$

This implies $\alpha\gamma_2(\gamma_2')^{-1} \in \Gamma_1\alpha$ and so $\gamma_2(\gamma_2')^{-1} \in \alpha^{-1}\Gamma_1\alpha$. But we also have $\gamma_2(\gamma_2')^{-1} \in \Gamma_2$ and these two facts together imply $\gamma_2(\gamma_2')^{-1} \in \Gamma_3$. Hence

$$\Gamma_3\gamma_2 = \Gamma_3\gamma_2',$$

which shows that the induced map is also an injection. □

With this lemma in hand, we can prove that the orbit decomposition of $\Gamma_1\alpha\Gamma_2$ is finite:

Proposition 4.4.1. *Let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Then the orbit decomposition*

$$\Gamma_1\alpha\Gamma_2 = \bigcup_j \Gamma_1\beta_j,$$

with respect to the action of Γ_1 by left multiplication, is a finite union.

Proof. Let $\Gamma_3 = \alpha\Gamma_1\alpha^{-1} \cap \Gamma_2$. Then Γ_3 acts on Γ_2 by left multiplication. By Lemma 4.4.1, the number of orbits of $\Gamma_1\backslash\Gamma_1\alpha\Gamma_2$ is the same as the number of cosets of $\Gamma_3\backslash\Gamma_2$ which is $[\Gamma_2 : \Gamma_3]$. By Lemma 3.1.1, $\alpha^{-1}\Gamma_1\alpha \cap \mathrm{PSL}_2(\mathbb{Z})$ is a congruence subgroup and hence $[\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\Gamma_1\alpha \cap \mathrm{PSL}_2(\mathbb{Z})]$ is finite. As $\Gamma_2 = \mathrm{PSL}_2(\mathbb{Z})\cap\Gamma_2$ and $\Gamma_3 = \alpha^{-1}\Gamma_1\alpha\cap\mathrm{PSL}_2(\mathbb{Z})\cap\Gamma_2$, it follows that $[\Gamma_2 : \Gamma_3] \le [\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\Gamma_1\alpha\cap\mathrm{PSL}_2(\mathbb{Z})]$ completing the proof. □

In light of Proposition 4.4.1, we will denote the orbit representatives by β_j to make it clear that there are finitely many. We can now introduce our double coset operators. Let Γ_1 and Γ_2 be two congruence subgroups and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. We define the **double coset operator** $[\Gamma_1\alpha\Gamma_2]_k$ on $\mathcal{S}_k(\Gamma_1)$ to be the linear operator given by

$$(f[\Gamma_1\alpha\Gamma_2]_k)(z) = \sum_j (f|_k\beta_j)(z) = \sum_j \det(\beta_j)^{k-1}j(\beta_j, z)^{-k}f(\beta_j z),$$

By Proposition 4.4.1 this sum is finite. It remains to check that $f[\Gamma_1\alpha\Gamma_2]_k$ is well-defined. Indeed, if β_j and β_j' belong to the same orbit, then $\beta_j'\beta_j^{-1} \in \Gamma_1$. But then as $f \in \mathcal{S}_k(\Gamma_1)$, is it invariant under the $|_k\beta_j'\beta_j^{-1}$ operator so that

$$(f|_k\beta_j)(z) = ((f|_k\beta_j'\beta_j^{-1})|_k\beta_j)(z) = (f|_k\beta_j')(z),$$

and therefore the $[\Gamma_1\alpha\Gamma_2]_k$ operator is well-defined. Actually, the map $[\Gamma_1\alpha\Gamma_2]_k$ preserves holomorphic forms:

Proposition 4.4.2. *For any two congruence subgroups Γ_1 and Γ_2, $[\Gamma_1\alpha\Gamma_2]_k$ maps $\mathcal{S}_k(\Gamma_1)$ into $\mathcal{S}_k(\Gamma_2)$.*

Proof. Holomorphy is immediate since the sum in $f[\Gamma_1\alpha\Gamma_2]_k$ is finite by Proposition 4.4.1. For modularity, let $\gamma \in \Gamma_2$. Then

$$\begin{aligned}
(f[\Gamma_1\alpha\Gamma_2]_k)(\gamma z) &= \sum_j \det(\beta_j)^{k-1}j(\beta_j, \gamma z)^{-k}f(\beta_j\gamma z) \\
&= \sum_j \det(\beta_j\gamma)^{k-1}j(\beta_j, \gamma z)^{-k}f(\beta_j\gamma z) && \det(\gamma) = 1 \\
&= \sum_j \det(\beta_j\gamma)^{k-1}\left(\frac{j(\gamma, z)}{j(\beta_j\gamma, z)}\right)^k f(\beta_j\gamma z) && \text{cocycle condition} \\
&= j(\gamma, z)^k\sum_j \det(\beta_j\gamma)^{k-1}j(\beta_j\gamma, z)^{-k}f(\beta_j\gamma z) \\
&= j(\gamma, z)^k\sum_j \det(\beta_j)^{k-1}j(\beta_j, z)^{-k}f(\beta_j z) && \beta_j \to \beta_j\gamma^{-1} \\
&= j(\gamma, z)^k\sum_j (f|_k\beta_j)(z) \\
&= j(\gamma, z)^k(f[\Gamma_1\alpha\Gamma_2]_k)(z).
\end{aligned}$$

This proves the modularity of $f[\Gamma_1\alpha\Gamma_2]_k$. For the growth condition, let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the cusp $\mathfrak{a}$ of $\Gamma_2\backslash\mathbb{H}$. For any orbit representative β_j, $\beta_j\sigma_{\mathfrak{a}}$ takes ∞ to an element of $\mathbb{Q}\cup\{\infty\}$ since $\beta_j\in\mathrm{GL}_2^+(\mathbb{Q})$. In other words, $\beta_j\sigma_{\mathfrak{a}}\infty=\mathfrak{b}$ for some cusp $\mathfrak{b}$ of $\Gamma_1\backslash\mathbb{H}$. Then by the cocycle condition, we have

$$j(\sigma_{\mathfrak{a}},z)^{-k}(f[\Gamma_1\alpha\Gamma_2]_k)(\sigma_{\mathfrak{a}}z)=\sum_j \det(\beta_j)^{k-1}j(\beta_j\sigma_{\mathfrak{a}},z)^{-k}f(\beta_j\sigma_{\mathfrak{a}}z),$$

and the growth condition follows from that of f. In particular, $f[\Gamma_1\alpha\Gamma_2]_k$ is a cusp form since f is. □

The double coset operators are the most basic types of operators on holomorphic forms. They are the building blocks needed to define the more important diamond and Hecke operators.

4.5 Diamond & Hecke Operators

The diamond and Hecke operators are special linear operators that are used to construct a linear theory of holomorphic forms. They will also help us understand the Fourier coefficients. Throughout this discussion, we will obtain corresponding results for holomorphic forms with nontrivial characters. We will discuss the diamond operator first. To define them, we need to consider both the congruence subgroups $\Gamma_1(N)$ and $\Gamma_0(N)$. Recall that $\Gamma_1(N)\le\Gamma_0(N)$ and consider the map

$$\Gamma_0(N)\to(\mathbb{Z}/N\mathbb{Z})^* \qquad \begin{pmatrix} a & b\\ c & d\end{pmatrix}\to d \pmod{N},$$

(d is invertible modulo N since $c\equiv 0 \pmod N$ and $ad-bc=1$). This is a surjective homomorphism and its kernel is exactly $\Gamma_1(N)$ so that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and $\Gamma_0(N)/\Gamma_1(N)\cong(\mathbb{Z}/N\mathbb{Z})^*$. Letting $\alpha=\left(\begin{smallmatrix} * & *\\ * & d\end{smallmatrix}\right)\in\Gamma_0(N)$ and $f\in\mathcal{S}_k(\Gamma_1)$, consider $\left(f\left[\Gamma_1(N)\alpha\Gamma_1(N)\right]_k\right)(z)$. This is only dependent upon the lower-right entry d of α taken modulo N. To see this, since $\Gamma_1(N)$ is normal in $\Gamma_0(N)$, $\Gamma_1(N)\alpha=\alpha\Gamma_1(N)$ so that $\Gamma_1(N)\alpha\Gamma_1(N)=\alpha\Gamma_1(N)$ and hence there is only one representative for the orbit decomposition. Therefore

$$\left(f\left[\Gamma_1(N)\alpha\Gamma_1(N)\right]_k\right)(z)=(f|_k\alpha)(z).$$

This induces an action of $\Gamma_0(N)$ on $\mathcal{S}_k(\Gamma_1)$ and since $\Gamma_1(N)$ acts trivially, this is really an action of $\Gamma_0(N)/\Gamma_1(N)\cong(\mathbb{Z}/N\mathbb{Z})^*$. In other words, we have an induced action that depends only upon the lower-right entry d of α taken modulo N. So for any d modulo N, we define the **diamond operator** $\langle d\rangle:\mathcal{S}_k(\Gamma_1(N))\to\mathcal{S}_k(\Gamma_1(N))$ to be the linear operator given by

$$(\langle d\rangle f)(z)=(f|_k\alpha)(z),$$

for any $\alpha=\left(\begin{smallmatrix} * & *\\ * & d\end{smallmatrix}\right)\in\Gamma_0(N)$. Our discussion above has already shown that the diamond operators $\langle d\rangle$ are well-defined. Moreover, the diamond operators are also invertible with $\langle\overline{d}\rangle$ serving as an inverse and α^{-1} as a representative for the definition. Also, since the operator $|_k\alpha$ is multiplicative and

$$\begin{pmatrix} * & *\\ 0 & d\end{pmatrix}\begin{pmatrix} * & *\\ 0 & e\end{pmatrix}\equiv\begin{pmatrix} * & *\\ 0 & de\end{pmatrix} \pmod{N},$$

the diamond operators are multiplicative. One reason the diamond operators are useful is that they decompose $\mathcal{S}_k(\Gamma_1(N))$ into eigenspaces. For any Dirichlet character χ modulo N, we let

$$\mathcal{S}_k(N,\chi)=\{f\in\mathcal{S}_k(\Gamma_1(N)):\langle d\rangle f=\chi(d)f \text{ for all } d\in(\mathbb{Z}/N\mathbb{Z})^*\},$$

be the χ-eigenspace. Also let $\mathcal{S}_k(N,\chi)$ be the corresponding subspace of cusp forms. Then $\mathcal{S}_k(\Gamma_1(N))$ admits a decomposition into these eigenspaces:

Proposition 4.5.1. *We have a direct sum decomposition*

$$\mathcal{S}_k(\Gamma_1(N)) = \bigoplus_{\chi \ (mod\ N)} \mathcal{S}_k(N,\chi).$$

Proof. The diamond operators give a representation of $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^*$ over $\mathcal{S}_k(\Gamma_1(N))$. Explicitly,

$$\Phi : (\mathbb{Z}/N\mathbb{Z})^* \times \mathcal{S}_k(\Gamma_1(N)) \to \mathcal{S}_k(\Gamma_1(N)) \qquad (d,f) \to \langle d \rangle f.$$

But any representation of a finite abelian group over $\mathbb{C}$ is completely reducible with respect to the characters of the group and every irreducible subrepresentation is 1-dimensional (see Theorem C.3.1). Since the characters of $(\mathbb{Z}/N\mathbb{Z})^*$ are given by Dirichlet characters, the decomposition as a direct sum follows. $\square$

Proposition 4.5.1 shows that the diamond operators sieve holomorphic forms on $\Gamma_1(N)\backslash\mathbb{H}$ with trivial character in terms of holomorphic forms on $\Gamma_0(N)\backslash\mathbb{H}$ with nontrivial characters. In particular, $\mathcal{S}_k(N,\chi) = \mathcal{S}_k(\Gamma_0(N),\chi)$. So by Proposition 4.5.1, we have

$$\mathcal{S}_k(\Gamma_1(N)) = \bigoplus_{\chi \ (\mathrm{mod}\ N)} \mathcal{S}_k(\Gamma_0(N),\chi).$$

This fact clarifies why it is necessary to consider holomorphic forms with nontrivial characters. Now it is time to define the Hecke operators. For a prime p, we define the p-th **Hecke operator** $T_p : \mathcal{S}_k(\Gamma_1(N)) \to \mathcal{S}_k(\Gamma_1(N))$ to be the linear operator given by

$$(T_p f)(z) = \left(\left[\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\Gamma_1(N)\right]_k f\right)(z).$$

We will start discussing properties of the diamond and Hecke operators, but first we prove an important lemma classifying the double coset $\Gamma_1(N)\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\Gamma_1(N)$:

Lemma 4.5.1. *Let p be a prime. Then*

$$\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\Gamma_1(N) = \left\{\gamma \in \mathrm{Mat}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix}1 & *\\ 0 & p\end{pmatrix} \ (mod\ N) \ and \ \det(\gamma) = p\right\}.$$

Proof. For the forward containment, it is clear that any element of $\Gamma_1(N)\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\Gamma_1(N)$ has determinant p and that

$$\begin{pmatrix}1 & *\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\begin{pmatrix}1 & *\\ 0 & 1\end{pmatrix} \equiv \begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix} \ (\mathrm{mod}\ N),$$

so the forward containment holds. For the reverse containment, let L be the set of 2×1 colum vectors with entries in $\mathbb{Z}$, in particular $L \cong \mathbb{Z}^2$, and set

$$L_0 = \left\{\begin{pmatrix}x\\ y\end{pmatrix} \in L : y \equiv 0 \ (\mathrm{mod}\ N)\right\}.$$

Then $\mathrm{Mat}_2(\mathbb{Z})$ acts on L_0 on the left by matrix multiplication. Now suppose $\gamma \in \mathrm{Mat}_2(\mathbb{Z})$ is such that $\gamma \equiv \left(\begin{smallmatrix}1 & *\\ 0 & p\end{smallmatrix}\right)$ (mod N) and $\det(\gamma) = p$. Clearly L_0 has index N in L. As the action of γ on L_0 multiplies the lower entry by p, γL_0 has index p in L_0. These two facts together imply

$$[L : \gamma L_0] = [L : L_0][L_0; \gamma L_0] = Np.$$

As γL_0 is a finitely generated abelian group (L is a rank 2 free module over $\mathbb{Z}$ and γL_0 is a subgroup), the structure theorem for finitely generated abelian groups implies that there exists a basis $\{u, v\}$ of L with $\det(u, v) = 1$, so $L = u\mathbb{Z} \oplus v\mathbb{Z}$, and positive integers m and n with $m \mid n$, $mn = Np$, and such that $\gamma L_0 = mu\mathbb{Z} \oplus nv\mathbb{Z}$. Moreover, as $m \mid n$ any element of γL_0 is zero modulo m. Now $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \in L_0$ and by our choice of γ, we have $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \in \gamma L_0$ as well. Thus $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right) \pmod{m}$. This forces $m = 1$ and $n = Np$ so that $\gamma L_0 = u\mathbb{Z} \oplus Npv\mathbb{Z}$. As there are unique subgroups of index N and p between $\mathbb{Z}$ and $Np\mathbb{Z}$, namely $N\mathbb{Z}$ and $p\mathbb{Z}$ respectively, we have

$$L_0 = u\mathbb{Z} \oplus Nv\mathbb{Z}, \quad \gamma L = u\mathbb{Z} \oplus pv\mathbb{Z}, \quad \text{and} \quad \gamma L_0 = u\mathbb{Z} \oplus Npv\mathbb{Z}.$$

Now let $\gamma_1 = (u, v)$. Since $u \in L_0$, we have that $\gamma_1 \in \Gamma_0(N)$. Now set $\gamma_2 = \left(\gamma_1 \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)\right)^{-1} \gamma$. Then $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ with $\det(\gamma_2) = 1$. It follows that

$$\gamma = \gamma_1 \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \gamma_2.$$

Let e_1 and e_2 be the standard basis vectors of L. Then $\gamma e_1 \in \gamma L_0$ implies that $\gamma e_1 = au + pcv$ with $a, c \in \mathbb{Z}$ and $N \mid c$. Similarly, $\gamma e_2 \in \gamma L$ implies that $\gamma e_2 = bu + pdv$ with $b, d \in \mathbb{Z}$. Letting $u = \left(\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right)$ and $v = \left(\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right)$, these facts together imply that $\gamma_2 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and the above identity can be written in the form

$$\gamma = \begin{pmatrix} au_1 + pcv_1 & bu_1 + pdv_1 \\ au_2 + pcv_2 & bu_2 + pdv_2 \end{pmatrix} = \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

As $\det(\gamma_2) = 1$ and $N \mid c$, we conclude that $\gamma_2 \in \Gamma_0(N)$. Hence $\gamma \in \Gamma_0(N) \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) \Gamma_0(N)$. Moreover, as $\gamma_1, \gamma_2 \in \Gamma_0(N)$ it follows that they belong to $\Gamma_1(N)$ if u_1 or a is equivalent to 1 modulo N respectively. But as $\gamma \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & p \end{smallmatrix}\right) \pmod{N}$ and $N \mid c$, we have $au_1 \equiv 1 \pmod{N}$. Thus if γ_1 or γ_2 belongs to $\Gamma_1(N)$ then they both do. So to complete the reverse containment it suffices to show

$$\Gamma_0(N) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_0(N) = \Gamma_1(N) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_0(N).$$

The reverse containment in this double coset equality is clear since $\Gamma_1(N) \subseteq \Gamma_0(N)$. For the forward containment, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$. Then we need to show that there exists a matrix $\delta \in \Gamma_1(N)$ such that

$$\gamma \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_0(N) = \delta \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_0(N).$$

Equivalently, there exists a matrix $\delta' \in \Gamma_1(N)$ (that is the inverse of δ) such that

$$\begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \delta' \gamma \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \in \Gamma_0(N).$$

If $p \mid N$, taking $\delta' = \left(\begin{smallmatrix} cd+1 & -1 \\ -cd & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$ gives

$$\begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \delta' \gamma \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \begin{pmatrix} cd+1 & -1 \\ -cd & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} * & * \\ p^{-1}c(1-ad) & * \end{pmatrix},$$

which belongs to $\Gamma_0(N)$ because $p \mid N$, $N \mid c$, and $N \mid 1 - ad$ (as $ad - bc = 1$). If $p \nmid N$, then the Chinese remainder theorem implies that there exists a $d' \in \mathbb{Z}$ with $d' \equiv 1 \pmod{N}$ and $d' \equiv -a \pmod{p}$. Necessarily $(c, d') = 1$ because $N \mid c$. Thus there exists δ' of the form $\delta' = \left(\begin{smallmatrix} a' & b' \\ c & d' \end{smallmatrix}\right) \in \Gamma_1(N)$, and we have

$$\begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \delta' \gamma \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \begin{pmatrix} a' & b' \\ c & d' \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} * & * \\ p^{-1}c(a+d') & * \end{pmatrix},$$

which again belongs to $\Gamma_0(N)$ because $p \mid d' + a$ and $N \mid c$. This prove the forward containment, and hence the original reverse containment, thus completing the proof. $\square$

With Lemma 4.5.1, it is not too hard to see that the diamond and Hecke operators commute:

Proposition 4.5.2. *For every $d \in (\mathbb{Z}/N\mathbb{Z})^*$ and prime p, the diamond operators $\langle d \rangle$ and Hecke operators T_p on $\mathcal{S}_k(\Gamma_1(N))$ commute:*

$$\langle d \rangle T_p = T_p \langle d \rangle$$

Proof. Let $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$\gamma\alpha\gamma^{-1} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 & (p-1)ab \\ 0 & p \end{pmatrix} \pmod{N},$$

because $c \equiv 0 \pmod{N}$, $ad - bc = 1$, and $ad \equiv 1 \pmod{N}$. By Lemma 4.5.1, $\gamma\alpha\gamma^{-1} \in \Gamma_1(N)\alpha\Gamma_1(N)$ and so we can use this representative in place of α. On the one hand,

$$\Gamma_1(N)\alpha\Gamma_1(N) = \bigcup_j \Gamma_1(N)\beta_j.$$

On the other hand, using $\gamma\alpha\gamma^{-1}$ in place of α and the normality of $\Gamma_1(N)$ in $\Gamma_0(N)$, we have

$$\Gamma_1(N)\alpha\Gamma_1(N) = \Gamma_1(N)\gamma\alpha\gamma^{-1}\Gamma_1(N) = \gamma\Gamma_1(N)\alpha\Gamma_1(N)\gamma^{-1} = \gamma\bigcup_j \Gamma_1(N)\beta_j\gamma^{-1} = \bigcup_j \Gamma_1(N)\gamma\beta_j\gamma^{-1}.$$

Upon comparing these two decompositions of $\Gamma_1(N)\alpha\Gamma_1(N)$ gives

$$\bigcup_j \Gamma_1(N)\beta_j = \bigcup_j \Gamma_1(N)\gamma\beta_j\gamma^{-1}.$$

Now let $f \in \mathcal{S}_k(\Gamma_1(N))$. Then this equivalence of unions implies

$$\langle d \rangle T_p f = \sum_j f|_k \beta_j\gamma = \sum_j f|_k \gamma\beta_j = T_p\langle d \rangle f.$$

□

Using Lemma 4.5.1 we can obtain an explicit description of the Hecke operator T_p:

Proposition 4.5.3. *Let $f \in \mathcal{S}_k(\Gamma_1(N))$. Then the Hecke operator T_p acts on f as follows:*

$$(T_p f)(z) = \begin{cases} \displaystyle\sum_{j \pmod p} \left(f\Big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}\right)(z) + \left(f\Big|_k \begin{pmatrix} m & n \\ N & p \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\right)(z) & \text{if } p \nmid N, \\ \displaystyle\sum_{j \pmod p} \left(f\Big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}\right)(z) & \text{if } p \mid N, \end{cases}$$

where m and n are chosen such that $\det\left(\left(\begin{smallmatrix} m & n \\ N & p \end{smallmatrix}\right)\right) = 1$.

Proof. Set $\Gamma_3 = \alpha^1\Gamma_1(N)\alpha \cap \Gamma_1(N)$ where $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$. Define

$$\beta_j = \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} \quad \text{and} \quad \beta_\infty = \begin{pmatrix} m & n \\ N & p \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} pm & n \\ Np & p \end{pmatrix},$$

for j taken modulo p and where m and n are chosen such that $\det\left(\left(\begin{smallmatrix} m & n \\ N & p \end{smallmatrix}\right)\right) = 1$. It suffices to show $\{\beta_1, \ldots, \beta_{p-1}\}$ and $\{\beta_1, \ldots, \beta_{p-1}, \beta_\infty\}$ are complete sets of orbit representatives for $\Gamma_1(N)\backslash\Gamma_1(N)\alpha\Gamma_1(N)$

depending on if $p \mid N$ or not. To accomplish this, we will find a complete set of coset representatives for $\Gamma_3\backslash\Gamma_1(N)$ and then use Lemma 4.4.1. First we require an explicit description of Γ_3. Let

$$\Gamma^0(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \pmod{p} \right\},$$

and define

$$\Gamma_1^0(N,p) = \Gamma_1(N) \cap \Gamma^0(p).$$

We claim $\Gamma_3 = \Gamma_1^0(N,p)$. For the forward containment, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$ and observe that

$$\alpha^{-1}\gamma\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} a & pd \\ p^{-1}c & d \end{pmatrix}.$$

If $\alpha^{-1}\gamma\alpha \in \Gamma_3$, then $\alpha^{-1}\gamma\alpha \in \Gamma_1(N)$ and thus $p \mid c$ so that $\alpha^{-1}\gamma\alpha \in \mathrm{PSL}_2(\mathbb{Z})$. Moreover, the previous computation implies $\alpha^{-1}\gamma\alpha \in \Gamma_1^0(N,p)$ and so $\Gamma_3 \subseteq \Gamma_1^0(N,p)$. For the reverse containment, suppose $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0^1(N,p)$. Then $b = pk$ for some $k \in \mathbb{Z}$. Now observe

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix}\begin{pmatrix} a & k \\ pc & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \alpha^{-1}\gamma\alpha,$$

where $\gamma = \left(\begin{smallmatrix} a & k \\ pc & d \end{smallmatrix}\right)$. As $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$ we conclude $\gamma \in \Gamma_1(N)$ as well. Now let

$$\alpha_j = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \alpha_\infty = \begin{pmatrix} pm & n \\ N & 1 \end{pmatrix},$$

for j taken modulo p and where m and n are as before. Clearly $\alpha_j \in \Gamma_1(N)$ for all j. As $pm - Nn = 1$, we have $pm \equiv 1 \pmod{N}$ so that $\alpha_\infty \in \Gamma_1(N)$ as well. We claim that $\{\alpha_1, \ldots, \alpha_{p-1}\}$ and $\{\alpha_1, \ldots, \alpha_{p-1}, \alpha_\infty\}$ are sets of coset representatives for $\Gamma_3\backslash\Gamma_1(N)$ depending on if $p \mid N$ or not. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$ and consider

$$\gamma\alpha_j^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & -j \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b - aj \\ c & d - cj \end{pmatrix}.$$

As $\gamma\alpha_j^{-1} \in \Gamma_1(N)$ because both γ and γ_i are, $\gamma\alpha_j^{-1} \in \Gamma_3 = \Gamma_1^0(N,p)$ for some i if and only if

$$b \equiv aj \pmod{p}.$$

First suppose $p \nmid a$. Then a is invertible modulo p so we may take $j = \overline{a}b \pmod{p}$. Now suppose $p \mid a$. If there is some i satisfying $b \equiv ai \pmod{p}$, then we also have $p \mid b$. But as $ad - bc = 1$, this is impossible and so no such i exists. As $a \equiv 1 \pmod{N}$, $p \mid a$ if and only if $p \nmid N$. In this case consider instead

$$\gamma\alpha_\infty^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & -n \\ -N & pm \end{pmatrix} = \begin{pmatrix} a - Nb & pmb - na \\ c - Nd & pmd - nc \end{pmatrix}.$$

Since $p \mid a$, we have $pmb - na \equiv 0 \mod p$ so that $\gamma\alpha_\infty^{-1} \in \Gamma_3 = \Gamma_1^0(N,p)$. Altogether, we have shown that $\{\alpha_1, \ldots, \alpha_{p-1}\}$ and $\{\alpha_1, \ldots, \alpha_{p-1}, \alpha_\infty\}$ are sets of coset representatives for $\Gamma_3\backslash\Gamma_1(N)$ depending on if $p \mid N$ or not. To show they are complete sets, we need to show that no two representatives belong to the same coset. To this end, suppose j and j' are distinct, taken modulo p, and consider

$$\alpha_j\alpha_{j'}^{-1} = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & j' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & j - j' \\ 0 & 1 \end{pmatrix}.$$

Then $\alpha_j\alpha_{j'}^{-1} \in \Gamma_3 = \Gamma_1^0(N,p)$ if and only if $j - j' \equiv 0 \pmod p$. This is impossible since j and j' are distinct. Hence α_j and $\alpha_{j'}$ represent distinct cosets. Now consider

$$\alpha_j\alpha_\infty^{-1} = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -n \\ -N & pm \end{pmatrix} = \begin{pmatrix} 1-Nj & pmj-n \\ -N & pm \end{pmatrix}.$$

Now $\alpha_j\alpha_\infty^{-1} \in \Gamma_3 = \Gamma_1^0(N,p)$ if and only if $pmj - n \equiv 0 \pmod p$. This is impossible since $pm - Nn = 1$ implies $p \nmid n$. Therefore α_j and α_∞ represent distinct cosets. It follows that $\{\alpha_1, \dots, \alpha_{p-1}\}$ and $\{\alpha_1, \dots, \alpha_{p-1}, \alpha_\infty\}$ are complete sets of coset representatives completing the proof. As

$$\alpha\alpha_j = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}\begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} = \beta_j \quad \text{and} \quad \alpha\alpha_\infty = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}\begin{pmatrix} pm & n \\ N & 1 \end{pmatrix} = \begin{pmatrix} pm & n \\ Np & p \end{pmatrix} = \beta_\infty,$$

Lemma 4.4.1 finishes the proof. □

This explicit definition of T_p can be used to compute how the Hecke operators act on the Fourier coefficients of a holomorphic form:

Proposition 4.5.4. *Let $f \in \mathcal{S}_k(\Gamma_1(N))$ have Fourier coefficients $a_n(f)$. Then for all primes p,*

$$(T_pf)(z) = \sum_{n \geq 1} \left(a_{np}(f) + \chi_{N,0}(p)p^{k-1}a_{\frac{n}{p}}(\langle p \rangle f)\right) e^{2\pi inz},$$

is the Fourier series of T_pf where it is understood that $a_{\frac{n}{p}}(f) = 0$ if $p \nmid n$. Moreover, if $f \in \mathcal{S}_k(N,\chi)$, then $T_pf \in \mathcal{S}_k(N,\chi)$ and

$$(T_pf)(z) = \sum_{n \geq 1} \left(a_{np}(f) + \chi(p)p^{k-1}a_{\frac{n}{p}}(f)\right) e^{2\pi inz},$$

where it is understood that $a_{\frac{n}{p}}(f) = 0$ if $p \nmid n$.

Proof. In view of Proposition 4.5.1 and the linearity of the Hecke operators, it suffices to assume $f \in \mathcal{S}_k(N,\chi)$. By Proposition 4.5.2, $T_pf \in \mathcal{S}_k(N,\chi)$. It remains to verify the second formula. Observe that

$$\left(f\Big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}\right)(z) = \frac{1}{p}\sum_{n \geq 1} a_n(f)e^{\frac{2\pi in(z+j)}{p}}.$$

Summing over all j modulo p gives

$$\sum_{j \pmod p} \left(f\Big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}\right)(z) = \sum_{n \geq 1} a_n(f)e^{\frac{2\pi inz}{p}} \frac{1}{p}\sum_{j \pmod p} e^{\frac{2\pi nij}{p}}.$$

If $p \nmid N$ then the inner sum vanishes because it is the sum over all p-th roots of unity. If $p \mid N$ then the sum is also p. Therefore

$$\sum_{j \pmod p} \left(f\Big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}\right)(z) = \sum_{n \geq 1} a_{np}(f)e^{2\pi inz}.$$

If $p \mid N$, then Proposition 4.5.3 implies

$$(T_pf)(z) = \sum_{n \geq 1} a_{np}(f)e^{2\pi inz}, \tag{4.2}$$

which is the claimed Fourier series of T_pf. If $p \nmid N$, then we have the additional term

$$\begin{aligned}\left(f\bigg|_k \begin{pmatrix} m & n \\ N & p \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\right)(z) &= \left(\langle p \rangle f\bigg|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\right)(z) \\ &= p^{k-1}(\langle p \rangle f)(pz) \\ &= \sum_{n \geq 1} p^{k-1} a_n(\langle p \rangle f) e^{2\pi i n p z} \\ &= \sum_{n \geq 1} \chi(p) p^{k-1} a_n(f) e^{2\pi i n p z},\end{aligned}$$

where the first equality holds because $\left(\begin{smallmatrix} m & n \\ N & p \end{smallmatrix}\right) \in \Gamma_0(N)$ and the last equality holds because $\langle p \rangle f = \chi(p) f$. In this case, Proposition 4.5.3 gives

$$(T_pf)(z) = \sum_{n \geq 1} a_{np}(f) + \chi(p)p^{k-1}a_{\frac{n}{p}}(f)e^{2\pi i n z}.$$

Since $\chi(p) = 0$ if $p \mid N$, these two cases can be expressed together as

$$(T_pf)(z) = \sum_{n \geq 1} \left(a_{np}(f) + \chi(p)p^{k-1}a_{\frac{n}{p}}(f)\right) e^{2\pi i n z}.$$

$\square$

We now mention the crucial result about Hecke operators which is that they form a simultaneously commuting family with the diamond operators:

Proposition 4.5.5. *Let p and q be primes and $d, e \in (\mathbb{Z}/N\mathbb{Z})^*$. The Hecke operators T_p and T_q and diamond operators $\langle d \rangle$ and $\langle e \rangle$ on $\mathcal{S}_k(\Gamma_1(N))$ form a simultaneously commuting family:*

$$T_pT_q = T_qT_p, \quad \langle d \rangle T_p = T_p \langle d \rangle, \quad \text{and} \quad \langle d \rangle \langle e \rangle = \langle e \rangle \langle d \rangle.$$

Proof. Showing the diamond and Hecke operators commute was Proposition 4.5.2. To show commutativity of the diamond operators, let $\gamma = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix}\right) \in \Gamma_0(N)$ and $\eta = \left(\begin{smallmatrix} * & * \\ * & e \end{smallmatrix}\right) \in \Gamma_0(N)$. Then

$$\gamma\eta \equiv \begin{pmatrix} * & * \\ 0 & de \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & ed \end{pmatrix} \equiv \eta\gamma \pmod{N}.$$

Therefore $\gamma\eta|_k = \eta\gamma|_k$ as operators and so for any $f \in \mathcal{S}_k(\Gamma_1(N))$, we have

$$\langle d \rangle \langle e \rangle f = f|_k \gamma\eta_k = f|_k \eta\gamma_k = \langle e \rangle \langle d \rangle f.$$

We now show that the Hecke operators commute. In view of Proposition 4.5.1 and linearity of the Hecke operators, it suffices to prove this for $f \in \mathcal{S}_k(N, \chi)$. Applying Proposition 4.5.4 twice, for any $n \geq 1$ we compute

$$\begin{aligned}a_n(T_pT_qf) &= a_{np}(T_qf) + \chi(p)p^{k-1}a_{\frac{n}{p}}(T_qf) \\ &= a_{npq}(f) + \chi(q)q^{k-1}a_{\frac{np}{q}}(f) + \chi(p)p^{k-1}(a_{\frac{nq}{p}}(f) + \chi(q)q^{k-1}a_{\frac{n}{pq}}(f)) \\ &= a_{npq}(f) + \chi(q)q^{k-1}a_{\frac{np}{q}}(f) + \chi(p)p^{k-1}a_{\frac{nq}{p}}(f) + \chi(pq)(pq)^{k-1}a_{\frac{n}{pq}}(f).\end{aligned}$$

The last expression is symmetric in p and q so that $a_n(T_pT_qf) = a_n(T_qT_pf)$ for all $n \geq 1$. Since all of the Fourier coefficients are equal, we get

$$T_pT_qf = T_qT_pf.$$

$\square$

We can use Proposition 4.5.5 to construct diamond operators $\langle m\rangle$ and Hecke operators T_m for all $m \ge 1$. The **diamond operator** $\langle m\rangle : \mathcal{S}_k(\Gamma_1(N)) \to \mathcal{S}_k(\Gamma_1(N))$ is defined to be the linear operator given by

$$\langle m\rangle = \begin{cases} \langle m\rangle \text{ with } m \text{ taken modulo } N & \text{if } (m,N)=1, \\ 0 & \text{if } (m,N)>1. \end{cases}$$

Now for the Hecke operators. If $m = p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$ is the prime decomposition of m, then we define the m-th **Hecke operator** $T_m : \mathcal{S}_k(\Gamma_1(N)) \to \mathcal{S}_k(\Gamma_1(N))$ to be the linear operator given by

$$T_m = \prod_{1\le i\le k} T_{p_i^{r_i}},$$

where T_{p^r} is defined inductively by

$$T_{p^r} = \begin{cases} T_pT_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}} & \text{if } p \nmid N, \\ T_p^r & \text{if } p \mid N, \end{cases}$$

for all $r \ge 2$. Note that when $m=1$, the product is empty and so T_1 is the identity operator. By Proposition 4.5.5, the Hecke operators T_m are multiplicative but not completely multiplicative in m. Moreover, they commute with the diamond operators $\langle m\rangle$. Using these definitions, Propositions 4.5.4 and 4.5.5, a more general formula for how the Hecke operators T_m act on Fourier coefficients can be derived:

Proposition 4.5.6. *Let $f \in \mathcal{S}_k(\Gamma_1(N))$ have Fourier coefficients $a_n(f)$. Then for $m \ge 1$ with $(m,N)=1$,*

$$(T_mf)(z) = \sum_{n\ge1}\left(\sum_{d\mid(n,m)} d^{k-1}a_{\frac{nm}{d^2}}(\langle d\rangle f)\right)e^{2\pi inz},$$

is the Fourier series of T_mf. Moreover, if $f \in \mathcal{S}_k(N,\chi)$, then

$$(T_mf)(z) = \sum_{n\ge1}\left(\sum_{d\mid(n,m)} \chi(d)d^{k-1}a_{\frac{nm}{d^2}}(f)\right)e^{2\pi inz}.$$

Proof. In view of Proposition 4.5.1 and linearity of the Hecke operators, we may assume $f \in \mathcal{S}_k(N,\chi)$. Therefore we only need to verify the second formula. When $m=1$ the result is obvious and when $m=p$, we have

$$\sum_{d\mid(n,p)} \chi(d)d^{k-1}a_{\frac{np}{d^2}}(f) = a_{np}(f) + \chi(p)p^{k-1}a_{\frac{n}{p}}(f),$$

which is the result obtained from Proposition 4.5.4. By induction assume that the desired formula holds for $m = 1, p, \dots, p^{r-1}$. Using the definition of T_{p^r} and Proposition 4.5.4, for any $n \ge 1$ we compute

$$\begin{aligned} a_n(T_{p^r}f) &= a_n(T_pT_{p^{r-1}}f) - \chi(p)p^{k-1}a_n(T_{p^{r-2}}f) \\ &= a_{np}(T_{p^{r-1}}f) + \chi(p)p^{k-1}a_{\frac{n}{p}}(T_{p^{r-1}}f) - \chi(p)p^{k-1}a_n(T_{p^{r-2}}f). \end{aligned}$$

By our induction hypothesis, this last expression is

$$\sum_{d\mid(np,p^{r-1})} \chi(d)d^{k-1}a_{\frac{np^r}{d^2}}(f) + \chi(p)p^{k-1}\sum_{d\mid\left(\frac{n}{p},p^{r-1}\right)} \chi(d)d^{k-1}a_{\frac{np^{r-2}}{d^2}}(f) - \chi(p)p^{k-1}\sum_{d\mid(n,p^{r-2})} \chi(d)d^{k-1}a_{\frac{np^{r-2}}{d^2}}(f).$$

Write the first sum as

$$\sum_{d\mid(np,p^{r-1})}\chi(d)d^{k-1}a_{\frac{np^r}{d^2}}(f)=a_{np^r}(f)+\sum_{d\mid(n,p^{r-2})}\chi(d)d^{k-1}a_{\frac{np^{r-2}}{d^2}}(f),$$

and observe that the sum on the right-hand side cancels the entire third term above. Therefore our expression reduces to

$$\begin{aligned}a_{np^r}(f)+\chi(p)p^{k-1}\sum_{d\mid\left(\frac{n}{p},p^{r-1}\right)}\chi(d)d^{k-1}a_{\frac{np^{r-2}}{d^2}}(f)&=a_{np^r}(f)+\sum_{d\mid\left(\frac{n}{p},p^{r-1}\right)}\chi(dp)(dp)^{k-1}a_{\frac{np^{r-2}}{d^2}}(f)\\&=a_{np^r}(f)+\sum_{\substack{d\mid(n,p^r)\\d\neq1}}\chi(d)d^{k-1}a_{\frac{np^r}{d^2}}(f)\\&=\sum_{d\mid(n,p^r)}\chi(d)d^{k-1}a_{\frac{np^r}{d^2}}(f),\end{aligned}$$

where in the second line we have performed the change of variables $dp\to d$ in the sum. This proves the claim when $m=p^r$ for all $r\geq 0$. By multiplicativity of the Hecke operators, it suffices to prove the claim when $m=p^rq^s$ for another prime q and some $s\geq 0$. We compute

$$\begin{aligned}a_n(T_{p^rq^s}f)&=a_n(T_{p^r}T_{q^s}f)\\&=\sum_{d_1\mid(n,p^r)}\chi(d_1)d_1^{k-1}a_{\frac{np^r}{d_1^2}}(T_{q^s}f)\\&=\sum_{d_1\mid(n,p^r)}\chi(d_1)d_1^{k-1}\sum_{d_2\mid\left(\frac{np^r}{d_1^2},q^s\right)}\chi(d_2)d_2^{k-1}a_{\frac{np^rq^s}{(d_1d_2)^2}}(f)\\&=\sum_{d_1\mid(n,p^r)}\sum_{d_2\mid\left(\frac{np^r}{d_1^2},q^s\right)}\chi(d_1d_2)(d_1d_2)^{k-1}a_{\frac{np^rq^s}{(d_1d_2)^2}}(f).\end{aligned}$$

Summing over pairs (d_1,d_2) of divisors of (n,p^r) and $\left(\frac{np^r}{d^2},q^s\right)$ respectively is the same as summing over divisors d of (n,p^rq^s). Indeed, because p and q are relative prime, any such d is of the form $d=d_1d_2$ where $d_1\mid(n,p^r)$ and $d_2\mid\left(\frac{np^r}{d_2^2},q^s\right)$. Therefore the double sum becomes

$$\sum_{d\mid(n,p^rq^s)}\chi(d)d^{k-1}a_{\frac{np^rq^s}{d^2}}(f).$$

This completes the proof. □

The diamond and Hecke operators turn out to be normal on the subspace of cusp forms. To prove this fact, we will require a lemma:

Lemma 4.5.2. *Let* Γ *be a congruence subgroup and let* $\alpha\in\mathrm{GL}_2^+(\mathbb{Q})$*. Then there exist* $\beta_1,\ldots,\beta_n\in\mathrm{GL}_2^+(\mathbb{Q})$*, where* $n=[\Gamma:\alpha^{-1}\Gamma\alpha\cap\Gamma]=[\Gamma:\alpha\Gamma\alpha^{-1}\cap\Gamma]$*, and such that*

$$\Gamma\alpha\Gamma=\bigcup_j\Gamma\beta_j=\bigcup_j\beta_j\Gamma.$$

Proof. Apply Lemma 3.3.1 with the congruence subgroup $\alpha\Gamma\alpha^{-1}\cap\Gamma$ in place of Γ to get

$$[\mathrm{PSL}_2(\mathbb{Z}) : \alpha^{-1}\Gamma\alpha\cap\Gamma] = [\mathrm{PSL}_2(\mathbb{Z}) : \alpha\Gamma\alpha^{-1}\cap\Gamma].$$

Dividing both sides by $[\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]$ gives

$$[\Gamma : \alpha^{-1}\Gamma\alpha\cap\Gamma] = [\Gamma : \alpha\Gamma\alpha^{-1}\cap\Gamma].$$

Therefore we can find coset representatives $\gamma_1,\ldots,\gamma_n\in\Gamma$ and $\widetilde{\gamma}_1,\ldots,\widetilde{\gamma}_n\in\Gamma$ such that

$$\Gamma = \bigcup_j(\alpha^{-1}\Gamma\alpha\cap\Gamma)\gamma_j = \bigcup_j(\alpha^{-1}\Gamma\alpha\cap\Gamma)\widetilde{\gamma}_j^{-1}.$$

Invoking Lemma 4.4.1 twice, we can express each of these coset decompositions as an orbit decomposition:

$$\bigcup_j(\alpha^{-1}\Gamma\alpha\cap\Gamma)\gamma_j = \bigcup_j\Gamma\alpha\gamma_j \quad \text{and} \quad \bigcup_j(\alpha^{-1}\Gamma\alpha\cap\Gamma)\widetilde{\gamma}_j^{-1} = \bigcup_j\Gamma\alpha^{-1}\widetilde{\gamma}_j^{-1}.$$

It follows that

$$\Gamma = \bigcup_j\Gamma\alpha\gamma_j = \bigcup_j\widetilde{\gamma}_j\alpha\Gamma.$$

For each j the orbit spaces $\Gamma\alpha\gamma_j$ and $\widetilde{\gamma}_j\alpha\Gamma$ have nonempty intersection. For if they did we would have $\Gamma\alpha\gamma_j\subseteq\bigcup_{i\neq j}\widetilde{\gamma}_i\alpha\Gamma$ and thus $\Gamma\alpha\Gamma\subseteq\bigcup_{i\neq j}\widetilde{\gamma}_i\alpha\Gamma$. This contradicts the previous decomposition of Γ. Therefore we can find representatives $\beta_j\in\Gamma\alpha\gamma_j\cap\widetilde{\gamma}_j\alpha\Gamma$ for every j. Then β_j

$$\Gamma = \bigcup_j\Gamma\beta_j = \bigcup_j\beta_j\Gamma.$$

□

We can use Lemma 4.5.2 to compute adjoints:

Proposition 4.5.7. *Let Γ be a congruence subgroup and let $\alpha\in\mathrm{GL}_2^+(\mathbb{Q})$. Set $\alpha' = \det(\alpha)\alpha^{-1}$. Then the following are true:*

(i) If $\alpha^{-1}\Gamma\alpha\subseteq\mathrm{PSL}_2(\mathbb{Z})$, then for all $f\in\mathcal{S}_k(\Gamma)$ and $g\in\mathcal{S}_k(\alpha^{-1}\Gamma\alpha)$, we have

$$\langle f|_k\alpha, g\rangle_{\alpha^{-1}\Gamma\alpha} = \langle f, g|_k\alpha'\rangle_\Gamma.$$

(ii) For all $f,g\in\mathcal{S}_k(\Gamma)$, we have

$$\langle f[\Gamma\alpha\Gamma]_k, g\rangle = \langle f, g[\Gamma\alpha'\Gamma]_k\rangle.$$

In particular, if $\alpha^{-1}\Gamma\alpha = \Gamma$ then $|_k\alpha^* = |_k\alpha'$ and $[\Gamma\alpha\Gamma]_k^* = [\Gamma\alpha'\Gamma]_k$ as operators.

Proof. To prove (i) we first compute

$$\begin{aligned}
\langle f|_k\alpha, g\rangle_{\alpha^{-1}\Gamma\alpha} &= \frac{1}{V_{\alpha^{-1}\Gamma\alpha}}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}}(f|_k\alpha)(z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_{\alpha^{-1}\Gamma\alpha}}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}}\det(\alpha)^{k-1}j(\alpha,z)^{-k}f(\alpha z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\alpha^{-1}\Gamma\alpha}}\det(\alpha)^{k-1}j(\alpha,z)^{-k}f(\alpha z)\overline{g(z)}\mathrm{Im}(z)^k\,d\mu && \text{Lemma 3.3.1} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma}\det(\alpha)^{k-1}j(\alpha,\alpha^{-1}z)^{-k}f(z)\overline{g(\alpha^{-1}z)}\mathrm{Im}(\alpha^{-1}z)^k\,d\mu && z\to\alpha^{-1}z.
\end{aligned}$$

As α' acts as α^{-1} on $\mathbb{H}$, this latter integral is equivalent to

$$\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \det(\alpha)^{k-1} j(\alpha,\alpha' z)^{-k} f(z)\overline{g(\alpha' z)}\mathrm{Im}(\alpha' z)^k\,d\mu.$$

Moreover, applying the cocycle condition and the identities $\mathrm{Im}(\alpha' z) = \det(\alpha')\frac{\mathrm{Im}(z)}{|j(\alpha',z)|^2}$, $j(\alpha\alpha',z) = \det(\alpha)$, and $\det(\alpha') = \det(\alpha)$ together, we can further rewrite the integral as

$$\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \det(\alpha')^{k-1}\overline{j(\alpha',z)^{-k}} f(z)\overline{g(\alpha' z)}\mathrm{Im}(z)^k\,d\mu.$$

Reversing the first computation in the start of the proof but applied to this integral shows that that

$$\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \det(\alpha')^{k-1}\overline{j(\alpha',z)^{-k}} f(z)\overline{g(\alpha' z)}\mathrm{Im}(z)^k\,d\mu = \langle f, g|_k\alpha'\rangle_\Gamma,$$

which completes the proof of (i). To prove (ii), Lemma 4.5.2 implies the coset decomposition $\Gamma\alpha\Gamma = \bigcup_j \Gamma\beta_j$ so that we can use the β_j as representatives for the $[\Gamma\alpha\Gamma]_k$ operator. As the β_j also satisfy $\Gamma\alpha\Gamma = \bigcup_j \beta_j\Gamma$, upon inverting β_j and noting that $\beta_j \in \Gamma\alpha$, we obtain $\Gamma\alpha^{-1}\Gamma = \bigcup_j \Gamma\beta_j^{-1}$. Since scalar multiplication commutes with matrices and the matrices in Γ have determinant 1, we conclude that $\Gamma\alpha'\Gamma = \bigcup_j \Gamma\beta_j'$ where $\beta_j' = \det(\beta_j)\beta_j^{-1}$ (also $\det(\beta_j) = \det(\alpha)$). So we can use the β_j' as representatives in the $[\Gamma\alpha'\Gamma]_k$ operator. The statement now follows from (i). The last statement is obvious. $\square$

We can now prove that the diamond and Hecke operators are normal:

Proposition 4.5.8. *On $\mathcal{S}_k(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m are normal for all $m \geq 1$ with $(m,N) = 1$. Moreover, their adjoints are given by*

$$\langle m\rangle^* = \langle \overline{m}\rangle \quad \text{and} \quad T_p^* = \langle\overline{p}\rangle T_p.$$

Proof. By multiplicativity of the diamond and Hecke operators, it suffices to prove the two formulas when $m = p$ for a prime $p \nmid N$. We will prove the adjoint formula for diamond operators first. Let $\alpha = \left(\begin{smallmatrix}\overline{p} & * \\ * & p\end{smallmatrix}\right) \in \Gamma_0(N)$ and $\alpha' = \det(\alpha)\alpha^{-1} = \left(\begin{smallmatrix}p & * \\ * & \overline{p}\end{smallmatrix}\right) \in \Gamma_0(N)$. Then Proposition 4.5.7 gives

$$\langle p\rangle^* = |_k\alpha' = \langle\overline{p}\rangle.$$

This proves the adjoint formula for the diamond operators and normality follows from multiplicativity. For the Hecke operators, let $\beta_j = \left(\begin{smallmatrix}1 & j \\ 0 & p\end{smallmatrix}\right)$ and $\beta_\infty = \left(\begin{smallmatrix}pm & n \\ Np & p\end{smallmatrix}\right)$ for j taken modulo p and where m and n are chosen such that $\det\left(\left(\begin{smallmatrix}m & n \\ N & p\end{smallmatrix}\right)\right) = 1$. By Proposition 4.5.3, $\{\beta_1,\ldots,\beta_{p-1},\beta_\infty\}$ is a complete set of orbit representatives for T_p. Now set $\alpha = \left(\begin{smallmatrix}1 & 0 \\ 0 & p\end{smallmatrix}\right)$ and $\alpha' = \det(\alpha)\alpha^{-1} = \left(\begin{smallmatrix}p & 0 \\ 0 & 1\end{smallmatrix}\right)$. Observe that

$$\begin{pmatrix}1 & n \\ N & pm\end{pmatrix}^{-1}\alpha\begin{pmatrix}p & n \\ N & m\end{pmatrix} = \begin{pmatrix}mp & -n \\ -N & 1\end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & p\end{pmatrix}\begin{pmatrix}p & n \\ N & m\end{pmatrix} = \alpha'.$$

As $\left(\begin{smallmatrix}1 & n \\ N & pm\end{smallmatrix}\right) \in \Gamma_1(N)$ (note that $pm \equiv 1 \pmod{N}$ since $pm - Nn = 1$) and $\left(\begin{smallmatrix}p & n \\ N & m\end{smallmatrix}\right) \in \Gamma_0(N)$, the fact that $\Gamma_1(N)$ is normal in $\Gamma_0(N)$ yields

$$\Gamma_1(N)\alpha'\Gamma_1(N) = \Gamma_1(N)\alpha\Gamma_1(N)\begin{pmatrix}p & n \\ N & m\end{pmatrix}.$$

As $m \equiv \overline{p} \pmod{N}$, the above identity and Proposition 4.5.7 together give

$$T_p^* = [\Gamma_1(N)\alpha'\Gamma_1(N)]_k = \langle\overline{p}\rangle T_p.$$

This proves the adjoint formula for the Hecke operators and normality follows from multiplicativity. $\square$

Note that on $\mathcal{S}_k(\Gamma_1(1))$, all of the diamond operators are the identity and therefore $T_p^* = T_p$ for all primes p. That is, the Hecke operators are self-adjoint (as are the diamond operators since they are the identity). Now suppose f is a non-constant cusp form. Let the eigenvalue of T_m for f be $\lambda_f(m)$. We say that the $\lambda_f(m)$ are the **Hecke eigenvalues** of f. If f is a simultaneous eigenfunction for all diamond operators $\langle m \rangle$ and Hecke operators T_m with $(m, N) = 1$, we call f an **eigenform**. If the condition $(m, N) = 1$ can be dropped, so that f is a simultaneous eigenfunction for all diamond and Hecke operators, we say f is a **Hecke eigenform**. In particular, on $\Gamma_1(1)\backslash\mathbb{H}$ all eigenforms are Hecke eigenforms. Now let f have Fourier coefficients $a_n(f)$. If f is a Hecke eigenform, then Proposition 4.5.6 immediately implies that the first Fourier coefficient of $T_m f$ is $a_m(f)$ and so

$$a_m(f) = \lambda_f(m)a_1(f),$$

for all $m \geq 1$. Therefore we cannot have $a_1(f) = 0$ for this would mean f is constant. So we can normalize f by dividing by $a_1(f)$ which guarantees that this Fourier coefficient is 1. It follows that

$$a_m(f) = \lambda_f(m),$$

for all $m \geq 1$. This normalization is called the **Hecke normalization** of f. The **Petersson normalization** of f is where we normalize so that $\langle f, f \rangle = 1$. In particular, any orthonormal basis of $\mathcal{S}_k(\Gamma_1(N))$ is Petersson normalized. From the spectral theorem we derive an important corollary:

Theorem 4.5.1. *$\mathcal{S}_k(\Gamma_1(N))$ admits an orthonormal basis of eigenforms.*

Proof. By Theorem 4.3.1, $\mathcal{S}_k(\Gamma_1(N))$ is finite dimensional. The claim then follows from the spectral theorem along with Propositions 4.5.5 and 4.5.8. □

The Hecke eigenvalues of Hecke eigenforms satisfy certain relations known as the **Hecke relations** for holomorphic forms:

Proposition 4.5.9 (Hecke relations, holomorphic version). *Let $f \in \mathcal{S}_k(N, \chi)$ be a Hecke eigenform with Hecke eigenvalues $\lambda_f(m)$. Then the Hecke eigenvalues are multiplicative and satisfy*

$$\lambda_f(n)\lambda_f(m) = \sum_{d \mid (n,m)} \chi(d)d^{k-1}\lambda_f\left(\frac{nm}{d^2}\right) \quad \text{and} \quad \lambda_f(nm) = \sum_{d \mid (n,m)} \mu(d)\chi(d)d^{k-1}\lambda_f\left(\frac{n}{d}\right)\lambda_f\left(\frac{m}{d}\right),$$

for all $n, m \geq 1$ with $(nm, N) = 1$. Moreover,

$$\lambda_f(p^r) = \lambda_f(p)^r,$$

for all $p \mid N$ and $r \geq 2$.

Proof. If necessary, Hecke normalize f. The multiplicativity of the Hecke eigenvalues now follows from the multiplicity of the Hecke operators. The first identity follows from computing the n-th Fourier coefficient of $T_m f$ in two different ways. On the one hand, use that f is a Hecke eigenform to get $\lambda_f(n)\lambda_f(m)$. On the other hand, use Proposition 4.5.6 to obtain $\sum_{d \mid (n,m)} \chi(d)d^{k-1}\lambda_f\left(\frac{nm}{d^2}\right)$. For the second identity, computing the p-th Fourier coefficient of T_p in two different ways just as before, we have

$$\chi(p)p^{k-1} = \lambda_f(p)^2 - \lambda_f(p^2),$$

provided $(p, N) = 1$. The second identity now follows from the first because $\lambda_f(n)$ is a specially multiplicative function in n (see Theorem A.2.2). The third identity follows from inspecting the first Fourier coefficient of $T_{p^r} f$ and noting that $T_{p^r} = T_p^r$ provided $p \mid N$. □

As an immediate consequence of the Hecke relations, the Hecke operators satisfy analogous relations:

Corollary 4.5.1. *The Hecke operators are multiplicative and satisfy*

$$T_nT_m = \sum_{d\mid(n,m)} \chi(d)d^{k-1}T_{\frac{nm}{d^2}} \quad \text{and} \quad T_{nm} = \sum_{d\mid(n,m)} \mu(d)\chi(d)d^{k-1}T_{\frac{n}{d}}T_{\frac{m}{d}},$$

for all $n, m \geq 1$ with $(nm, N) = 1$.

Proof. This is immediate from Theorem 4.5.1 and the Hecke relations. □

The identities in Corollary 4.5.1 can also be established directly. Moreover, the first identity is symmetric in n and m so it can be used to show that the Hecke operators commute.

4.6 Atkin-Lehner Theory

So far, our entire theory of holomorphic forms has started with a fixed congruence subgroup of some level. Atkin-Lehner theory, or the theory of oldforms & newforms, allows us to discuss holomorphic forms in the context of moving between levels. In this setting, we will only deal with congruence subgroups of the form $\Gamma_1(N)$ and $\Gamma_0(N)$. The easiest way lift a holomorphic form from a smaller level to a larger level is to observe that if $M \mid N$, then $\Gamma_1(N) \leq \Gamma_1(M)$ so there is a natural inclusion $\mathcal{S}_k(\Gamma_1(M)) \subseteq \mathcal{S}_k(\Gamma_1(N))$. There is a less trivial way of lifting from $\mathcal{S}_k(\Gamma_1(M))$ to $\mathcal{S}_k(\Gamma_1(N))$. For any $d \mid \frac{N}{M}$, let $\alpha_d = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If $f \in \mathcal{S}_k(\Gamma_1(M))$, we consider

$$(f|_k\alpha_d)(z) = \det(\alpha_d)^{k-1}j(\alpha_d, z)^{-k}f(\alpha_d z) = d^{k-1}f(dz).$$

It turns out that $|_k\alpha_d$ maps $\mathcal{S}_k(\Gamma_1(M))$ into $\mathcal{S}_k(\Gamma_1(N))$ and more:

Proposition 4.6.1. *Let M and N be positive integers such that $M \mid N$. For any $d \mid \frac{N}{M}$, $|_k\alpha_d$ maps $\mathcal{S}_k(\Gamma_1(M))$ into $\mathcal{S}_k(\Gamma_1(N))$. In particular, $|_k\alpha_d$ takes $\mathcal{S}_k(M, \chi)$ into $\mathcal{S}_k(N, \chi)$.*

Proof. By Proposition 4.5.1, it suffices to verify only the latter statement so we may assume $f \in \mathcal{S}_k(M, \chi)$. It is clear that holomorphy is satisfied for $f|_k\alpha_d$. To verify modularity, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right) \in \Gamma_1(N)$. Then

$$\alpha_d\gamma\alpha_d^{-1} = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d' \end{pmatrix}\begin{pmatrix} d^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & bd \\ d^{-1}c & d' \end{pmatrix} = \gamma',$$

where $\gamma' = \left(\begin{smallmatrix} a & bd \\ d^{-1}c & d' \end{smallmatrix}\right)$. Since $c \equiv 0 \pmod N$ and $d \mid \frac{N}{M}$, we deduce that $d^{-1}c \equiv 0 \pmod M$. So $\gamma' \in \Gamma_1(M)$ and therefore $\alpha_d\Gamma_1(N)\alpha_d^{-1} \subseteq \Gamma_1(M)$, or equivalently, $\Gamma_1(N) \subseteq \alpha_d^{-1}\Gamma_1(M)\alpha_d$. Writing $\gamma = \alpha_d^{-1}\gamma'\alpha_d$, we see that $\chi(\gamma') = \chi(\gamma)$ and $j(\gamma', \alpha_d z) = j(\gamma, z)$ and so

$$\begin{aligned} (f|_k\alpha_d)(\gamma z) &= d^{k-1}f(d\gamma z) \\ &= d^{k-1}f(d\alpha_d^{-1}\gamma'\alpha_d z) \\ &= d^{k-1}f(\gamma'\alpha_d z) \\ &= \chi(\gamma')j(\gamma', \alpha_d z)d^{k-1}f(\alpha_d z) && \text{modularity} \\ &= \chi(\gamma)j(\gamma, z)d^{k-1}f(\alpha_d z) \\ &= \chi(\gamma)j(\gamma, z)d^{k-1}f(dz) \\ &= \chi(\gamma)j(\gamma, z)(f|_k\alpha_d)(z). \end{aligned}$$

This verifies the modularity of $f|_k\alpha_d$. For the growth condition, let $\sigma_\mathfrak{a}$ be a scaling matrix for the cusp $\mathfrak{a}$ of $\Gamma_1(M)\backslash\mathbb{H}$. Then $\alpha_d\sigma_\mathfrak{a}$ takes ∞ to an element of $\mathbb{Q}\cup\{\infty\}$ since $\alpha_d \in \mathrm{GL}_2^+(\mathbb{Q})$. In other words, $\alpha_d\sigma_\mathfrak{a}\infty = \mathfrak{b}$ for some cusp $\mathfrak{b}$ of $\Gamma_1(N)\backslash\mathbb{H}$. Then the cocycle condition implies

$$j(\sigma_\mathfrak{a}, z)^{-k}(f|_k\alpha_d)(\sigma_\mathfrak{a}z) = \det(\alpha_d)^{k-1}j(\alpha_d\sigma_\mathfrak{a}, z)^{-k}f(\alpha_d\sigma_\mathfrak{a}z),$$

and the growth condition follows from that of f. In particular, $f|_k\alpha_d$ is a cusp form since f is. $\square$

We can now define oldforms and newforms. For each divisor d of N, set

$$i_d : \mathcal{S}_k\left(\Gamma_1\left(\frac{N}{d}\right)\right) \times \mathcal{S}_k\left(\Gamma_1\left(\frac{N}{d}\right)\right) \to \mathcal{S}_k(\Gamma_1(N)) \qquad (f, g) \mapsto f + g|_k\alpha_d.$$

This map is well-defined by Proposition 4.6.1. The subspace of **oldforms** of level N is

$$\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N)) = \bigoplus_{p|N} \mathrm{Im}(i_p),$$

and the subspace of **newforms** of level N is

$$\mathcal{S}_k^{\mathrm{new}}(\Gamma_1(N)) = \mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))^\perp.$$

An element of these subspaces is called an **oldform** or **newform** respectively. Note that there are no oldforms of level 1. We will need a useful operator for the study of oldforms and newforms. Let

$$W_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix},$$

and note that $\det(W_N) = N$. We define the **Atkin-Lehner operator** ω_N to be the linear operator on $\mathcal{S}_k(\Gamma_1(N))$ given by

$$(\omega_N f)(z) = N^{1-\frac{k}{2}}(f|_kW_N)(z) = N^{\frac{k}{2}}j(W_N, z)^{-k}f(W_Nz) = (\sqrt{N}z)^{-k}f\left(-\frac{1}{Nz}\right).$$

As W_N is invertible, so is the Atkin-Lehner operator ω_N. It is not difficult to see how ω_N acts on $\mathcal{S}_k(\Gamma_1(N))$:

Proposition 4.6.2. *ω_N maps $\mathcal{S}_k(\Gamma_1(N))$ into itself. In particular, ω_N takes $\mathcal{S}_k(N,\chi)$ into $\mathcal{S}_k(N,\overline{\chi})$ and preserves the subspaces of oldforms and newforms. Moreover, ω_N is self-adjoint and*

$$\omega_N^2 f = (-1)^k f.$$

Proof. In light of Proposition 4.5.1, the first statement is a consequence of the latter ones. Therefore we may assume $f \in \mathcal{S}_k(N,\chi)$. Holomorphy is obvious. For modularity, note that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, we have

$$W_N\gamma W_N^{-1} = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & N^{-1} \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} d & -N^{-1}c \\ -Nb & a \end{pmatrix} = \gamma',$$

where $\gamma' = \left(\begin{smallmatrix} d & -N^{-1}c \\ -Nb & a \end{smallmatrix}\right) \in \Gamma_0(N)$. Thus $W_N\gamma = \gamma' W_N$ and it follows that

$$\begin{aligned}
(\omega_N f)(\gamma z) &= (\sqrt{N}\gamma z)^{-k} f(W_N\gamma z) \\
&= (\sqrt{N}\gamma z)^{-k} f(\gamma' W_N z) \\
&= \chi(\gamma')\left(\sqrt{N}\frac{az+b}{cz+d}\right)^{-k}\left(\frac{b}{z}+a\right)^{k} f\left(-\frac{1}{Nz}\right) && \text{modularity} \\
&= \chi(\gamma')\left(\sqrt{N}\frac{az+b}{cz+d}\right)^{-k}\left(\frac{z}{az+b}\right)^{-k} f\left(-\frac{1}{Nz}\right) \\
&= \chi(\gamma') j(\gamma,z)^k (\sqrt{N}z)^{-k} f\left(-\frac{1}{Nz}\right) \\
&= \overline{\chi}(\gamma) j(\gamma,z)^k (\sqrt{N}z)^{-k} f\left(-\frac{1}{Nz}\right) && ad \equiv 1 \pmod{N} \\
&= \overline{\chi}(\gamma) j(\gamma,z)^k (\omega_N f)(z).
\end{aligned}$$

This verifies modularity of $\omega_N f$. As for the growth condition, let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the cusp $\mathfrak{a}$. Then $W_N\sigma_{\mathfrak{a}}$ takes ∞ to an element of $\mathbb{Q}\cup\{\infty\}$ since $W_N \in \mathrm{GL}_2^+(\mathbb{Q})$. In other words, $W_N\sigma_{\mathfrak{a}}\infty = \mathfrak{b}$ for some cusp $\mathfrak{b}$. Then the cocycle condition implies

$$j(\sigma_{\mathfrak{a}},z)^{-k}(\omega_N f)(\sigma_{\mathfrak{a}} z) = N^{\frac{k}{2}} j(W_N\sigma_{\mathfrak{a}},z)^{-k} f(W_N\sigma_{\mathfrak{a}} z),$$

and the growth condition follows from that of f. In particular, $\omega_N f$ is a cusp form because f is. It follows that $\omega_N f \in \mathcal{S}_k(N,\overline{\chi})$. We now show that ω_N is self-adjoint. Indeed, since $\omega_N' = \det(W_N)\omega_N^{-1} = \omega_N$ (recall we are working in $\mathrm{PSL}_2(\mathbb{Z})$) Proposition 4.5.7 implies

$$\omega_N^* = N^{1-\frac{k}{2}}|_k W_N' = \omega_N.$$

Thus ω_N is self-adjoint. We now show that ω_N preserves the subspaces of oldforms and newforms. To show that ω_N preserves $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$, let $h = f + g|_k\alpha_p$ be in the image of i_p, with $p \mid N$, so that $i_p(f,g) = h$. Then it suffices to show

$$\omega_N i_p(f,g) = i_p(\omega_{\frac{N}{p}} p^{\frac{3k}{2}-1} g, \omega_{\frac{N}{p}} p^{1-\frac{k}{2}} f),$$

which will follow from the formulas

$$\omega_N f = (\omega_{\frac{N}{p}} p^{1-\frac{k}{2}} f)|_k\alpha_p \quad \text{and} \quad \omega_N(g|_k\alpha_p) = \omega_{\frac{N}{p}} p^{\frac{3k}{2}-1} g.$$

Both formulas follow immediately from the identities

$$W_N = W_{\frac{N}{p}}\alpha_p \quad \text{and} \quad \alpha_p W_N = pW_{\frac{N}{p}}.$$

Thus ω_N preserves the subspace of oldforms. To see that ω_N preserves the subspace of newforms as well, let f and g be a newform and an oldform respectively. The fact that ω_N is self-adjoint and preserves the subspace of oldforms gives

$$\langle \omega_N f, g\rangle = \langle f, \omega_N g\rangle = 0.$$

Hence $\omega_N f$ must also be a newform and so ω_N preserves the subspace of newforms. It remains to prove the formula. For this, observe $W_N^2 z = z$ and $j(W_N^2,z) = (-N)^k$ so that

$$(\omega_N^2 f)(z) = N^{2-k}(f|_k W_N^2)(z) = N^k j(W_N^2,z) f(W_N^2 z) = (-1)^k f(z). \qquad \square$$

Proposition 4.6.2 shows that ω_N is an involution if k is even and is at most of order 4. We now need to understand how the Atkin-Lehner operator interacts with the diamond and Hecke operators:

Proposition 4.6.3. *On $\mathcal{S}_k(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m satisfy the following adjoint formulas for all $m \geq 1$:*

$$\langle m\rangle^* = \omega_N\langle m\rangle\omega_N^{-1} \quad \text{and} \quad T_m^* = \omega_N T_m \omega_N^{-1}.$$

Proof. By multiplicativity of the diamond and Hecke operators, it suffices to prove the two adjoint formulas when $m = p$ for a prime p. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$. Then

$$W_N\gamma W_N^{-1} = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & N^{-1} \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} d & -N^{-1}c \\ -Nb & a \end{pmatrix} = \gamma',$$

where $\gamma' = \left(\begin{smallmatrix} d & -N^{-1}c \\ -Nb & a \end{smallmatrix}\right) \in \Gamma_1(N)$. In other words, W_N normalizes $\Gamma_1(N)$ so that $W_N\Gamma_1(N)W_N^{-1} = \Gamma_1(N)$. As $\Gamma_0(N) \leq \Gamma_1(N)$, the same holds for $\Gamma_0(N)$ as well. For the diamond operators, the formula is obvious when $p \mid N$ since $\langle p\rangle$ is the zero operator. So suppose $p \nmid N$ and let $\alpha = \left(\begin{smallmatrix} \overline{p} & * \\ * & p \end{smallmatrix}\right) \in \Gamma_0(N)$ and $\alpha' = \det(\alpha)\alpha^{-1} = \left(\begin{smallmatrix} p & * \\ * & \overline{p} \end{smallmatrix}\right) \in \Gamma_0(N)$. As $W_N\alpha W_N^{-1} = \alpha'$, Proposition 4.5.7 gives

$$\langle p\rangle^* = |_k\alpha' = \omega_N\langle p\rangle\omega_N^{-1}.$$

This is the desired adjoint formula for the diamond operators when $p \mid N$. Thus the adjoint formula holds for all the diamond operators. For the Hecke operators, let $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ and set $\alpha' = \det(\alpha)\alpha^{-1} = \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then $W_N\alpha W_N^{-1} = \alpha'$, and as W_N normalizes $\Gamma_1(N)$, Proposition 4.5.7 gives

$$T_p^* = [\Gamma_1(N)\alpha'\Gamma_1(N)]_k = \omega_N T_p\omega_N^{-1}.$$

This is the desired adjoint formula for the Hecke operators. $\square$

It turns out that the spaces of oldforms and newforms are invariant under the diamond and Hecke operators:

Proposition 4.6.4. *On $\mathcal{S}_k(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m preserve the subspaces of oldforms and newforms for all $m \geq 1$.*

Proof. By multiplicativity of the diamond and Hecke operators, it suffices to prove this when m is prime. Moreover, if $N = 1$ the result is trivial because there are no oldforms. Therefore we may assume $N > 1$ and hence consider a prime p with $p \mid N$. We will first show that the diamond and Hecke operators preserve $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$. Let $h = f + g|_k\alpha_p$ be in the image of i_p so that $i_p(f,g) = h$ and let q be a prime. For the diamond operators, it suffices to show

$$\langle q\rangle i_p(f,g) = i_p(\langle q\rangle f, \langle q\rangle g),$$

which will follow from the formulas

$$\langle q\rangle f = \langle q\rangle f \quad \text{and} \quad \langle q\rangle(g|_k\alpha_p) = (\langle q\rangle g)|_k\alpha_p.$$

In both of these formulas, the diamond operators on the left-hand sides and right-hand sides are on levels N and $\frac{N}{p}$ respectively. Both formulas are trivial if $q \mid N$ because then $\langle q\rangle$ is the zero operator. So suppose $q \nmid N$. If $\alpha = \left(\begin{smallmatrix} * & * \\ * & q \end{smallmatrix}\right) \in \Gamma_0(N)$, then $\alpha \in \Gamma_0\left(\frac{N}{p}\right)$ and the first formula follows. The second formula follows from the identity

$$\begin{pmatrix} * & * \\ * & q \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} * & * \\ * & q \end{pmatrix}.$$

Therefore the diamond operators preserve $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$. The case for the Hecke operators is slightly more involved. Now consider the Hecke operator T_q and suppose $q \neq p$. For this Hecke operator, it suffices to show

$$T_q i_p(f,g) = i_p(T_q f, T_q g),$$

which will follow from the formulas

$$T_q f = T_q f \quad \text{and} \quad T_q(g|_k\alpha_p) = (T_q g)|_k\alpha_p.$$

In both of these formulas, the Hecke operators on the left-hand sides and right-hand sides are on levels N and $\frac{N}{p}$ respectively. The first formula follows immediately from Proposition 4.5.4 since this shows that the action of the Hecke operators on the Fourier coefficients is identical. For the second formula, by Proposition 4.5.1 we may assume that $g \in \mathcal{S}\left(\frac{N}{p},\chi\right)$. Then $g|_k\alpha_p \in \mathcal{S}_k(N,\chi)$ (recall Proposition 4.6.1) and the second formula follows by comparing Fourier coefficients using Proposition 4.5.4. Thus the Hecke operators T_q, save for $q = p$, preserve $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$. Lastly, we consider the Hecke operator T_p. For this Hecke operator, it suffices to show

$$T_p i_p(f,g) = i_p(T_p f + p^{k-1}g, -\langle p\rangle f),$$

which will follow from the formulas

$$T_p f = T_p f - (\langle p\rangle f)|_k\alpha_p \quad \text{and} \quad T_p(g|_k\alpha_p) = p^{k-1}g.$$

In both of these formulas, the diamond and Hecke operators on the left-hand sides and right-hand sides are on levels N and $\frac{N}{p}$ respectively. Since $p \mid N$, Proposition 4.5.3 gives

$$T_p f = \sum_{j \pmod p} f\bigg|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} = T_p f - (\langle p\rangle f)|_k\alpha_p,$$

where the second equality follows regardless of whether $p \mid \frac{N}{p}$ or not because either $\langle p\rangle$ is the zero operator on level $\frac{N}{p}$ (in the case $p \mid \frac{N}{p}$) or it annihilates the extra term in the Hecke operator T_p on level $\frac{N}{p}$ (in the case $p \nmid \frac{N}{p}$). Therefore the first formula holds. Similarly, since $p \mid N$, Proposition 4.5.3 gives

$$T_p(g|_k\alpha_p)(z) = \sum_{j \pmod p} g\bigg|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix} = \sum_{j \pmod p} g\bigg|_k \begin{pmatrix} p & pj \\ 0 & p \end{pmatrix} = p^{k-1}g,$$

where the last equality follows because g is 1-periodic. This shows that the second formula holds. Thus the Hecke operator T_p preserves $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$. We have now shown that all of the diamond and Hecke operators preserve $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$. To show that they also preserve $\mathcal{S}_k^{\mathrm{new}}(\Gamma_1(N))$, let f and g be a newform and an oldform respectively. Then for any $m \geq 1$, Proposition 4.6.3 gives

$$\langle\langle m\rangle f, g\rangle = \langle f, \omega_N\langle m\rangle\omega_N^{-1}g\rangle = 0 \quad \text{and} \quad \langle T_m f, g\rangle = \langle f, \omega_N T_m \omega_N^{-1}g\rangle = 0,$$

where the second equality in either identity holds because the diamond and Hecke operators preserve $\mathcal{S}_k^{\mathrm{old}}(\Gamma_1(N))$ and so does the Atkin-Lehner operator ω_N (and hence its inverse as well) by Proposition 4.6.2. It follows that the diamond and Hecke operators preserve $\mathcal{S}_k^{\mathrm{new}}(\Gamma_1(N))$ and this completes the proof. $\square$

As a corollary, we deduce that these subspaces admit orthogonal bases of eigenforms:

Corollary 4.6.1. *$\mathcal{S}_k^{\text{old}}(\Gamma_1(N))$ and $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ admit orthonormal bases of eigenforms.*

Proof. This follows immediately from Theorem 4.5.1 and Proposition 4.6.4 □

Something quite amazing happens for the subspace in $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$; the condition $(m, N) = 1$ for eigenforms in a base can be removed. Therefore the eigenforms are actually eigenfunctions for all of the diamond and Hecke operators. We require a preliminary result whose proof is quite involved but it is not beyond the scope of this text (see [DS05] for a proof):

Lemma 4.6.1. *If $f \in \mathcal{S}_k(\Gamma_1(N))$ has Fourier coefficients $a_n(f)$ and is such that $a_n(f) = 0$ for all $n \geq 1$ whenever $(n, N) = 1$, then*

$$f = \sum_{p \mid N} p^{k-1} f_p|_k \alpha_p,$$

for some $f_p \in \mathcal{S}_k\left(\Gamma_1\left(\frac{N}{p}\right)\right)$.

The important observation to make about Lemma 4.6.1 is that if $f \in \mathcal{S}_k(\Gamma_1(N))$ is such that its n-th Fourier coefficients vanish when n is relatively prime to the level, then f must be an oldform. With this lemma we can prove the main theorem about $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$. The introduction of some language will be useful for the statement and its proof. We say that f is a **primitive Hecke eigenform** if it is a nonzero Hecke normalized Hecke eigenform in $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$. We can now prove the main result about newforms which is that Hecke eigenforms exist:

Theorem 4.6.1. *Let $f \in \mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ be an eigenform. Then the following hold:*

(i) f is a Hecke eigenform.

(ii) If g is any cusp form with the same Hecke eigenvalues at all primes, then $g = cf$ for some nonzero $c \in \mathbb{C}$.

Moreover, the primitive Hecke eigenforms in $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ form an orthogonal basis and each such eigenform lies in an eigenspace $\mathcal{S}_k(N, \chi)$.

Proof. First suppose $f \in \mathcal{S}_k(\Gamma_1(N))$ is an eigenform with Fourier coefficients $a_n(f)$. For $m \geq 1$ with $(m, N) = 1$, there exists $\lambda_f(m), \mu_f(m) \in \mathbb{C}$ such that $T_m f = \lambda_f(m) f$ and $\langle m \rangle f = \mu_f(m) f$. Actually, $\langle m \rangle f = \mu_f(m) f$ holds for all $m \geq 1$ because $\langle m \rangle$ is the zero operator if $(m, N) > 1$ and in this case we can take $\mu_f(m) = 0$. If we set $\chi(n) = \mu_f(m)$, then χ is a Dirichlet character modulo N. This follows because multiplicativity of $\langle m \rangle$ implies the same for χ and χ is N-periodic since $\langle m \rangle$ is N-periodic ($\langle m \rangle$ is defined by m taken modulo N). But then $\langle m \rangle f = \chi(m) f$ so that $f \in \mathcal{S}_k(N, \chi)$. As f is an eigenform, we also have $a_m(f) = \lambda_f(m) a_1(f)$ provided $(m, N) = 1$. So if $a_1(f) = 0$, Lemma 4.6.1 implies $f \in \mathcal{S}_k^{\text{old}}(\Gamma_1(N))$. With this fact in hand, we can prove the statements.

(i) The claim is trivial if f is zero, so assume otherwise. If $f \in \mathcal{S}_k^{\text{new}}(\Gamma_1(N))$, then $f \notin \mathcal{S}_k^{\text{old}}(\Gamma_1(N))$ and so by what we have shown $a_1(f) \neq 0$. Therefore we may Hecke normalize f so that $a_1(f) = 1$ and $a_m(f) = \lambda_f(m)$. Now set $g_m = T_m f - \lambda_f(m) f$ for any $m \geq 1$. By Proposition 4.6.4, $g_m \in \mathcal{S}_k^{\text{new}}(\Gamma_1(N))$. Moreover, g_m is an eigenform and its first Fourier coefficient is zero. But then $g_m \in \mathcal{S}_k^{\text{old}}(\Gamma_1(N))$ too and so $g_m = 0$ because $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ and $\mathcal{S}_k^{\text{old}}(\Gamma_1(N))$ are orthogonal subspaces. This means $T_m f = \lambda_f(m) f$ for any $m \geq 1$. Therefore f is a primitive Hecke eigenform and so is a Hecke eigenform before Hecke normalization.

(ii) Suppose g has the same Hecke eigenvalues at all primes. By multiplicativity of the Hecke operators, g is a Hecke eigenform. After Hecke normalization, f and g have the same Fourier coefficients and so are identical. It follows that before Hecke normalization $f = cg$ for some nonzero $c \in \mathbb{C}$.

Note that our initial remarks together with (i) show that each primitive Hecke eigenform f belongs to some eigenspace $\mathcal{S}_k(N, \chi)$. By Corollary 4.6.1, $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ admits an orthogonal basis of eigenforms which by (i) are Hecke eigenforms. As $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ is finite dimensional (because $\mathcal{S}_k(\Gamma_1(N))$ is), it follows that all of the primitive Hecke eigenforms form an orthogonal basis for $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ if we can show that they are linearly independent. So suppose to the contrary that we have a nontrivial linear relation

$$\sum_{1 \le i \le r} c_i f_i = 0,$$

for some primitive Hecke eigenforms f_i, nonzero constants c_i, and with r minimal. Note that $r \geq 2$ for else we do not have a nontrivial linear relation. Letting $m \geq 1$ applying the operator $T_m - \lambda_{f_1}(m)$ to our nontrivial linear relation gives

$$\sum_{2 \le i \le r} c_i(\lambda_{f_i}(m) - \lambda_{f_1}(m)) f_i = 0,$$

which has one less term. Since r was chosen to be minimal, this implies $\lambda_{f_i}(m) - \lambda_{f_1}(m)$ for all i. But m was arbitrary, so $f_i = f_1$ for all i by (ii). Hence $r = 1$ which is a contradiction. $\square$

Statement (i) in Theorem 4.6.1 implies that primitive Hecke eigenforms satisfy the Hecke relations for all $n, m \geq 1$. Statement (ii) is known as **multiplicity one** for holomorphic forms and can be interpreted as saying that a basis of newforms for $\mathcal{S}_k^{\text{new}}(\Gamma_1(N))$ contains one element per set of eigenvalues for the Hecke operators. We will now discuss conjugate cusp forms. For any $f \in \mathcal{S}_k(N, \chi)$, we define the **conjugate** $\overline{f}$ of f by

$$\overline{f}(z) = \overline{f(-\overline{z})}.$$

Note that if f has Fourier coefficients $a_n(f)$, then $\overline{f}$ has Fourier coefficients $\overline{a_n(f)}$. It turns out that $\overline{f}$ is indeed a holomorphic cusp form and behaves well with respect to the Hecke operators:

Proposition 4.6.5. *If $f \in \mathcal{S}_k(N, \chi)$, then $\overline{f} \in \mathcal{S}_k(N, \overline{\chi})$. Moreover,*

$$T_m \overline{f} = \overline{T_m f},$$

for all $m \geq 1$ with $(m, N) = 1$. In particular, if f is an eigenform with Hecke eigenvalues $\lambda_f(m)$ then f is too but with Hecke eigenvalues $\overline{\lambda_f(m)}$.

Proof. Holomorphy is clear so we need next need to verify the modularity. For this, let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ and note that $\gamma = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix} \in \Gamma_0(N)$. Then we compute

$$\begin{aligned}
\overline{f}(\gamma z) &= \overline{f(-\overline{\gamma z})} \\
&= \overline{f(\gamma'(-\overline{z}))} \\
&= \overline{\chi(\gamma) j(\gamma', -\overline{z})^k f(-\overline{z})} \\
&= \overline{\chi(\gamma) j(\gamma, z)^k f(-\overline{z})} \\
&= \overline{\chi}(\gamma) j(\gamma, z)^k \overline{f(-\overline{z})} \\
&= \overline{\chi}(\gamma) j(\gamma, z)^k \overline{f}(z),
\end{aligned}$$

which proves modularity. The growth condition follows immediately from that of f and thus $\overline{f}$ is a cusp form since f is. This proves the first statement. The second statement is immediate from Proposition 4.5.6. For the last statement, suppose $f \in \mathcal{S}_k(N,\chi)$ is an eigenform with Hecke eigenvalues $\lambda_f(m)$. Then by what we have already shown,

$$(T_m\overline{f})(z) = (\overline{T_m f})(z) = \overline{T_m f(-\overline{z})} = \overline{\lambda_f(m) f(-\overline{z})} = \overline{\lambda_f(m)}\overline{f}(z),$$

which completes the proof. □

In conjunction with Theorem 4.6.1, Proposition 4.6.5 implies that the primitive Hecke eigenforms in $\mathcal{S}_k^{\mathrm{new}}(\Gamma_1(N))$ are conjugate invariant and if $f \in \mathcal{S}_k(N,\chi)$ is such an eigenform then $\overline{f} \in \mathcal{S}_k(N,\overline{\chi})$ is as well. The crucial fact we need is how $\omega_N f$ is related to $\overline{f}$ when f is a primitive Hecke eigenform:

Proposition 4.6.6. *If $f \in \mathcal{S}_k(N,\chi)$ is a primitive Hecke eigenform, then*

$$\omega_N f = \omega_N(f)\overline{f},$$

where $\overline{f} \in \mathcal{S}_k(N,\overline{\chi})$ is a primitive Hecke eigenform and $\omega_N(f) \in \mathbb{C}$ is nonzero with $|\omega_N(f)| = 1$.

Proof. Let f have Hecke eigenvalues $\lambda_f(m)$. On the one hand, Theorem 4.6.1 and Proposition 4.6.5 together imply that $\overline{f}$ is a primitive Hecke eigenform with Hecke eigenvalues $\overline{\lambda_f(m)}$. One the other hand, Proposition 4.6.3 implies $\omega_N T_m = T_m^*\omega_N$ for all $m \geq 1$. Then

$$\langle T_m\omega_N f, \omega_N f\rangle = \langle \omega_N f, T_m^*\omega_N f\rangle = \langle \omega_N f, \omega_N T_m f\rangle = \langle \omega_N f, \lambda_m(f)\omega_N f\rangle = \overline{\lambda_f(m)}\langle \omega_N f, \omega_N f\rangle,$$

and it follows that $T_m\omega_N f = \overline{\lambda_f(m)}\omega_N f$. In other words, $\omega_N f$ is a Hecke eigenform with Hecke eigenvalues $\overline{\lambda_f(m)}$. Then multiplicity one gives

$$\omega_N f = \omega_N(f)\overline{f},$$

for some nonzero $\omega_N(f) \in \mathbb{C}$. Actually, by Proposition 4.6.2 we see that ω_N is at most of order 4 so necessarily $|\omega_N(f)| = 1$. □

4.7 The Ramanujan-Petersson Conjecture

We will now discuss a famous conjecture about the size of the Hecke eigenvalues of primitive Hecke eigenforms. Historically the conjecture was born from conjectures made about the **modular discriminant** Δ given by

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

which is a weight 12 primitive Hecke eigenform on $\Gamma_1(1)\backslash\mathbb{H}$ (see [DS05] for a proof). Therefore it is natural to begin our discussion here. It can be shown that the Fourier series of the modular discriminant is

$$\Delta(z) = \sum_{n\geq 1}\tau(n)e^{2\pi i n z},$$

where the $\tau(n)$ are integers with $\tau(1) = 1$ and $\tau(2) = -24$ (see [DS05] for a proof). The function $\tau : \mathbb{N} \to \mathbb{Z}$ is called **Ramanujan's τ function**. Ramanujan himself studied this function in his 1916 paper (see [Ram16]), and computed $\tau(n)$ for $1 \leq n \leq 30$. From these computations he conjectured that τ should satisfy the following three properties:

(i) If $(n,m) = 1$, then $\tau(nm) = \tau(n)\tau(m)$.

(ii) $\tau(p^n) = \tau(p^{n-1})\tau(p) - p^{11}\tau(p^{n-2})$ for all prime p.

(iii) $|\tau(p)| \le 2p^{\frac{11}{2}}$ for all prime p.

Note that (i) means τ is multiplicative. Moreover, (i) and (ii) are strikingly similar to the properties satisfied by the Hecke operators. In fact, (i) and (ii) are follow from the fact that Δ is a Hecke eigenform. Property (iii) turned out to be drastically more difficult to prove and is known as the classical **Ramanujan-Petersson conjecture**. To state the Ramanujan-Petersson conjecture for holomorphic forms, suppose $f \in \mathcal{S}_k(N,\chi)$ is a primitive Hecke eigenform with Hecke eigenvalues $\lambda_f(m)$. For each prime p, consider the polynomial

$$1 - \lambda_f(p)p^{-\frac{k-1}{2}}p^{-s} + \chi(p)p^{-2s}.$$

We call this the p-th **Hecke polynomial** of f. Let $\alpha_1(p)$ and $\alpha_2(p)$ denote the roots. From this quadratic, we have

$$\alpha_1(p) + \alpha_2(p) = \lambda_f(p)p^{-\frac{k-1}{2}} \quad \text{and} \quad \alpha_1(p)\alpha_2(p) = \chi(p).$$

Then the more general **Ramanujan-Petersson conjecture** for holomorphic forms is following statement:

Theorem 4.7.1 (Ramanujan-Petersson conjecture, holomorphic version). *Suppose $f \in \mathcal{S}_k(N,\chi)$ is a primitive Hecke eigenform with Hecke eigenvalues $\lambda_f(m)$. Let $\alpha_1(p)$ and $\alpha_2(p)$ be the roots of the p-th Hecke polynomial. Then for all primes p,*

$$|\lambda_f(p)| \le 2p^{\frac{k-1}{2}}.$$

Moreover, if $p \nmid N$, then

$$|\alpha_1(p)| = |\alpha_2(p)| = 1.$$

In the 1970's Deligne proved the Ramanujan-Petersson conjecture (see [Del71] and [Del74] for the full proof). The argument is significantly beyond the scope of this text, and in actuality follows from Deligne's work on the Weil conjectures (except in the case $k = 1$ which requires a modified argument). This requires understanding classical algebraic topology and ℓ-acid cohomology in addition to analytic number theory. As such, the proof of the Ramanujan-Petersson conjecture has been one of the biggest advances in analytic number theory in recent decades. Note that the Ramanujan-Petersson conjecture and the Hecke relations together give the bound $\lambda_f(m) \ll \sigma_0(m)m^{\frac{k-1}{2}} \ll_\varepsilon m^{\frac{k-1}{2}+\varepsilon}$ (recall Proposition A.3.1).

4.8 Twists of Holomorphic Forms

We can also twist holomorphic forms by Dirichlet characters. Let $f \in \mathcal{S}_k(N,\chi)$ have Fourier series

$$f(z) = \sum_{n\ge1} a_n(f)e^{2\pi inz},$$

and let ψ be a Dirichlet character modulo M. We define the **twisted holomorphic form** $f \otimes \psi$ of f twisted by ψ by the Fourier series

$$(f \otimes \psi)(z) = \sum_{n\ge1} a_n(f)\psi(n)e^{2\pi inz}.$$

In order for $f \otimes \psi$ to be well-defined, we need to prove that it is a holomorphic form. The following proposition proves this and more when ψ is primitive:

Proposition 4.8.1. *Suppose $f \in \mathcal{S}_k(N,\chi)$ and ψ is a primitive Dirichlet character of conductor q. Then $f \otimes \psi \in \mathcal{S}_k(Nq^2, \chi\psi^2)$.*

Proof. By Corollary 1.4.1, we can write

$$\begin{aligned}
(f \otimes \psi)(z) &= \sum_{n \ge 1} a_n(f)\psi(n)e^{2\pi i n z} \\
&= \sum_{n \ge 1} a_n(f)\left(\frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)e^{\frac{2\pi i r n}{q}}\right)e^{2\pi i n z} \\
&= \frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)\sum_{n \ge 1} a_n(f)e^{2\pi i n\left(z+\frac{r}{q}\right)} \\
&= \frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)f\left(z+\frac{r}{q}\right).
\end{aligned}$$

From this last expression, holomorphy is immediate since the sum is finite. For modularity, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma_0(Nq^2)$ and set $\gamma_r = \left(\begin{smallmatrix} 1 & \frac{r}{q} \\ 0 & 1\end{smallmatrix}\right)$ for every r modulo q. Then for r and r' modulo q, we compute

$$\gamma_r\gamma\gamma_{r'}^{-1} = \begin{pmatrix} 1 & \frac{r}{q} \\ 0 & 1\end{pmatrix}\begin{pmatrix} a & b \\ c & d\end{pmatrix}\begin{pmatrix} 1 & -\frac{r'}{q} \\ 0 & 1\end{pmatrix} = \begin{pmatrix} a+\frac{cr}{q} & b-\frac{ar'-dr}{q}-\frac{crr'}{q^2} \\ c & d-\frac{cr'}{q}\end{pmatrix}.$$

Since $c \equiv 0 \pmod{Nq^2}$, if we choose r' (for each r) such that $ar' \equiv dr \pmod q$, then $\gamma_r\gamma\gamma_r^{-1} \in \Gamma_0(N)$. Such a choice exists and is unique by Bézout's identity because a and d are relatively prime to q as $ad \equiv 1 \pmod{Nq^2}$. Making this choice and setting $\eta_r = \gamma_r\gamma\gamma_{r'}^{-1}$, we compute

$$f\left(\gamma z + \frac{r}{q}\right) = f(\gamma_r\gamma z) = f(\eta_r\gamma_{r'}z) = \chi(\eta_r)j(\eta_r,\gamma_{r'}z)^k f(\gamma_{r'}z) = \chi(\eta_r)j(\eta_r,\gamma_{r'}z)^k f\left(z+\frac{r'}{q}\right).$$

Moreover,

$$\chi(\eta_r)j(\eta_r,\gamma_{r'}z) = \chi\left(d-\frac{cr'}{q}\right)\left(c\gamma_{r'}z+d-\frac{cr'}{q}\right) = \chi(d)(cz+d) = \chi(\gamma)j(\gamma,z).$$

Together these two computations imply

$$f\left(\gamma z+\frac{r}{q}\right) = \chi(\gamma)j(\gamma,z)^k f\left(z+\frac{r'}{q}\right).$$

Now, as $ar' \equiv dr \pmod q$ and $ad \equiv 1 \pmod q$, we have

$$\overline{\psi}(r) = \overline{\psi}(a\overline{d}r') = \psi^2(d)\overline{\psi}(r') = \psi^2(\gamma)\overline{\psi}(r').$$

Putting everything together,

$$\begin{aligned}
(f\otimes\psi)(\gamma z) &= \frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)f\left(\gamma z+\frac{r}{q}\right) \\
&= \chi(\gamma)j(\gamma,z)^k\frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)f\left(z+\frac{r'}{q}\right) \\
&= \chi\psi^2(\gamma)j(\gamma,z)^k\frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r')f\left(z+\frac{r'}{q}\right) \\
&= \chi\psi^2(\gamma)j(\gamma,z)^k(f\otimes\psi)(z).
\end{aligned}$$

from which the modularity of $f \otimes \psi$ follows. For the growth condition, let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the cusp $\mathfrak{a}$ of $\Gamma_0(Nq^2)\backslash\mathbb{H}$. As $\gamma_r \in \mathrm{GL}_2^+(\mathbb{Q})$, $\gamma_r\sigma_{\mathfrak{a}}$ takes ∞ to an element of $\mathbb{Q} \cup \{\infty\}$. Thus $\gamma_r\sigma_{\mathfrak{a}}\infty = \mathfrak{b}$ for some cusp $\mathfrak{b}$ of $\Gamma_0(N)\backslash\mathbb{H}$. Then as $j(\gamma_r, \sigma_{\mathfrak{a}}z) = 1$, our previous work and the cocycle condition together imply

$$j(\sigma_{\mathfrak{a}}, z)^{-k}(f \otimes \psi)(\sigma_{\mathfrak{a}}z) = \frac{1}{\tau(\overline{\psi})} \sum_{r \pmod q} \overline{\psi}(r) j(\gamma_r\sigma_{\mathfrak{a}}, z)^{-k} f(\gamma_r\sigma_{\mathfrak{a}}z),$$

and the growth condition follows from that of f. Thus $f \otimes \psi$ is a cusp form since f is. $\square$

The generalization of Proposition 4.8.1 to all characters is slightly more involved. To this end, define operators U_p and V_p on $\mathcal{S}_k(\Gamma_1(N))$ to be the linear operators given by

$$(U_pf)(z) = \sum_{n \geq 1} a_{np}(f)e^{2\pi inz},$$

and

$$(V_pf)(z) = \sum_{n \geq 1} a_n(f)e^{2\pi inpz},$$

if f has Fourier series

$$f(z) = \sum_{n \geq 1} a_n(f)e^{2\pi inz}.$$

We will show that both U_p and V_p map $\mathcal{S}_k(\Gamma_1(N))$ into $\mathcal{S}_k(\Gamma_1(Np))$ and more:

Lemma 4.8.1. *For any prime p, U_p and V_p map $\mathcal{S}_k(\Gamma_1(N))$ into $\mathcal{S}_k(\Gamma_1(Np))$. In particular, U_p and V_p map $\mathcal{S}_k(N, \chi)$ into $\mathcal{S}_k(Np, \chi\chi_{p,0})$.*

Proof. In light of Proposition 4.5.1, the first statement follows from the second. As $N \mid Np$, $\Gamma_1(Np) \leq \Gamma_1(N)$ so that $f \in \mathcal{S}_k(\Gamma_1(Np))$ if $f \in \mathcal{S}_k(\Gamma_1(N))$. Now suppose $f \in \mathcal{S}_k(N, \chi)$. Similarly, $N \mid Np$ implies $\Gamma_0(Np) \leq \Gamma_0(N)$ so that $f \in \mathcal{S}_k(Np, \chi\chi_{p,0})$ for the modulus Np character $\chi\chi_{p,0}$. Therefore we may assume $f \in \mathcal{S}_k(Np, \chi\chi_{p,0})$. Now consider U_p. As $p \mid Np$, Equation (4.2) implies $U_p = T_p$ is the p-th Hecke operator on $\mathcal{S}_k(\Gamma_1(Np))$ and the claim follows from the definition of the Hecke operators and Proposition 4.5.4. Now consider V_p. We have

$$(V_pf)(z) = f(pz),$$

and the claim follows by regarding $f \in \mathcal{S}_k(Np, \chi\chi_{p,0})$ and that $\left(\begin{smallmatrix} p & 0 \\ 0 & p \end{smallmatrix}\right)$ lies in the center of $\mathrm{PSL}_2(\mathbb{Z})$. $\square$

We can now generalize Proposition 4.8.1 to all characters:

Proposition 4.8.2. *Suppose $f \in \mathcal{S}_k(N, \chi)$ and ψ is a Dirichlet character modulo M. Then $f \otimes \psi \in \mathcal{S}_k(NM^2, \chi\psi^2)$.*

Proof. Let $\widetilde{\psi}$ be the primitive character of conductor q inducing ψ. Then $\psi = \widetilde{\psi}\psi_{\frac{M}{q},0}$. As $\psi_{\frac{M}{q},0} = \prod_{p \mid \frac{M}{q}} \psi_{p,0}$, it suffices to prove the claim when ψ is primitive and when $\psi = \psi_{p,0}$. The primitive case follows from Proposition 4.8.1. So suppose $\psi = \psi_{p,0}$. Then

$$f \otimes \psi_{p,0} = f - V_pU_pf.$$

Now by Lemma 4.8.1, $V_pU_pf \in \mathcal{S}_k(Np^2, \chi\psi_{p,0}^2)$ (where we have written $\psi_{p,0}^2$ in place of $\chi_{p,0}$). Since we also have $f \in \mathcal{S}_k(Np^2, \chi\psi_{p,0}^2)$ (because $N \mid Np^2$ so that $\Gamma_1(Np^2) \leq \Gamma_1(N)$ and $\Gamma_0(Np^2) \leq \Gamma_0(N)$ and again writing $\psi_{p,0}^2$ in place of $\chi_{p,0}$), it follows that $f \otimes \psi_{p,0} \in \mathcal{S}_k(Np^2, \chi\psi_{p,0}^2)$. This proves the claim in the case $\psi = \psi_{p,0}$ and thus completes the proof. $\square$

In particular, Proposition 4.8.2 shows that $f \otimes \psi$ is well-defined for any Dirichlet character ψ.

Chapter 5

The Theory of Automorphic & Maass Forms

Maass forms are the non-holomorphic analog to holomorphic forms. They are real-analytic, eigenfunctions for a differential operator, invariant with respect to a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$, and satisfy a growth condition. The closely related automorphic forms satisfy fewer conditions and are necessary for the discussion of Maass forms in full generality. We introduce both Maass forms, automorphic forms, and their general theory. Throughout we assume that all of our congruence subgroups are reduced at infinity.

5.1 Automorphic & Maass Forms

Define $\varepsilon(\gamma, z)$ by

$$\varepsilon(\gamma, z) = \left(\frac{cz+d}{|cz+d|}\right),$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ and $z \in \mathbb{H}$. Note that $|\varepsilon(\gamma, z)| = 1$. Moreover, we have the relation

$$\varepsilon(\gamma, z) = \left(\frac{j(\gamma, z)}{|j(\gamma, z)|}\right).$$

As a consequence, the cocycle condition for $j(\gamma, z)$ implies

$$\varepsilon(\gamma'\gamma, z) = \varepsilon(\gamma', \gamma z)\varepsilon(\gamma, z),$$

and this is called the **cocycle condition** for $\varepsilon(\gamma, z)$. For any $k \in \mathbb{Z}$ and any $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ we define the **slash operator** $|_{\varepsilon,k} : C(\mathbb{H}) \to C(\mathbb{H})$ to be the linear operator given by

$$(f|_{\varepsilon,k}\gamma)(z) = \det(\gamma)^{-1}\varepsilon(\gamma, z)^{-k} f(\gamma z).$$

If ε is clear from content we will suppress this dependence accordingly. Note that if $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$, the slash operator takes the simpler form

$$(f|_{\varepsilon,k}\gamma)(z) = \varepsilon(\gamma, z)^{-k} f(\gamma z).$$

The cocycle condition implies that the slash operator is multiplicative. Indeed, if $\gamma, \gamma' \in \mathrm{GL}_2^+(\mathbb{Q})$, then

$$\begin{aligned}
((f|_{\varepsilon,k}\gamma')|_{\varepsilon,k}\gamma)(z) &= \det(\gamma)^{-1}\varepsilon(\gamma, z)^{-k}(f|_{\varepsilon,k}\gamma')(\gamma z) \\
&= \det(\gamma'\gamma)^{-1}\varepsilon(\gamma', \gamma z)^{-k}\varepsilon(\gamma, z)^{-k} f(\gamma'\gamma z) \\
&= \det(\gamma'\gamma)^{-1}\varepsilon(\gamma'\gamma, z)^{-k} f(\gamma'\gamma z) && \text{cocycle condition} \\
&= (f|_{\varepsilon,k}\gamma'\gamma)(z).
\end{aligned}$$

Any operator that commutes with the slash operators $|_{\varepsilon,k}\gamma$ for every $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ is said to be **invariant**. We define differential operators $R_k : C^\infty(\mathbb{H}) \to C^\infty(\mathbb{H})$ and $L_k : C^\infty(\mathbb{H}) \to C^\infty(\mathbb{H})$ to be the linear operators given by

$$R_k = \frac{k}{2} + y\left(i\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) \quad \text{and} \quad L_k = \frac{k}{2} + y\left(i\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right).$$

We call these operators the **Maass differential operators**. In particular, R_k is the **Maass raising operator** and L_k is the **Mass lowering operator**. The **Laplace operator** $\Delta_k : C^\infty(\mathbb{H}) \to C^\infty(\mathbb{H})$ is the linear operator given by

$$\Delta_k = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + iky\frac{\partial}{\partial x}.$$

When $k = 0$, we will suppress this dependence. Note that Δ is the usual Laplace operator on $\mathbb{H}$. Expanding the products $R_{k-2}L_k$ and $L_{k+2}R_k$ and invoking the Cauchy-Riemann equation

$$\frac{\partial}{\partial y} = i\frac{\partial}{\partial x},$$

we arrive at the identities

$$\Delta_k = -R_{k-2}L_k + \frac{k}{2}\left(1 - \frac{k}{2}\right) \quad \text{and} \quad \Delta_k = -L_{k+2}R_k - \frac{k}{2}\left(1 - \frac{k}{2}\right).$$

The Mass differential operators and the Laplace operator satisfy important relations (see [Bum97] for a proof):

Proposition 5.1.1. *The Laplace operator Δ_k is invariant. That is,*

$$\Delta_k(f|_{\varepsilon,k}\gamma) = \Delta_k(f)|_{\varepsilon,k}\gamma,$$

for all $f \in C^\infty(\mathbb{H})$ and $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$. Moreover, the Maass differential operators R_k and L_k satisfy

$$(R_kf)|_{\varepsilon,k+2}\gamma = R_k(f|_{\varepsilon,k}\gamma) \quad \text{and} \quad (R_kf)|_{\varepsilon,k-2}\gamma = L_k(f|_{\varepsilon,k}\gamma).$$

We will now introduce automorphic functions, automorphic forms, and Maass forms. Let Γ be a congruence subgroup of level N that is reduced at infinity and let χ be a Dirichlet character of conductor $q \mid N$. Set $\chi(\gamma) = \chi(d)$ for all for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. First up are the automorphic functions. We say that a function $f : \mathbb{H} \to \mathbb{C}$ is an **automorphic function** of **weight** k, **level** N, and **character** χ if following property is satisfied:

(i) $(f|_{\varepsilon,k}\gamma)(z) = \chi(\gamma)f(z)$ for all $\gamma \in \Gamma$.

We call property (i) is called the **automorphy condition** and we say that f is **automorphic**. The automorphy condition can equivalently be expressed as

$$f(\gamma z) = \chi(\gamma)\varepsilon(\gamma, z)^k f(z).$$

Note that automorphic functions admit Fourier series. Indeed, automorphy implies

$$f(z+1) = f\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} z\right) = f(z),$$

so that f is 1-periodic. Let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the $\mathfrak{a}$ cusp. As Lemma 3.1.1 implies $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}$ is a congruence subgroup, it follows by the cocycle condition that $f|_k\sigma_{\mathfrak{a}}$ is an automorphic function on $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}\backslash\mathbb{H}$ of the same weight and character as f. In particular, $f|_k\sigma_{\mathfrak{a}}$ is 1-periodic. Thus f admits a **Fourier series** at the $\mathfrak{a}$ cusp given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = \sum_{n\in\mathbb{Z}} a_{\mathfrak{a}}(n,y)e^{2\pi inx},$$

with **Fourier coefficients** $a_{\mathfrak{a}}(n,y)$. Observe that the sum is over all $n \in \mathbb{Z}$ since f may be unbounded as $z \to \infty$. If $\mathfrak{a} = \infty$, we will drop this dependence and in this case $(f|_k\sigma_{\mathfrak{a}}) = f$. Unlike holomorphic forms, this Fourier series need not converge pointwise anywhere. Next up are the automorphic forms. We say that a function $f : \mathbb{H} \to \mathbb{C}$ is an **automorphic form** of **weight** k, **eigenvalue** λ, **level** N, and **character** χ if following properties are satisfied:

(i) f is smooth on $\mathbb{H}$.

(ii) $(f|_{\varepsilon,k}\gamma)(z) = \chi(\gamma)f(z)$ for all $\gamma \in \Gamma$.

(iii) f is an eigenfunction for Δ_k with eigenvalue λ.

In property (iii), we will often let $s \in \mathbb{C}$ be such that $\lambda = s(1-s)$ and write $\lambda = \lambda(s)$ so that the eigenvalue can be determined by s. It turns turns out that Property (i) is implied by (iii). This is because Δ_k is an elliptic operator and any eigenfunction of an elliptic operator is automatically real-analytic and hence smooth (see [Eva22] for a proof in the weight zero case and [DFI02] for notes on the general case). Let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the $\mathfrak{a}$ cusp. As automorphic forms are automorphic functions, it follows by Proposition 5.1.1 that $f|_k\sigma_{\mathfrak{a}}$ is an automorphic form on $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}\backslash\mathbb{H}$ of the same weight, eigenvalue, and character as f. Moreover, $f|_k\sigma_{\mathfrak{a}}$ is also 1-periodic and so f admits a **Fourier series** at the $\mathfrak{a}$ cusp given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = \sum_{n\in\mathbb{Z}} a_{\mathfrak{a}}(n,y,s)e^{2\pi inx},$$

with **Fourier coefficients** $a_{\mathfrak{a}}(n,y,s)$. If $\mathfrak{a} = \infty$ or s is fixed, we will drop these dependencies accordingly and in this case $f|_k\sigma_{\mathfrak{a}} = f$. As f (and hence $f|_k\sigma_{\mathfrak{a}}$ too) is smooth, it converges uniformly to its Fourier series everywhere. The Fourier coefficients $a_{\mathfrak{a}}(n,y,s)$ are mostly determined by Δ_k. To see this, since $f|_k\sigma_{\mathfrak{a}}$ is smooth we may differentiate the Fourier series of $f|_k\sigma_{\mathfrak{a}}$ termwise. The fact that $f|_k\sigma_{\mathfrak{a}}$ is an eigenfunction for Δ_k with eigenvalue $\lambda(s)$ gives the ODE

$$(4\pi^2n^2y^2 - 2\pi nky)a_{\mathfrak{a}}(n,y,s) - y^2a_{\mathfrak{a},yy}(n,y,s) = \lambda(s)a_{\mathfrak{a}}(n,y,s).$$

If $n \neq 0$, this is a Whittaker equation. To see this, first put the ODE in homogeneous form

$$y^2a_{\mathfrak{a},yy}(n,y,s) - (4\pi^2n^2y^2 - 2\pi nky - \lambda(s))a_{\mathfrak{a}}(n,y,s) = 0.$$

Now make the change of variables $y \to \frac{y}{4\pi|n|}$ to get

$$y^2a_{\mathfrak{a},yy}(n,4\pi|n|y,s) - \left(\frac{y^2}{4} - \operatorname{sgn}(n)\frac{k}{2}y - \lambda(s)\right)a_{\mathfrak{a}}(n,4\pi|n|y,s) = 0,$$

where $\operatorname{sgn}(n) = \pm 1$ if n is positive or negative respectively. Diving by y^2 results in

$$a_{\mathfrak{a},yy}(n,4\pi|n|y,s) + \left(\frac{1}{4} - \frac{\operatorname{sgn}(n)\frac{k}{2}}{y} - \frac{\lambda(s)}{y^2}\right)a_{\mathfrak{a}}(n,4\pi|n|y,s) = 0.$$

As $\lambda(s) = s(1-s) = \frac{1}{4} - (s-\frac{1}{2})^2$, the above equation becomes

$$a_{\mathfrak{a},yy}(n, 4\pi|n|y, s) + \left(\frac{1}{4} - \frac{\operatorname{sgn}(n)\frac{k}{2}}{y} - \frac{\frac{1}{4} - \left(s-\frac{1}{2}\right)^2}{y^2}\right) a_{\mathfrak{a}}(n, 4\pi|n|y, s) = 0.$$

This is the Whittaker equation (see Appendix B.7). Since f has moderate growth at the cusps, the general solution is the Whittaker function $W_{\operatorname{sgn}(n)\frac{k}{2}, s-\frac{1}{2}}(4\pi|n|y)$. Therefore

$$a_{\mathfrak{a}}(n, y, s) = a_{\mathfrak{a}}(n, s) W_{\operatorname{sgn}(n)\frac{k}{2}, s-\frac{1}{2}}(4\pi|n|y),$$

for some coefficients $a_{\mathfrak{a}}(n, s)$. If $n = 0$, then the differential equation is a second order linear ODE which is

$$-y^2 a_{\mathfrak{a},yy}(0, y, s) = \lambda(s) a_{\mathfrak{a}}(0, y, s).$$

This is a Cauchy-Euler equation, and since s and $1-s$ are the two roots of $z^2 - z + \lambda$, the general solution is

$$a_{\mathfrak{a}}(0, y, s) = a_{\mathfrak{a}}^{+}(s) y^s + a_{\mathfrak{a}}^{-}(s) y^{1-s},$$

The coefficients $a_{\mathfrak{a}}(n, s)$ and $a_{\mathfrak{a}}^{\pm}(s)$ are the only part of the Fourier series that actually depend on the implicit congruence subgroup Γ. Using these coefficients, f admits **Fourier-Whittaker series** at the $\mathfrak{a}$ cusp given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = a_{\mathfrak{a}}^{+}(s) y^s + a_{\mathfrak{a}}^{-}(s) y^{1-s} + \sum_{n \neq 0} a_{\mathfrak{a}}(n, s) W_{\operatorname{sgn}(n)\frac{k}{2}, s-\frac{1}{2}}(4\pi|n|y) e^{2\pi i n x},$$

with **Fourier-Whittaker coefficients** $a_{\mathfrak{a}}^{\pm}(s)$ and $a_{\mathfrak{a}}(n, s)$. If $\mathfrak{a} = \infty$ or s is fixed, we will drop these dependencies accordingly and in this case $f|_k\sigma_{\mathfrak{a}} = f$. As f (and hence $f|_k\sigma_{\mathfrak{a}}$ too) is smooth, it converges uniformly to its Fourier-Whittaker series everywhere. Last up are the Maass forms. We say that a function $f : \mathbb{H} \to \mathbb{C}$ is a **Maass form** on $\Gamma\backslash\mathbb{H}$ of **weight** k, **eigenvalue** λ, **level** N, and **character** χ if the following properties are satisfied:

(i) f is smooth on $\mathbb{H}$.

(ii) $(f|_{\varepsilon,k}\gamma)(z) = \chi(\gamma) f(z)$ for all $\gamma \in \Gamma$.

(iii) f is an eigenfunction for Δ_k with eigenvalue λ.

(iv) $(f|_{\varepsilon,k}\alpha)(z) = O(y^n)$ for some $n \geq 1$ and all $\alpha \in \mathrm{PSL}_2(\mathbb{Z})$ (or equivalently $\alpha \in \mathrm{GL}_2^{+}(\mathbb{Q})$).

We say f is a **(Maass) cusp form** if the additional property is satisfied:

(v) For all cusps $\mathfrak{a}$ and any $y > 0$, we have

$$\int_0^1 (f|_k\sigma_{\mathfrak{a}})(x+iy)\,dx = 0.$$

Property (iv) is called the **growth condition** for Maass forms and we say f has **moderate growth at the cusps**. Clearly we only need to verify the growth condition on a set of scaling matrices for the cusps. Moreover, the equivalence in the growth condition follows exactly in the same way as for holomorphic forms. Indeed, the decomposition $\alpha = \gamma\eta$ for any $\alpha \in \mathrm{GL}_2^{+}(\mathbb{Q})$ with $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ and $\eta \in \mathrm{GL}_2^{+}(\mathbb{Q})$ of the form $\eta = \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$ along with the cocycle condition together imply

$$\varepsilon(\alpha, z) = \varepsilon(\gamma, \eta z),$$

and it follows that $(f|_{\varepsilon,k}\alpha)(z) = o(e^{2\pi y})$ for all $\alpha \in \mathrm{GL}_2^{+}(\mathbb{Q})$ which proves the forward implication. The reverse implication is trivial since $\mathrm{PSL}_2(\mathbb{Z}) \subset \mathrm{GL}_2^{+}(\mathbb{Q})$.

Remark 5.1.1. *Holomorphic forms embed into Maass forms. If $f(z)$ is a weight k holomorphic form on $\Gamma\backslash\mathbb{H}$, then $F(z) = \mathrm{Im}(4\pi z)^{\frac{k}{2}} f(z)$ is a weight k Maass form on $\Gamma\backslash\mathbb{H}$. This is because $\mathrm{Im}(\gamma z) = \frac{\mathrm{Im}(z)}{|j(\gamma,z)|^2}$ and $F(z)$ clearly has polynomial growth in $\mathrm{Im}(z)$. Moreover, as $f(z)$ is holomorphic, it satisfies the Cauchy-Riemann equations so that*

$$L_k(F(z)) = \left(\frac{k}{2} + y\left(i\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\right)(F(z)) = \frac{k}{2}F(z) - \frac{k}{2}F(z) = 0.$$

Therefore

$$\Delta_k(F(z)) = \left(-R_{k-2}L_k + \frac{k}{2}\left(1 - \frac{k}{2}\right)\right)(F(z)) = \frac{k}{2}\left(1 - \frac{k}{2}\right)F(z).$$

This means $F(z)$ is an eigenfunction for Δ_k with eigenvalue $\lambda\left(\frac{k}{2}\right)$.

One might expect that the Maass raising and lowering operators R_k and L_k act by changing the weight of a Maass form by ± 2 respectively. This is indeed the case:

Proposition 5.1.2. *If f is a weight k Maass form on $\Gamma\backslash\mathbb{H}$, then $R_k f$ and $L_k f$ are Maass forms on $\Gamma\backslash\mathbb{H}$ of weight $k+2$ and $k-2$ respectively and of the same eigenvalue, level, and character as f.*

Proof. This is immediate from the definition of Maass forms and Proposition 5.1.1. □

As a consequence of Proposition 5.1.2, it suffices to study Maass forms of weights $k = 0, 1$ although imposing this additional restriction is usually unnecessary. Let $\sigma_{\mathfrak{a}}$ be a scaling matrix for the $\mathfrak{a}$ cusp. As Maass forms are automorphic forms, it follows by Proposition 5.1.1 that $f|_k\sigma_{\mathfrak{a}}$ is an automorphic form on $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}}\backslash\mathbb{H}$ of the same weight, eigenvalue, and character as f. Moreover, $f|_k\sigma_{\mathfrak{a}}$ is also 1-periodic. Note that this means we only need to verify the growth condition as $y \to \infty$. As $f|_k\sigma_{\mathfrak{a}}$ is 1-periodic, f admits a **Fourier-Whittaker series** at the $\mathfrak{a}$ cusp given by

$$(f|_k\sigma_{\mathfrak{a}})(z) = a_{\mathfrak{a}}^{+}(s)y^{s} + a_{\mathfrak{a}}^{-}(s)y^{1-s} + \sum_{n\neq 0} a_{\mathfrak{a}}(n,s)W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)e^{2\pi inx},$$

with **Fourier-Whittaker coefficients** $a_{\mathfrak{a}}^{\pm}(s)$ and $a_{\mathfrak{a}}(n,s)$. As f (and hence $f|_k\sigma_{\mathfrak{a}}$ too) is smooth, it converges uniformly to its Fourier-Whittaker series everywhere. Moreover, property (v) implies that f is a cusp form if and only if $a_{\mathfrak{a}}^{\pm}(s) = 0$ for every cusp $\mathfrak{a}$. It is useful to specify the Whittaker function in the case of weight zero Maass forms. When $k = 0$, Theorem B.7.1 implies that the Fourier-Whittaker series of f at the $\mathfrak{a}$ cusp takes the form

$$(f|_k\sigma_{\mathfrak{a}})(z) = a_{\mathfrak{a}}^{+}(s)y^{s} + a_{\mathfrak{a}}^{-}(s)y^{1-s} + \sum_{n\neq 0} a_{\mathfrak{a}}(n,s)\sqrt{4|n|y}K_{s-\frac{1}{2}}(2\pi|n|y)e^{2\pi inx}.$$

In any case, we can also easily derive a bound for the size of the Fourier-Whittaker coefficients of cusp forms. Fix some $Y > 0$ and consider

$$\int_{\Gamma_\infty\backslash\mathbb{H}_Y} |(f|_k\sigma_{\mathfrak{a}})(z)|^2\,d\mu,$$

where $\mathbb{H}_Y$ is the half-plane defined by $Y \le \mathrm{Im}(z) \le 2Y$. This integral exists since $\Gamma_\infty\backslash\mathbb{H}_Y$ is compact. Substituting in the Fourier-Whittaker series of f at the $\mathfrak{a}$ cusp, this integral can be expressed as

$$\int_{Y}^{2Y}\int_{0}^{1}\sum_{n,m\neq 0} a_{\mathfrak{a}}(n,s)\overline{a_{\mathfrak{a}}(m,s)}W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)\overline{W_{\mathrm{sgn}(m)\frac{k}{2},s-\frac{1}{2}}(4\pi|m|y)}e^{2\pi i(n-m)x}\,\frac{dy}{y^2}.$$

Appealing to Fubini's theorem, we can interchange the sum and the two integrals. Upon making this interchange, the identity Equation (4.1) implies that the inner integral cuts off all of the terms except the diagonal $n = m$, resulting in

$$\sum_{n\neq 0}\int_{Y}^{2Y} |a_{\mathfrak{a}}(n,s)|^2|W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)|^2\,\frac{dy}{y^2}.$$

In particular, we see that this is a sum of nonnegative terms. Retaining only a single term in the sum, we have

$$|a_{\mathfrak{a}}(n,s)|^2\int_{Y}^{2Y} |W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)|^2\,\frac{dy}{y^2} \ll \int_{\Gamma_\infty\backslash\mathbb{H}_Y} |(f|_k\sigma_{\mathfrak{a}})(z)|^2\,d\mu.$$

Moreover, $|(f|_k\sigma_{\mathfrak{a}})(z)|^2$ is bounded on $\Gamma_\infty\backslash\mathbb{H}_Y$ because this space is compact, so that

$$\int_{\Gamma_\infty\backslash\mathbb{H}_Y} |(f|_k\sigma_{\mathfrak{a}})(z)|^2\,d\mu \ll \int_{Y}^{2Y}\int_{0}^{1}\frac{dx\,dy}{y^2} \ll \frac{1}{Y}.$$

Putting these two estimates together gives

$$|a_{\mathfrak{a}}(n,s)|^2\int_{Y}^{2Y} |W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)|^2\,\frac{dy}{y^2} \ll \frac{1}{Y}.$$

Taking $Y = \frac{1}{|n|}$ and making the change of variables $y \to \frac{y}{|n|}$, we obtain

$$a_{\mathfrak{a}}(n,s) \ll 1.$$

This bound is known as the **Hecke bound** for Maass forms. Using Lemma B.7.1, we have the estimate $W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y) = O((|n|y)^{\frac{k}{2}}e^{-2\pi|n|y})$. This estimate together with the Hecke bound and the Taylor series of the k-th derivative of $\frac{e^y}{1-e^y}$ gives

$$(f|_k\sigma_{\mathfrak{a}})(z) = O\left(y^{\frac{k}{2}}\sum_{n\neq 0}|n|^{\frac{k+1}{2}}e^{-2\pi|n|y}\right) = O\left(y^{\frac{k}{2}}\sum_{n\geq 1}n^k e^{-2\pi ny}\right) = O(y^{\frac{k}{2}}e^{-2\pi y}).$$

This implies $(f|_k\sigma_{\mathfrak{a}})(z)$ exhibits rapid decay. Accordingly, we say that f exhibits **rapid decay at the cusps**. Observe that $f|_k\sigma_{\mathfrak{a}}$ is then bounded on $\mathbb{H}$ and, in particular, f is bounded on $\mathbb{H}$.

5.2 Poincaré & Eisenstein Series

Let Γ be a congruence subgroup of level N. We will introduce two important classes of automorphic functions on $\Gamma\backslash\mathbb{H}$ namely the Poincaré and Eisenstein series. The Eisenstein series will be Maass forms while the Poincaré series will only be automorphic functions. Both of these classes are defined on a larger space $\mathbb{H}\times\{s\in\mathbb{C} : \sigma > 1\}$ and hence are functions of two variables. Let $m \geq 0$, $k \geq 0$, χ be a Dirichlet character with conductor $q \mid N$, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. Then the m-th **(automorphic) Poincaré series** $P_{m,k,\chi,\mathfrak{a}}(z,s)$ of weight k and character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp is defined by

$$P_{m,k,\chi,\mathfrak{a}}(z,s) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma}\overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma,z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)^s e^{2\pi im\sigma_{\mathfrak{a}}^{-1}\gamma z}.$$

We call m the **index** of $P_{m,k,\chi,\mathfrak{a}}(z,s)$. If $k = 0$, χ is the trivial character, or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly. We first show that $P_{m,k,\chi,\mathfrak{a}}(z,s)$ is well-defined. It suffices to show that the summands are independent of the representatives γ and $\sigma_{\mathfrak{a}}$. This has already been accomplished when we introduced the holomorphic Poincaré series for $\overline{\chi}(\gamma)$ and $e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}$. Now just as with the holomorphic Poincaré series, the set of representatives of $\sigma_{\mathfrak{a}}^{-1}\gamma$ is $\Gamma_\infty\sigma_{\mathfrak{a}}^{-1}\gamma$ and it remains to verify independence from multiplication on the left by an element of Γ_∞ namely η_∞. The cocycle relation implies

$$\varepsilon(\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma, z) = \varepsilon(\eta_\infty, \sigma_{\mathfrak{a}}^{-1}\gamma z)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z) = \varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z),$$

where the last equality follows because $\varepsilon(\eta_\infty, \sigma_{\mathfrak{a}}^{-1}\gamma z) = 1$ as $j(\eta_\infty, \sigma_{\mathfrak{a}}^{-1}\gamma z) = 1$. Thus $\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)$ is independent of the representatives γ and $\sigma_{\mathfrak{a}}$. Lastly, we have

$$\mathrm{Im}(\eta_\infty\sigma_{\mathfrak{a}}^{-1}\gamma z) = \mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z),$$

because η_∞ does not affect the imaginary part as it acts by translation. Therefore $\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)$ is independent of the representatives γ and $\sigma_{\mathfrak{a}}$ as well. We conclude that $P_{m,k,\chi,\mathfrak{a}}(z,s)$ is well-defined. We claim $P_{m,k,\chi,\mathfrak{a}}(z,s)$ is also locally absolutely uniformly convergent for $z \in \mathbb{H}$ and $\sigma > 1$. To see this, first recall that $|e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}| = e^{-2\pi m\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)} < 1$. Then the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma$ yields

$$P_{m,k,\chi,\mathfrak{a}}(z,s) \ll \sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{\mathrm{Im}(z)^\sigma}{|cz+d|^{2\sigma}},$$

and this latter series is locally absolutely uniformly convergent for $z \in \mathbb{H}$ and $\sigma > 1$ by Proposition B.8.1. Hence the same holds for $P_{m,k,\chi,\mathfrak{a}}(z,s)$. Verifying automorphy amounts to a computation:

$$\begin{aligned}
P_{m,k,\chi,\mathfrak{a}}(\gamma z,s) &= \sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma')\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma', \gamma z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma')\left(\frac{\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)}{\varepsilon(\gamma,z)}\right)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \varepsilon(\gamma,z)^k\sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma')\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma)\varepsilon(\gamma,z)^k\sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma')\overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma)\varepsilon(\gamma,z)^k\sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma'\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma)\varepsilon(\gamma,z)^k\sum_{\gamma'\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma')\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma', z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma' z)^s e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma'\gamma z} \\
&= \chi(\gamma)\varepsilon(\gamma,z)^k P_{m,k,\chi,\mathfrak{a}}(z,s),
\end{aligned}$$

where in the second line we have used the cocycle condition and in the second to last line we have used that $\gamma' \to \gamma'\gamma^{-1}$ is a bijection on Γ. As for the growth condition, let $\sigma_{\mathfrak{b}}$ be a scaling matrix for the cusp $\mathfrak{b}$. Then the bound $|e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{b}} z}| = e^{-2\pi m\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma\sigma_{\mathfrak{b}} z)} < 1$, cocycle condition, and the Bruhat decomposition for $\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{b}}$ together give

$$\varepsilon(\sigma_{\mathfrak{b}}, z)^{-k}P_{m,k,\chi,\mathfrak{a}}(\sigma_{\mathfrak{b}} z,s) \ll \mathrm{Im}(z)^\sigma \sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{1}{|cz+d|^{2\sigma}}.$$

Now decompose this sum as

$$\sum_{(c,d)\in\mathbb{Z}^2-\{\mathbf{0}\}} \frac{1}{|cz+d|^{2\sigma}} = \sum_{d\neq 0}\frac{1}{d^{2\sigma}} + \sum_{c\neq 0}\sum_{d\in\mathbb{Z}}\frac{1}{|cz+d|^{2\sigma}} = 2\sum_{d\geq 1}\frac{1}{d^{2\sigma}} + 2\sum_{c\geq 1}\sum_{d\in\mathbb{Z}}\frac{1}{|cz+d|^{2\sigma}}.$$

Notice that the first sum is absolutely uniformly bounded provided $\sigma > 1$. Moreover, the exact same argument as for holomorphic Eisenstein series shows that the second sum is too. So for all $\mathrm{Im}(z) \geq 1$ and $\sigma > 1$, we have

$$\varepsilon(\sigma_{\mathfrak{b}}, z)^{-k} P_{m,k,\chi,\mathfrak{a}}(\sigma_{\mathfrak{b}} z, s) \ll \mathrm{Im}(z)^{\sigma} = o(e^{2\pi \mathrm{Im}(z)}),$$

provided $\mathrm{Im}(z) \geq 1$ and $\sigma > 1$. This verifies the growth condition. We collect this work as a theorem:

Theorem 5.2.1. *Let $m \geq 0$, $k \geq 0$, χ be a Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. For $\sigma > 1$, the Poincaré series*

$$P_{m,k,\chi,\mathfrak{a}}(z, s) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)^{s} e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma z},$$

is a smooth automorphic function on $\Gamma\backslash\mathbb{H}$.

For $m = 0$, we write $E_{k,\chi,\mathfrak{a}}(z, s) = P_{0,k,\chi,\mathfrak{a}}(z, s)$ and call $E_{k,\chi}(z)$ the **(Maass) Eisenstein series** of weight k and character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp. It is defined by

$$E_{k,\chi,\mathfrak{a}}(z, s) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)^{s}.$$

If $k = 0$, χ is the trivial character, or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly. It turns out that $E_{k,\chi,\mathfrak{a}}(z, s)$ is actually a Maass form. The only thing left to verify is that $E_{k,\chi,\mathfrak{a}}(z, s)$ is an eigenfunction for Δ_k. To see this, first observe that

$$\Delta_k(y^s) = \left(-y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + iky\frac{\partial}{\partial x}\right)(y^s) = \lambda(s)y^s.$$

Therefore $\mathrm{Im}(z)^s$ is an eigenfunction for Δ_k with eigenvalue $\lambda(s)$. Since Δ_k is invariant,

$$\Delta_k((\mathrm{Im}(\cdot)^s|_{\varepsilon,k}\gamma)(z)) = ((\Delta_k \mathrm{Im}(\cdot)^s)|_{\varepsilon,k}\gamma)(z) = \lambda(s)(\mathrm{Im}(\cdot)^s|_{\varepsilon,k}\gamma)(z),$$

and so $(\mathrm{Im}(\cdot)^s|_{\varepsilon,k}\gamma)(z) = \varepsilon(\gamma, z)^{-k}\mathrm{Im}(\gamma z)^s$ is also an eigenfunction for Δ_k with eigenvalue $\lambda(s)$ for all $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$. We immediately conclude that

$$\Delta_k(E_{k,\chi,\mathfrak{a}}(z, s)) = \lambda(s)E_{k,\chi,\mathfrak{a}}(z, s),$$

which shows $E_{k,\chi,\mathfrak{a}}(z, s)$ is also an eigenfunction for Δ_k with eigenvalue $\lambda(s)$. We collect this work as a theorem:

Theorem 5.2.2. *Let $k \geq 0$, χ be a Dirichlet character with conductor dividing the level, and $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$. For $\sigma > 1$, the Eisenstein series*

$$E_{k,\chi,\mathfrak{a}}(z, s) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)^{s},$$

is a weight k Maass form with eigenvalue $\lambda(s)$ and character χ on $\Gamma\backslash\mathbb{H}$.

5.3 Inner Product Spaces of Automorphic Functions

Let $\mathcal{A}_k(\Gamma,\chi)$ denote the space of all weight k automorphic functions with character χ on $\Gamma\backslash\mathbb{H}$ and let $\mathcal{A}_{k,\lambda}(\Gamma,\chi)$, $\mathcal{M}_{k,\lambda}(\Gamma,\chi)$, and $\mathcal{C}_{k,\lambda}(\Gamma,\chi)$ denote the associated subspaces of automorphic functions, Maass forms, and cusp forms of eigenvalue λ respectively. If $k=0$ or χ is the trivial character, we will suppress these dependencies. Note that if Γ_1 and Γ_2 are two congruence subgroups such that $\Gamma_1 \le \Gamma_2$, then we have the inclusion

$$\mathcal{A}_k(\Gamma_2,\chi) \subseteq \mathcal{A}_k(\Gamma_1,\chi),$$

and this respects the subspaces of automorphic forms, Maass forms, and cusp forms. So in general, the smaller the congruence subgroup the more automorphic functions there are. Our goal is to construct a complex Hilbert space containing $\mathcal{C}_{k,\lambda}(\Gamma,\chi)$ for which we can apply a linear theory. The natural space to consider is the L^2-space for automorphic functions. We define the L^2-norm $||\cdot||_\Gamma$ for $f \in \mathcal{A}_k(\Gamma,\chi)$ by

$$||f||_\Gamma = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z)|^2\,d\mu\right)^{\frac{1}{2}}.$$

If the congruence subgroup is clear from context we will suppress the dependence upon Γ. As f is automorphic, the norm is independent of the choice of fundamental domain and hence well-defined. Let $\mathcal{L}_k(\Gamma,\chi)$ be the subspace of $\mathcal{A}_k(\Gamma,\chi)$ consisting of those functions with bounded L^2-norm and let $\mathcal{L}_{k,\lambda}(\Gamma,\chi)$ denote the associated subspace of automorphic forms. Moreover, if χ is the trivial character or if $k=0$, we will suppress these dependencies accordingly. Since this is an L^2-space, $\mathcal{L}_k(\Gamma,\chi)$ is an induced inner product space (because the parallelogram law is satisfied). In particular, for any $f,g \in \mathcal{L}_k(\Gamma,\chi)$ we define their **Petersson inner product** to be

$$\langle f,g\rangle_\Gamma = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)}\,d\mu.$$

If the congruence subgroup is clear from context we will suppress the dependence upon Γ. The integral is locally absolutely uniformly convergent by the Cauchy–Schwarz inequality and that $f,g \in \mathcal{L}_k(\Gamma,\chi)$. As f and g are automorphic, the integral is independent of the choice of fundamental domain. These two facts imply that the Petersson inner product is well-defined. We will continue to use this notion even if f and g do not belong to $\mathcal{L}_k(\Gamma,\chi)$ provided the integral is locally absolutely uniformly convergent. Just as was the case for holomorphic forms, the Petersson inner product is invariant with respect to the slash operator:

Proposition 5.3.1. *For any $f,g \in \mathcal{L}_k(\Gamma,\chi)$ and $\alpha \in \mathrm{PSL}_2(\mathbb{Z})$, we have*

$$\langle f|_k\alpha, g|_k\alpha\rangle_{\alpha^{-1}\Gamma\alpha} = \langle f,g\rangle_\Gamma.$$

Proof. The argument used in the proof of Proposition 4.3.1 holds verbatim. □

More importantly, the Petersson inner product turns $\mathcal{L}_k(\Gamma,\chi)$ into a complex Hilbert space:

Proposition 5.3.2. *$\mathcal{L}_k(\Gamma,\chi)$ is a complex Hilbert space with respect to the Petersson inner product.*

Proof. Let $f,g \in \mathcal{L}_k(\Gamma,\chi)$. Linearity of the integral immediately implies that the Petersson inner product is linear on $\mathcal{L}_k(\Gamma,\chi)$. It is also positive definite since

$$\langle f,f\rangle = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{f(z)}\,d\mu = \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z)|^2\,d\mu \ge 0,$$

with equality if and only if f is identically zero. To see that it is conjugate symmetric, observe

$$
\begin{aligned}
\overline{\langle g, f\rangle} &= \overline{\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} g(z)\overline{f(z)}\,d\mu} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \overline{g(z)}f(z)\,\overline{d\mu} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \overline{g(z)}f(z)\,d\mu && d\mu = \frac{dx\,dy}{y^2} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)}\,d\mu \\
&= \langle f, g\rangle.
\end{aligned}
$$

So the Petersson inner product is a Hermitian inner product on $\mathcal{L}_k(\Gamma,\chi)$. We now show that $\mathcal{L}_k(\Gamma,\chi)$ is complete. Let $(f_n)_{n\ge1}$ be a Cauchy sequence in $\mathcal{L}_k(\Gamma,\chi)$. Then $||f_n - f_m|| \to 0$ as $n, m \to \infty$. But

$$
||f_n - f_m|| = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f_n(z) - f_m(z)|^2\,d\mu\right)^{\frac{1}{2}},
$$

and this integral tends to zero if and only if $|f_n(z) - f_m(z)| \to 0$ as $n, m \to \infty$. Therefore $\lim_{n\to\infty} f_n(z)$ exists and we define the limiting function f by $f(z) = \lim_{n\to\infty} f_n(z)$. We claim that f is automorphic. Indeed, as the f_n are automorphic, we have

$$
f(\gamma z) = \lim_{n\to\infty} f_n(\gamma z) = \lim_{n\to\infty} \chi(\gamma)\varepsilon(\gamma,z)^k f_n(z) = \chi(\gamma)\varepsilon(\gamma,z)^k \lim_{n\to\infty} f_n(z) = \chi(\gamma)\varepsilon(\gamma,z)^k f(z),
$$

for any $\gamma \in \Gamma$. Also, $||f|| < \infty$. To see this, since $(f_n)_{n\ge1}$ is Cauchy we know $(||f_n||)_{n\ge1}$ converges. In particular, $\lim_{n\to\infty} ||f_n|| < \infty$. But

$$
\lim_{n\to\infty} ||f_n|| = \lim_{n\to\infty}\left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f_n(z)|^2\,d\mu\right)^{\frac{1}{2}} = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \left|\lim_{n\to\infty} f_n(z)\right|^2\,d\mu\right)^{\frac{1}{2}} = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z)|^2\,d\mu\right)^{\frac{1}{2}} = ||f||,
$$

where the second equality holds by the dominated convergence theorem. Hence $||f|| < \infty$ as desired and so $f \in \mathcal{L}_k(\Gamma,\chi)$. We now show that $f_n \to f$ in the L^2-norm. Indeed,

$$
||f(z) - f_n(z)|| = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z) - f_n(z)|^2\,d\mu\right)^{\frac{1}{2}},
$$

and it follows that $||f(z) - f_n(z)|| \to 0$ as $n \to \infty$ so that the Cauchy sequence $(f_n)_{n\ge1}$ converges. □

We will need two more subspaces. Let $\mathcal{B}_k(\Gamma,\chi)$ be the subspace of $\mathcal{A}_k(\Gamma,\chi)$ such that f is smooth and bounded and let $\mathcal{D}_k(\Gamma,\chi)$ be the subspace of $\mathcal{A}_k(\Gamma,\chi)$ such that f and $\Delta_k f$ are smooth and bounded. If χ is the trivial character or if $k = 0$, we will suppress the dependencies accordingly. Since boundedness on $\mathbb{H}$ implies square-integrability over $\mathcal{F}_\Gamma$, we have the following chain of inclusions:

$$
\mathcal{D}_k(\Gamma,\chi) \subseteq \mathcal{B}_k(\Gamma,\chi) \subseteq \mathcal{L}_k(\Gamma,\chi) \subseteq \mathcal{A}_k(\Gamma,\chi).
$$

Moreover, $\mathcal{D}_k(\Gamma,\chi)$ is almost all of $\mathcal{L}_k(\Gamma,\chi)$ as the following proposition shows:

Proposition 5.3.3. *$\mathcal{D}_k(\Gamma,\chi)$ is dense in $\mathcal{L}_k(\Gamma,\chi)$.*

Proof. Note that $\mathcal{D}_k(\Gamma,\chi)$ is an algebra of functions that vanish at infinity. We will show that $\mathcal{D}_k(\Gamma,\chi)$ is nowhere vanishing, separates points, and self-adjoint. For nowhere vanishing fix a $z \in \mathbb{H}$. Let φ_z be a bump function defined on some sufficiently small neighborhood U_z of z. Then

$$\Phi(v) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \overline{\chi}(\gamma)\varepsilon(\gamma, v)^{-k}\varphi_z(\gamma v),$$

belongs to $\mathcal{D}_k(\Gamma,\chi)$ and is nonzero at z (the automorphy follows exactly as in the case of Eisenstein series). We now show $\mathcal{D}_k(\Gamma,\chi)$ separates points. To see this, consider two distinct points $z, w \in \mathbb{H}$. Let $U_{z,w}$ be a small neighborhood of z not containing w. Then $\Phi_z \mid_{U_{z,w}}$ belongs to $\mathcal{D}_k(\Gamma,\chi)$ with $\Phi_z \mid_{U_{z,w}} (z) \neq 0$ and $\varphi_z \mid_{U_{z,w}} (w) = 0$. To see that $\mathcal{D}_k(\Gamma,\chi)$ is self-adjoint, recall that complex conjugation is smooth and commutes with partial derivatives so that if f belongs to $\mathcal{D}_k(\Gamma,\chi)$ then so does $\overline{f}$. Therefore the Stone–Weierstrass theorem for complex functions defined on locally compact Hausdorff spaces (as $\mathbb{H}$ is a locally compact Hausdorff space) implies that $\mathcal{D}_k(\Gamma,\chi)$ is dense in $C_0(\mathbb{H})$ with the supremum norm. Note that $\mathcal{L}_k(\Gamma,\chi) \subseteq C_0(\mathbb{H})$. Now we show $\mathcal{D}_k(\Gamma,\chi)$ is dense in $\mathcal{L}_k(\Gamma,\chi)$. Let $f \in \mathcal{L}_k(\Gamma,\chi)$. By what we have just shown, there exists a sequence $(f_n)_{n\geq 1}$ in $\mathcal{D}_k(\Gamma,\chi)$ converging to f in the supremum norm. But

$$||f - f_n|| = \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} |f(z) - f_n(z)|^2 \, d\mu\right)^{\frac{1}{2}} \leq \left(\frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sup_{z \in \mathcal{F}_\Gamma} |f(z) - f_n(z)|^2 \, d\mu\right)^{\frac{1}{2}},$$

and the last expression tends to zero as $n \to \infty$ because $f_n \to f$ in the supremum norm. □

As $\mathcal{D}_k(\Gamma,\chi) \subseteq \mathcal{B}_k(\Gamma,\chi)$, Proposition 5.3.3 implies that $\mathcal{B}_k(\Gamma,\chi)$ is dense in $\mathcal{L}_k(\Gamma,\chi)$ too. It can be shown that the Laplace operator Δ_k is bounded from below and symmetric on $\mathcal{D}_k(\Gamma,\chi)$ and hence admits a self-adjoint extension to $\mathcal{L}_k(\Gamma,\chi)$ (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Proposition 5.3.4. *On $\mathcal{L}_k(\Gamma,\chi)$, the Laplace operator Δ_k is bounded from below by $\lambda\left(\frac{|k|}{2}\right)$ and self-adjoint.*

In particular, Δ is positive. If we suppose $f \in \mathcal{L}_k(\Gamma,\chi)$ is an eigenfunction for Δ_k with eigenvalue λ, then Proposition 5.3.4 implies λ is real and $\lambda \geq \lambda\left(\frac{|k|}{2}\right)$. Since $\lambda = s(1-s)$ and λ is real, s and $1-s$ are either conjugates or real. In the former case, $s = 1 - \overline{s}$ and we find that

$$\sigma = 1 - \sigma \quad \text{and} \quad t = t.$$

Therefore $s = \frac{1}{2} + it$. In the later case, s is real. It follows that in either case, we may write $\lambda = \frac{1}{4} + r^2$ and $s = \frac{1}{2} + \nu$ for unique r and ν with r real or purely imaginary and ν purely imaginary or real corresponding to the two cases respectively. In particular, we also have $\lambda = \frac{1}{4} - \nu^2$ and $\nu = ir$. We refer to r as the **spectral parameter** of f and ν as the **type** of f. We collect the ways of expressing λ below:

$$\lambda = s(1-s) = \frac{1}{4} + r^2 = \frac{1}{4} - \nu^2.$$

Therefore to specific λ it suffices to specify either s, the spectral parameter r, or the type ν. We will often replace λ with one of these parameters in $\mathcal{A}_{k,\lambda}(\Gamma,\chi)$, $\mathcal{M}_{k,\lambda}(\Gamma,\chi)$, $\mathcal{C}_{k,\lambda}(\Gamma,\chi)$, and $\mathcal{L}_{k,\lambda}(\Gamma,\chi)$.

Remark 5.3.1. *In the case of embedding weight k holomorphic forms into Maass forms, we have*

$$\lambda = \frac{k}{2}\left(1 - \frac{k}{2}\right) = \frac{1}{4} + \left(i\frac{1-k}{2}\right)^2 + \frac{1}{4} - \left(\frac{1-k}{2}\right)^2,$$

so that $r = i\frac{1-k}{2}$ and $\nu = \frac{k-1}{2}$.

We now introduce variations of the Poincaré and Eisenstein series. Let $m \geq 0$, $k \geq 0$, χ be a Dirichlet character with conductor $q \mid N$, $\sigma_{\mathfrak{a}}$ be a scaling matrix for the $\mathfrak{a}$ cusp, and $\psi(y)$ be a smooth function such that $\psi(y) \ll_\varepsilon y^{1+\varepsilon}$ as $y \to 0$ Then the m-th **(automorphic) Poincaré series** $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ of weight k and character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp and with respect to $\psi(y)$ is defined by

$$P_{m,k,\chi,\mathfrak{a}}(z,\psi) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z))e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}.$$

If $k = 0$, χ is the trivial character, or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly. Moreover, if $\psi(y)$ is a bump function, we say that $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is **incomplete**. We claim that $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is well-defined. This is easy to see as we have already showed $\overline{\chi}(\gamma)$, $\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}$, $\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)$, and $e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}$, are all independent of representatives for γ and $\sigma_{\mathfrak{a}}$ when discussing the automorphic Poincaré series. So $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is well-defined. We claim $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is also locally absolutely uniformly convergent for $z \in \mathbb{H}$. To see this, we require a technical lemma:

Lemma 5.3.1. *For any compact subset K of $\mathbb{H}$, there are finitely many pairs $(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}$, with $c \neq 0$, for which*

$$\frac{\mathrm{Im}(z)}{|cz+d|^2} > 1,$$

for all $z \in K$.

Proof. Let $\beta = \sup_{z \in K} |z|$. As $|cz+d| \geq |cz| > 0$ and $\mathrm{Im}(z) < |z|$, we have

$$\frac{\mathrm{Im}(z)}{|cz+d|^2} \leq \frac{1}{|c^2 z|} \leq \frac{1}{|c|^2\beta}.$$

So if $\frac{\mathrm{Im}(z)}{|cz+d|^2} > 1$, then $\frac{1}{|c|^2\beta} > 1$ which is to say $|c| < \frac{1}{\sqrt{\beta}}$ and therefore $|c|$ is bounded. On the other hand, $|cz+d| \geq |d| \geq 0$. Excluding the finitely many terms $(c,0)$, we may assume $|d| > 0$. In this case, similarly

$$\frac{\mathrm{Im}(z)}{|cz+d|^2} \leq \left|\frac{z}{d^2}\right| \leq \frac{\beta}{|d|^2}.$$

So if $\frac{\mathrm{Im}(z)}{|cz+d|^2} > 1$, then $\frac{\beta}{|d|^2} > 1$ which is to say $|d| < \sqrt{\beta}$. So $|d|$ is also bounded. Since both $|c|$ and $|d|$ are bounded, the claim follows. $\square$

Now we are ready to show that $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is locally absolutely uniformly convergent for $z \in \mathbb{H}$. Let K be a compact subset of $\mathbb{H}$. Then it suffices to show $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is absolutely uniformly convergent on K. The bound $|e^{2\pi i m\sigma_{\mathfrak{a}}^{-1}\gamma z}| = e^{-2\pi m \mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)} < 1$ and the Bruhat decomposition applied to $\sigma_{\mathfrak{a}}^{-1}\Gamma$ together give

$$P_{m,k,\chi,\mathfrak{a}}(z,\psi) \ll \psi(\mathrm{Im}(z)) + \sum_{(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}} \psi\left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right).$$

It now further suffices to show that the latter series above is absolutely uniformly convergent on K. By Lemma 5.3.1, there are all but finitely many terms in the sum with $\psi\left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right) \ll_\varepsilon \left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right)^{1+\varepsilon}$. But the finitely many other terms are all uniformly bounded on K because $\psi(y)$ is continuous (as it is smooth). Therefore

$$\sum_{(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}} \psi\left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right) \ll \sum_{(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}} \left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right)^{1+\varepsilon} \ll \sum_{(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}} \left(\frac{\mathrm{Im}(z)}{|cz+d|^2}\right)^{1+\varepsilon},$$

and this last series is locally absolutely uniformly convergent for $z \in \mathbb{H}$ by Proposition B.8.1. It follows that $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is too. Actually, we can do better if $\psi(y)$ is a bump function since finitely many terms will be nonzero. Indeed, $\sigma_{\mathfrak{a}}^{-1}\Gamma$ is a Fuchsian group because it is a subset of the modular group. So from $\sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\backslash\Gamma = \Gamma_\infty\backslash\sigma_{\mathfrak{a}}^{-1}\Gamma$ we see that $\{\sigma_{\mathfrak{a}}^{-1}\gamma z : \gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma\}$ is discrete. Since $\mathrm{Im}(z)$ is an open map it takes discrete sets to discrete sets so that $\{\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z) : \gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma\}$ is also discrete. Now $\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z))$ is nonzero if and only if $\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z) \in \mathrm{Supp}(\psi)$ and $\{\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z) : \gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma\} \cap \mathrm{Supp}(\psi)$ is finite as it is a discrete subset of a compact set (since $\psi(y)$ has compact support). Hence finitely many of the terms are nonzero. Moreover, the compact support of $\psi(y)$ then implies that $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is also compactly supported (since the function $\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z))$ is continuous and $\mathbb{C}$ is Hausdorff) and hence bounded on $\mathbb{H}$. As a consequence, $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is L^2-integrable. We collect this work as a theorem:

Theorem 5.3.1. *Let $m \geq 0$, $k \geq 0$, χ be a Dirichlet character with conductor dividing the level, $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$, and $\psi(y)$ be a smooth function such that $\psi(y) \ll_\varepsilon y^{1+\varepsilon}$ as $y \to 0$. The Poincaré series*

$$P_{m,k,\chi,\mathfrak{a}}(z,\psi) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z))e^{2\pi i m \sigma_{\mathfrak{a}}^{-1}\gamma z},$$

is a smooth automorphic function on $\Gamma\backslash\mathbb{H}$. If $\psi(y)$ is a bump function, $P_{m,k,\chi,\mathfrak{a}}(z,\psi)$ is L^2-integrable.

For $m = 0$, we write $E_{k,\chi,\mathfrak{a}}(z,\psi) = P_{0,k,\chi,\mathfrak{a}}(z,\psi)$ and call $E_{k,\chi,\mathfrak{a}}(z,\psi)$ the **(automorphic) Eisenstein series** of weight k and character χ on $\Gamma\backslash\mathbb{H}$ at the $\mathfrak{a}$ cusp and with respect to $\psi(y)$. It is defined by

$$E_{k,\chi,\mathfrak{a}}(z,\psi) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)).$$

If $k = 0$, χ is the trivial character, or $\mathfrak{a} = \infty$, we will drop these dependencies accordingly. Moreover, if $\psi(y)$ is a bump function, we say that $E_{k,\chi,\mathfrak{a}}(z,\psi)$ is **incomplete**. We have already verified the following theorem:

Theorem 5.3.2. *Let $k \geq 0$, χ be a Dirichlet character with conductor dividing the level, $\mathfrak{a}$ be a cusp of $\Gamma\backslash\mathbb{H}$, and $\psi(y)$ be a smooth function such that $\psi(y) \ll_\varepsilon y^{1+\varepsilon}$ as $y \to 0$. The Eisenstein series*

$$E_{k,\chi,\mathfrak{a}}(z,\psi) = \sum_{\gamma \in \Gamma_{\mathfrak{a}}\backslash\Gamma} \overline{\chi}(\gamma)\varepsilon(\sigma_{\mathfrak{a}}^{-1}\gamma, z)^{-k}\psi(\mathrm{Im}(\sigma_{\mathfrak{a}}^{-1}\gamma z)),$$

is a smooth automorphic function on $\Gamma\backslash\mathbb{H}$. If $\psi(y)$ is a bump function, $E_{k,\chi,\mathfrak{a}}(z,\psi)$ is also L^2-integrable.

Unfortunately, the Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,\psi)$ fail to be Maass forms because they are not eigenfunctions for the Laplace operator. This is because compactly supported functions cannot be real-analytic (which as we have already mentioned is implied for any eigenfunction of the Laplace operator). However, the incomplete Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,\psi)$ are L^2-integrable where as the Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ are not. This is the advantage in working with incomplete Eisenstein series. We will compute their inner product against an arbitrary element of $\mathcal{B}_k(\Gamma,\chi)$. Let $f \in \mathcal{B}_k(\Gamma,\chi)$ and consider $E_{k,\chi,\mathfrak{a}}(\cdot,\psi)$. We compute their inner product as follows:

$$
\begin{aligned}
\langle f, E_{k,\chi,\mathfrak{a}}(\cdot,\psi)\rangle &= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} f(z)\overline{E_{k,\chi,\mathfrak{a}}(z,\psi)}\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \chi(\gamma)\overline{\varepsilon(\sigma_\mathfrak{a}^{-1}\gamma,z)^{-k}} f(z)\overline{\psi(\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z))}\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \chi(\gamma)\varepsilon(\sigma_\mathfrak{a}^{-1}\gamma,z)^{k} f(z)\overline{\psi(\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z))}\,d\mu && \frac{\overline{\varepsilon(\sigma_\mathfrak{a}^{-1}\gamma,z)}}{\varepsilon(\sigma_\mathfrak{a}^{-1}\gamma,z)} = 1 \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \left(\frac{\varepsilon(\sigma_\mathfrak{a}^{-1}\gamma,z)}{\varepsilon(\gamma,z)}\right)^{k} f(\gamma z)\overline{\psi(\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z))}\,d\mu && \text{automorphy} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \varepsilon(\sigma_\mathfrak{a},\sigma_\mathfrak{a}^{-1}\gamma z)^{-k} f(\gamma z)\overline{\psi(\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma z))}\,d\mu && \text{cocycle condition} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} \varepsilon(\sigma_\mathfrak{a},\sigma_\mathfrak{a}^{-1}\gamma\sigma_\mathfrak{a} z)^{-k} f(\gamma\sigma_\mathfrak{a} z)\overline{\psi(\mathrm{Im}(\sigma_\mathfrak{a}^{-1}\gamma\sigma_\mathfrak{a} z))}\,d\mu && z\to\sigma_\mathfrak{a} z \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\infty\backslash\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}} \varepsilon(\sigma_\mathfrak{a},\gamma z)^{-k} f(\sigma_\mathfrak{a}\gamma z)\overline{\psi(\mathrm{Im}(\gamma z))}\,d\mu && \gamma\to\sigma_\mathfrak{a}\gamma\sigma_\mathfrak{a}^{-1} \\
&= \frac{1}{V_\Gamma}\int_{\mathcal{F}_{\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}}} \sum_{\gamma\in\Gamma_\infty\backslash\sigma_\mathfrak{a}^{-1}\Gamma\sigma_\mathfrak{a}} (f|_k\sigma_\mathfrak{a})(\gamma z)\overline{\psi(\mathrm{Im}(\gamma z))}\,d\mu \\
&= \frac{1}{V_\Gamma}\int_{\Gamma_\infty\backslash\mathbb{H}} (f|_k\sigma_\mathfrak{a})(z)\overline{\psi(\mathrm{Im}(z))}\,d\mu && \text{unfolding.}
\end{aligned}
$$

Substituting in the Fourier series of f at the $\mathfrak{a}$ cusp, we obtain

$$\frac{1}{V_\Gamma}\int_0^\infty\int_0^1 \left(\sum_{n\in\mathbb{Z}} a_\mathfrak{a}(n,y)e^{2\pi i n x}\right)\overline{\psi(y)}\,\frac{dx\,dy}{y^2}.$$

By Fubini's theorem, we can interchange the sum and the two integrals. Upon making this interchange, the identity Equation (4.1) implies that the inner integral cuts off all of the terms in the sum except the diagonal $n=0$, resulting in

$$\frac{1}{V_\Gamma}\int_0^\infty a_\mathfrak{a}(0,y)\overline{\psi(y)}\,\frac{dy}{y^2}.$$

This latter integral is precisely the constant term in the Fourier series of f at the $\mathfrak{a}$ cusp. It follows that f is orthogonal to $\mathcal{E}_k(\Gamma,\chi)$ if and only if $a_\mathfrak{a}(0,y)=0$ for all cusps $\mathfrak{a}$. To state this property in another way, let $\mathcal{E}_k(\Gamma,\chi)$ and $\mathcal{C}_k(\Gamma,\chi)$ denote the subspaces of $\mathcal{B}_k(\Gamma,\chi)$ generated by such forms respectively. Moreover, let $\mathcal{C}_{k,\nu}(\Gamma,\chi)$ and $\mathcal{A}_{k,\nu}(\Gamma,\chi)$ denote the corresponding subspaces of $\mathcal{C}_k(\Gamma,\chi)$ and $\mathcal{A}_k(\Gamma,\chi)$ whose type is ν. If $k=0$ or χ is the trivial character, we will suppress these dependencies. Then we have shown that

$$\mathcal{B}_k(\Gamma,\chi) = \mathcal{E}_k(\Gamma,\chi)\oplus\mathcal{C}_k(\Gamma,\chi).$$

Moreover, as $\mathcal{B}_k(\Gamma,\chi)$ is dense in $\mathcal{L}_k(\Gamma,\chi)$, we have

$$\mathcal{L}_k(\Gamma,\chi) = \overline{\mathcal{E}_k(\Gamma,\chi)}\oplus\overline{\mathcal{C}_k(\Gamma,\chi)},$$

where the closure is with respect to the topology induced by the L^2-norm. The essential fact is that $\mathcal{C}_k(\Gamma,\chi)$ will turn out to be the space of weight k cusp forms with character χ on $\Gamma\backslash\mathbb{H}$ and the corresponding subspaces $\mathcal{C}_{k,\nu}(\Gamma,\chi)$ are finite dimensional. Thus all cusp forms are L^2-integrable and we can apply a linear theory to $\mathcal{C}_{k,\nu}(\Gamma,\chi)$.

5.4 Spectral Theory of the Laplace Operator

We are now ready to discuss the spectral theory of the Laplace operator Δ_k. What we want to do is to decompose $\mathcal{L}_k(\Gamma,\chi)$ into subspaces invariant under Δ_k such that on each subspace Δ_k has either pure point spectrum, absolutely continuous spectrum, or residual spectrum. Although the proof is beyond the scope of this text, the spectral resolution of the Laplace operator on $\mathcal{C}_k(\Gamma,\chi)$ is as follows (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Theorem 5.4.1. *The Laplace operator Δ_k has pure point spectrum on $\mathcal{C}_k(\Gamma,\chi)$. The corresponding subspaces $\mathcal{C}_{k,\nu}(\Gamma,\chi)$ are finite dimensional and mutually orthogonal. Letting $\{u_j\}_{j\geq 1}$ be an orthonormal basis of cusp forms for $\mathcal{C}_k(\Gamma,\chi)$, every $f\in\mathcal{C}_k(\Gamma,\chi)$ admits a series of the form*

$$f(z)=\sum_{j\geq 1}\langle f,u_j\rangle u_j(z),$$

which is locally absolutely uniformly convergent if $f\in\mathcal{D}_k(\Gamma,\chi)$ and convergent in the L^2-norm otherwise.

We will now discuss the spectrum of the Laplace operator on $\mathcal{E}_k(\Gamma,\chi)$. Essential is the meromorphic continuation of the Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Theorem 5.4.2. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be cusps of $\Gamma\backslash\mathbb{H}$. The Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ admits meromorphic continuation to $\mathbb{C}$, via a Fourier-Whittaker series at the $\mathfrak{b}$ cusp given by*

$$E_{k,\chi,\mathfrak{a}}(\sigma_{\mathfrak{b}}z,s)=\delta_{\mathfrak{a},\mathfrak{b}}y^s+\tau_{\mathfrak{a},\mathfrak{b}}(s)y^{1-s}+\sum_{n\neq 0}\tau_{\mathfrak{a},\mathfrak{b}}(n,s)W_{\mathrm{sgn}(n)\frac{k}{2},s-\frac{1}{2}}(4\pi|n|y)e^{2\pi inx},$$

where $\tau_{\mathfrak{a},\mathfrak{b}}(s)$ and $\tau_{\mathfrak{a},\mathfrak{b}}(n,s)$ are meromorphic functions.

The Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ also satisfy a functional equation. To state it we need some notation. Fix an ordering of the cusps $\mathfrak{a}$ of $\Gamma\backslash\mathbb{H}$ and define

$$\mathcal{E}(z,s)=(E_{k,\chi,\mathfrak{a}}(z,s))^t_{\mathfrak{a}}\quad\text{and}\quad\Phi(s)=(\tau_{\mathfrak{a},\mathfrak{b}}(s))_{\mathfrak{a},\mathfrak{b}}.$$

In other words, $\mathcal{E}(z,s)$ is the column vector of the Eisenstein series and $\Phi(s)$ is the square matrix of meromorphic functions $\tau_{\mathfrak{a},\mathfrak{b}}(s)$ described in Theorem 5.4.2. Then we have the following (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Theorem 5.4.3. *The Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ of weight k and character χ on $\Gamma\backslash\mathbb{H}$ satisfy the functional equation*

$$\mathcal{E}(z,s)=\Phi(s)\mathcal{E}(z,1-s).$$

The matrix $\Phi(s)$ is symmetric and satisfies the functional equation

$$\Phi(s)\Phi(1-s)=I.$$

Moreover, it is unitary on the line $\sigma=\frac{1}{2}$ and hermitian if s is real.

As $\Phi(s)$ is symmetric by Theorem 5.4.3, if $\mathfrak{a}=\infty$ or $\mathfrak{b}=\infty$, we will suppress these dependencies for $\tau_{\mathfrak{a},\mathfrak{b}}$. Understanding the poles of $\tau_{\mathfrak{a},\mathfrak{b}}$ are also important for understanding the poles of the Eisenstein series $E_{k,\chi,\mathfrak{a}}(z,s)$ (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Theorem 5.4.4. *The functions $\tau_{\mathfrak{a},\mathfrak{b}}(s)$ are meromorphic for $\sigma \geq \frac{1}{2}$ with a finite number of simple poles in the segment $(\frac{1}{2}, 1]$. A pole of $\tau_{\mathfrak{a},\mathfrak{b}}(s)$ is also a pole of $\tau_{\mathfrak{a},\mathfrak{a}}(s)$. Moreover, the poles of $E_{k,\chi,\mathfrak{a}}(z,s)$ are among the poles of $\tau_{\mathfrak{a},\mathfrak{a}}(s)$, $E_{k,\chi,\mathfrak{a}}(z,s)$ has no poles on the line $\sigma = \frac{1}{2}$, and the residues of $E_{k,\chi,\mathfrak{a}}(z,s)$ are Maass forms in $\mathcal{E}_k(\Gamma,\chi)$.*

To begin decomposing the space $\mathcal{E}_k(\Gamma,\chi)$, consider the subspace $C_0^\infty(\mathbb{R}_{>0})$ of $\mathcal{L}^2(\mathbb{R}_{>0})$ with the normalized standard inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^\infty f(r)\overline{g(r)}\,dr,$$

for any $f, g \in C_0^\infty(\mathbb{R}_{>0})$. For each cusps $\mathfrak{a}$ of $\Gamma\backslash\mathbb{H}$ we associate the **Eisenstein transform** $E_{k,\chi,\mathfrak{a}} : C_0^\infty(\mathbb{R}_{>0}) \to \mathcal{A}_k(\Gamma,\chi)$ defined by

$$(E_{k,\chi,\mathfrak{a}}f)(z) = \frac{1}{4\pi}\int_0^\infty f(r)E_{k,\chi,\mathfrak{a}}\left(z, \frac{1}{2}+ir\right)\,dr.$$

Clearly $E_{k,\chi,\mathfrak{a}}f$ is automorphic because $E_{k,\chi,\mathfrak{a}}(z,s)$ is. It is not too hard to show the following (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Proposition 5.4.1. *If $f \in C_0^\infty(\mathbb{R}_{>0})$, then $E_{\mathfrak{a}}f$ is L^2-integrable over $\mathcal{F}_\Gamma$. That is, $E_{k,\chi,\mathfrak{a}}$ maps $C_0^\infty(\mathbb{R}_{>0})$ into $\mathcal{L}_k(\Gamma,\chi)$. Moreover,*

$$\langle E_{k,\chi,\mathfrak{a}}f, E_{k,\chi,\mathfrak{b}}g\rangle = \delta_{\mathfrak{a},\mathfrak{b}}\langle f, g\rangle,$$

for any $f, g \in C_0^\infty(\mathbb{R}_{>0})$ and any two cusps $\mathfrak{a}$ and $\mathfrak{b}$.

We let $\mathcal{E}_{k,\mathfrak{a}}(\Gamma,\chi)$ denote the image of the Eisenstein transform $E_{k,\chi,\mathfrak{a}}$. We call $\mathcal{E}_{k,\mathfrak{a}}(\Gamma,\chi)$ the **Eisenstein space** of $E_{k,\chi,\mathfrak{a}}(z,s)$. An immediate consequence of Proposition 5.4.1 is that the Eisenstein spaces for distinct cusps are orthogonal. Moreover, since $E_{k,\chi,\mathfrak{a}}\left(z, \frac{1}{2}+ir\right)$ is an eigenfunction for the Laplace operator with eigenvalue $\lambda = \frac{1}{4}+r^2$, and f and $E_{k,\chi,\mathfrak{a}}\left(z, \frac{1}{2}+ir\right)$ are smooth, the Leibniz integral rule implies

$$\Delta E_{k,\chi,\mathfrak{a}} = E_{k,\chi,\mathfrak{a}}M,$$

where $M : C_0^\infty(\mathbb{R}_{>0}) \to C_0^\infty(\mathbb{R}_{>0})$ is the multiplication operator given by

$$(Mf)(r) = \left(\frac{1}{4}+r^2\right)f(r),$$

for all $f \in C_0^\infty(\mathbb{R}_{>0})$. Therefore if $E_{k,\chi,\mathfrak{a}}f$ belongs to $\mathcal{E}_{k,\mathfrak{a}}(\Gamma,\chi)$ then so does $E_{k,\chi,\mathfrak{a}}(Mf)$. But as $f, Mf \in C_0^\infty(\mathbb{R}_{>0})$, this means $\mathcal{E}_{k,\mathfrak{a}}(\Gamma,\chi)$ is invariant under the Laplace operator. While the Eisenstein spaces are invariant, they do not make up all of $\mathcal{E}_k(\Gamma,\chi)$. By Theorem 5.4.4, the residues of the Eisenstein series belong to $\mathcal{E}_k(\Gamma,\chi)$. Let $\mathcal{R}_k(\Gamma,\chi)$ denote the subspace generated by the residues of these Eisenstein series. We call any element of $\mathcal{R}_k(\Gamma,\chi)$ a **(residual) Maass form** (by Theorem 5.4.4 they are Maass forms). Also let $\mathcal{R}_{k,s_j}(\Gamma,\chi)$ denote the subspace generated by those residues taken at $s = s_j$. For both of these subspaces, if χ is the trivial character of if $k = 0$, we will suppress the dependencies accordingly. Since there are finitely many cusps of $\Gamma\backslash\mathbb{H}$, each $\mathcal{R}_{k,s_j}(\Gamma,\chi)$ is finite dimensional. As the number of residues in $(\frac{1}{2}, 1]$ is finite by Theorem 5.4.4, it follows that $\mathcal{R}_k(\Gamma,\chi)$ is finite dimensional too. So $\mathcal{R}_k(\Gamma,\chi)$ decomposes as

$$\mathcal{R}_k(\Gamma,\chi) = \bigoplus_{\frac{1}{2}<s_j\leq 1}\mathcal{R}_{k,s_j}(\Gamma,\chi).$$

This decomposition is orthogonal because the Maass forms belonging to distinct subspaces $\mathcal{R}_{k,s_j}(\Gamma,\chi)$ have distinct eigenvalues and eigenfunctions of self-adjoint operators are orthogonal (recall that Δ_k is self-adjoint by Proposition 5.3.4). Also, each subspace $\mathcal{R}_{k,s_j}(\Gamma,\chi)$ is clearly invariant under the Laplace operator because its elements are Maass forms. The residual forms are particularly simple in the weight zero case (see [Iwa02] for a proof):

Proposition 5.4.2. *There is only one residual form in $\mathcal{R}(\Gamma, \chi)$. It is obtained from the residue at $s = 1$ and it is a constant function.*

We are now ready for the spectral resolution. Although the proof is beyond the scope of this text, the spectral resolution of the Laplace operator on $\mathcal{E}_k(\Gamma, \chi)$ is as follows (see [Iwa02] for a proof in the weight zero case and [DFI02] for notes on the general case):

Theorem 5.4.5. *$\mathcal{E}_k(\Gamma, \chi)$ admits the orthogonal decomposition*

$$\mathcal{E}_k(\Gamma, \chi) = \mathcal{R}_k(\Gamma, \chi) \bigoplus_{\mathfrak{a}} \mathcal{E}_{k,\mathfrak{a}}(\Gamma, \chi),$$

where the direct sum is over the cusps of $\Gamma\backslash\mathbb{H}$. The Laplace operator Δ_k has discrete spectrum on $\mathcal{R}_k(\Gamma, \chi)$ in the segment $[0, \frac{1}{4})$ and has pure continuous spectrum on each Eisenstein space $\mathcal{E}_{k,\mathfrak{a}}(\Gamma, \chi)$ covering the segment $\left[\frac{1}{4}, \infty\right)$ uniformly with multiplicity one. Letting $\{u_j\}_{j\geq 1}$ be an orthonormal basis residual Maass forms for $\mathcal{R}_k(\Gamma, \chi)$, every $f \in \mathcal{E}_{k,\mathfrak{a}}(\Gamma, \chi)$ admits a decomposition of the form

$$f(z) = \sum_{j\geq 1} \langle f, u_j \rangle u_j(z) + \sum_{\mathfrak{a}} \frac{1}{4\pi} \int_{\mathbb{R}} \left\langle f, E_{k,\chi,\mathfrak{a}}\left(\cdot, \frac{1}{2} + \nu\right)\right\rangle E_{k,\chi,\mathfrak{a}}\left(z, \frac{1}{2} + ir\right)\, dr.$$

The series and integrals are locally absolutely uniformly convergent if $f \in \mathcal{D}_k(\Gamma, \chi)$ and convergent in the L^2-norm otherwise.

Combining Theorems 5.4.1 and 5.4.5 gives the full spectral resolution of $\mathcal{L}_k(\Gamma, \chi)$:

Theorem 5.4.6. *$\mathcal{B}_k(\Gamma, \chi)$ admits the orthogonal decomposition*

$$\mathcal{B}_k(\Gamma, \chi) = \mathcal{C}_k(\Gamma, \chi) \oplus \mathcal{R}_k(\Gamma, \chi) \bigoplus_{\mathfrak{a}} \mathcal{E}_{k,\mathfrak{a}}(\Gamma, \chi),$$

where the sum is over all cusps of $\Gamma\backslash\mathbb{H}$. The Laplace operator has pure point spectrum on $\mathcal{C}_k(\Gamma, \chi)$, discrete spectrum on $\mathcal{R}_k(\Gamma, \chi)$, and absolutely continuous spectrum on $\mathcal{E}_k(\Gamma, \chi)$. Letting $\{u_j\}_{j\geq 1}$ be an orthonormal basis of Maass forms for $\mathcal{C}_k(\Gamma, \chi) \oplus \mathcal{R}_k(\Gamma, \chi)$, any $f \in \mathcal{L}_k(\Gamma, \chi)$ has a series of the form

$$f(z) = \sum_{j\geq 1} \langle f, u_j \rangle u_j(z) + \sum_{\mathfrak{a}} \frac{1}{4\pi} \int_{\mathbb{R}} \left\langle f, E_{k,\chi,\mathfrak{a}}\left(\cdot, \frac{1}{2} + \nu\right)\right\rangle E_{k,\chi,\mathfrak{a}}\left(z, \frac{1}{2} + ir\right)\, dr,$$

which is locally absolutely uniformly convergent if $f \in \mathcal{D}_k(\Gamma, \chi)$ and convergent in the L^2-norm otherwise. Moreover,

$$\mathcal{L}_k(\Gamma, \chi) = \overline{\mathcal{C}_k(\Gamma, \chi)} \oplus \overline{\mathcal{R}_k(\Gamma, \chi)} \bigoplus_{\mathfrak{a}} \overline{\mathcal{E}_{k,\mathfrak{a}}(\Gamma, \chi)},$$

where the closure is with respect to the topology induced by the L^2-norm.

Proof. Combine Theorems 5.4.1 and 5.4.5 and use the fact that $\mathcal{B}_k(\Gamma, \chi) = \mathcal{E}_k(\Gamma, \chi) \oplus \mathcal{C}_k(\Gamma, \chi)$ for the first statement. The last statement holds because $\mathcal{B}_k(\Gamma, \chi)$ is dense in $\mathcal{L}_k(\Gamma, \chi)$. $\square$

5.5 Double Coset Operators

We can extend the theory of double coset operators to Maass form just as we did for holomorphic forms. For any two congruence subgroups Γ_1 and Γ_2 (not necessarily of the same level) and any $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, we define the **double coset operator** $[\Gamma_1\alpha\Gamma_2]_k$ to be the linear operator on $\mathcal{C}_{k,\nu}(\Gamma_1)$ given by

$$(f[\Gamma_1\alpha\Gamma_2]_k)(z) = \sum_j (f|_k\beta_j)(z) = \sum_j \det(\beta_j)^{-1}\varepsilon(\beta_j, z)^{-k} f(\beta_j z).$$

As was the case for holomorphic forms, Proposition 4.4.1 implies that this sum is finite. It remains to check that $f[\Gamma_1\alpha\Gamma_2]_k$ is well-defined. Indeed, if β_j and β_j' belong to the same orbit, then $\beta_j'\beta_j^{-1} \in \Gamma_1$. But then as $f \in \mathcal{C}_{k,\nu}(\Gamma_1)$, is it invariant under the $|_k\beta_j'\beta_j^{-1}$ operator so that

$$(f|_k\beta_j)(z) = ((f|_k\beta_j'\beta_j^{-1})|_k\beta_j)(z) = (f|_k\beta_j')(z),$$

and therefore the $[\Gamma_1\alpha\Gamma_2]_k$ operator is well-defined. There is also an analogous statement about the double coset operators for Maass forms:

Proposition 5.5.1. *For any two congruence subgroups Γ_1 and Γ_2, $[\Gamma_1\alpha\Gamma_2]_k$ maps $\mathcal{C}_{k,\nu}(\Gamma_1)$ into $\mathcal{C}_{k,\nu}(\Gamma_2)$.*

Proof. Arguing as in the proof of Proposition 4.4.2 with smoothness replacing holomorphy, automorphy replacing modularity, and the analogous growth condition for Maass forms, the only piece left to verify is that $f[\Gamma_1\alpha\Gamma_2]_k$ is an eigenfunction for Δ with eigenvalue λ if f is. This is easy since the invariance of Δ implies

$$\Delta(f[\Gamma_1\alpha\Gamma_2]_k)(z) = \sum_j \Delta(f|_k\beta_j)(z) = \lambda\sum_j \det(\beta_j)^{-1}\varepsilon(\beta_j, z)^{-k} f(\beta_j z) = \lambda(f[\Gamma_1\alpha\Gamma_2]_k)(z).$$

Thus $f[\Gamma_1\alpha\Gamma_2]_k$ is an eigenfunction for Δ with eigenvalue λ. This completes the proof. □

5.6 Diamond & Hecke Operators

Extending the theory of diamond operators and Hecke operators is also fairly straightforward. To see this, we have already shown that $\Gamma_1(N)$ is normal in $\Gamma_0(N)$ so that

$$\left(f\left[\Gamma_1(N)\alpha\Gamma_1(N)\right]_k\right)(z) = (f|_k\alpha)(z),$$

for any $\alpha = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix}\right) \in \Gamma_0(N)$. Therefore, for any d taken modulo N, we define the **diamond operator** $\langle d\rangle : \mathcal{C}_{k,\nu}(\Gamma_1(N)) \to \mathcal{C}_{k,\nu}(\Gamma_1(N))$ to be the linear operative given by

$$(\langle d\rangle f)(z) = (f|_k\alpha)(z),$$

for any $\alpha = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix}\right) \in \Gamma_0(N)$. As for holomorphic forms, the diamond operators are multiplicative and invertible. They also decompose $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ into eigenspaces. For any Dirichlet character modulo N, let

$$\mathcal{C}_{k,\nu}(N,\chi) = \{f \in \mathcal{C}_{k,\nu}(\Gamma_1(N)) : \langle d\rangle f = \chi(d)f \text{ for all } d \in (\mathbb{Z}/N\mathbb{Z})^*\},$$

be the χ-eigenspace. Also let $\mathcal{C}_{k,\nu}(N,\chi)$ be the corresponding subspace of cusp forms. Then $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ admits a decomposition into these eigenspaces:

Proposition 5.6.1. *We have a direct sum decomposition*

$$\mathcal{C}_{k,\nu}(\Gamma_1(N)) = \bigoplus_{\chi \ (mod\ N)} \mathcal{C}_{k,\nu}(N,\chi).$$

Proof. The argument used in the proof of Proposition 4.5.1 holds verbatim. $\square$

Just as for holomorphic forms, Proposition 5.6.1 shows that the diamond operators sieve Maass forms on $\Gamma_1(N)\backslash\mathbb{H}$ with trivial character in terms of Maass forms on $\Gamma_0(N)\backslash\mathbb{H}$ with nontrivial characters. Precisely, $\mathcal{C}_{k,\nu}(N,\chi) = \mathcal{C}_{k,\nu}(\Gamma_0(N),\chi)$ and $\mathcal{C}_{k,\nu}(N,\chi) = \mathcal{C}_{k,\nu}(\Gamma_0(N),\chi)$. So by Proposition 5.6.1, we have

$$\mathcal{C}_{k,\nu}(\Gamma_1(N)) = \bigoplus_{\chi \ (\mathrm{mod}\ N)} \mathcal{C}_{k,\nu}(\Gamma_0(N),\chi).$$

As for holomorphic forms, this decomposition helps clarify why we consider Maass forms with nontrivial characters. We define the Hecke operators in the same way as for holomorphic forms. For a prime p, we define the p-th **Hecke operator** $T_p : \mathcal{C}_{k,\nu}(\Gamma_1(N)) \to \mathcal{C}_{k,\nu}(\Gamma_1(N))$ to be the linear operator given by

$$(T_pf)(z) = \left(f\left[\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\Gamma_1(N)\right]_k\right)(z).$$

The diamond and Hecke operators commute:

Proposition 5.6.2. *For every $d \in (\mathbb{Z}/N\mathbb{Z})^*$ and prime p, the diamond operators $\langle d\rangle$ and Hecke operators T_p on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ commute:*

$$\langle d\rangle T_p = T_p\langle d\rangle$$

Proof. The argument used in the proof of Proposition 4.5.2 holds verbatim. $\square$

Exactly as for holomorphic forms, Lemma 4.5.1 will give an explicit description of the Hecke operator T_p:

Proposition 5.6.3. *Let $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$. Then the Hecke operator T_p acts on f as follows:*

$$(T_pf)(z) = \begin{cases} \displaystyle\sum_{j \ (mod\ p)} \left(f\Big|_k \begin{pmatrix}1 & j\\ 0 & p\end{pmatrix}\right)(z) + \left(f\Big|_k \begin{pmatrix}m & n\\ N & p\end{pmatrix}\begin{pmatrix}p & 0\\ 0 & 1\end{pmatrix}\right)(z) & \textit{if } p \nmid N,\\ \displaystyle\sum_{j \ (mod\ p)} \left(f\Big|_k \begin{pmatrix}1 & j\\ 0 & p\end{pmatrix}\right)(z) & \textit{if } p \mid N,\end{cases}$$

where m and n are chosen such that $\det\left(\left(\begin{smallmatrix}m & n\\ N & p\end{smallmatrix}\right)\right) = 1$.

Proof. The argument used in the proof of Proposition 4.5.3 holds verbatim. $\square$

We use Proposition 5.6.3 to understand how the Hecke operators act on the Fourier-Whittaker coefficients of Maass forms:

Proposition 5.6.4. *Let $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ have Fourier-Whittaker coefficients $a_n(f)$. Then for all primes p,*

$$(T_pf)(z) = \sum_{n\neq 0}\left(a_{np}(f) + \chi_{N,0}(p)p^{-1}a_{\frac{n}{p}}(\langle p\rangle f)\right)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx},$$

is the Fourier-Whittaker series of T_pf *where it is understood that* $a_{\frac{n}{p}}(f) = 0$ *if* $p \nmid |n|$. *Moreover, if* $f \in \mathcal{C}_{k,\nu}(N,\chi)$, *then* $T_pf \in \mathcal{C}_{k,\nu}(N,\chi)$ *and*

$$(T_pf)(z) = \sum_{n \neq 0} \left(a_{np}(f) + \chi(p)p^{-1}a_{\frac{n}{p}}(f)\right) W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx},$$

where it is understood that $a_{\frac{n}{p}}(f) = 0$ *if* $p \nmid |n|$.

Proof. Argue as in the proof of Proposition 4.5.4. □

As for holomorphic forms, the Hecke operators form a simultaneously commuting family with the diamond operators:

Proposition 5.6.5. *Let* p *and* q *be primes and* $d, e \in (\mathbb{Z}/N\mathbb{Z})^*$. *Then the Hecke operators* T_p *and* T_q *and diamond operators* $\langle d \rangle$ *and* $\langle e \rangle$ *on* $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ *form a simultaneously commuting family:*

$$T_pT_q = T_qT_p, \quad \langle d \rangle T_p = T_p \langle d \rangle, \quad \text{and} \quad \langle d \rangle \langle e \rangle = \langle e \rangle \langle d \rangle.$$

Proof. Argue as in the proof of Proposition 4.5.5. □

We use Proposition 5.6.5 to construct diamond operators $\langle m \rangle$ and Hecke operators T_m for all $m \geq 1$ exactly as for holomorphic forms. Explicitly, the **diamond operator** $\langle m \rangle : \mathcal{C}_{k,\nu}(\Gamma_1(N)) \to \mathcal{C}_{k,\nu}(\Gamma_1(N))$ is defined to be the linear operator given by

$$\langle m \rangle = \begin{cases} \langle m \rangle \text{ with } m \text{ taken modulo } N & \text{if } (m,N) = 1, \\ 0 & \text{if } (m,N) > 1. \end{cases}$$

For the Hecke operators, if $m = p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$ is the prime decomposition of m, then the m-th **Hecke operator** $T_m : \mathcal{C}_{k,\nu}(\Gamma,\chi) \to \mathcal{C}_{k,\nu}(\Gamma,\chi)$ is the linear operator given by

$$T_m = \prod_{1 \leq i \leq k} T_{p_i^{r_i}},$$

where T_{p^r} is defined inductively by

$$T_{p^r} = \begin{cases} T_pT_{p^{r-1}} - p^{-1}\langle p \rangle T_{p^{r-2}} & \text{if } p \nmid N, \\ T_p^r & \text{if } p \mid N, \end{cases}$$

for all $r \geq 2$. Note that when $m = 1$, the product is empty and so T_1 is the identity operator. By Proposition 5.6.5, the Hecke operators T_m are multiplicative but not completely multiplicative in m and they commute with the diamond operators $\langle m \rangle$. Moreover, a more general formula for how the Hecke operators T_m act on the Fourier-Whittaker coefficients can be derived:

Proposition 5.6.6. *Let* $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ *have Fourier-Whittaker coefficients* $a_n(f)$. *Then for* $m \geq 1$ *with* $(m,N) = 1$,

$$(T_mf)(z) = \sum_{n \neq 0} \left(\sum_{d \mid (|n|,m)} d^{-1}a_{\frac{nm}{d^2}}(\langle d \rangle f) \right) W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx},$$

is the Fourier-Whittaker series of T_mf. *Moreover, if* $f \in \mathcal{C}_{k,\nu}(N,\chi)$, *then*

$$(T_mf)(z) = \sum_{n \neq 0} \left(\sum_{d \mid (|n|,m)} \chi(d)d^{-1}a_{\frac{nm}{d^2}}(f) \right) W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx}.$$

Proof. Argue as in the proof of Proposition 4.5.6. $\square$

The diamond and Hecke operators turn out to be normal on the subspace of cusp forms. Just as with holomorphic forms, we can use Lemma 4.5.2 to compute adjoints:

Proposition 5.6.7. *Let Γ be a congruence subgroup and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Set $\alpha' = \det(\alpha)\alpha^{-1}$. Then the following are true:*

(i) If $\alpha^{-1}\Gamma\alpha \subseteq \mathrm{PSL}_2(\mathbb{Z})$, then for all $f \in \mathcal{C}_{k,\nu}(\Gamma,\chi)$ and $g \in \mathcal{C}_{k,\nu}(\alpha^{-1}\Gamma\alpha)$, we have

$$\langle f|_k\alpha, g\rangle_{\alpha^{-1}\Gamma\alpha} = \langle f, g|_k\alpha'\rangle_\Gamma.$$

(ii) For all $f, g \in \mathcal{C}_{k,\nu}(\Gamma,\chi)$, we have

$$\langle f[\Gamma\alpha\Gamma]_k, g\rangle = \langle f, g[\Gamma\alpha'\Gamma]_k\rangle.$$

In particular, if $\alpha^{-1}\Gamma\alpha = \Gamma$ then $|_k\alpha^* = |_k\alpha'$ and $[\Gamma\alpha\Gamma]_k^* = [\Gamma\alpha'\Gamma]_k$ as operators.

Proof. Argue as in the proof of Proposition 4.5.7. $\square$

We can now prove that the diamond and Hecke operators are normal:

Proposition 5.6.8. *On $\mathcal{C}_{k,\nu}(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m are normal for all $m \geq 1$ with $(m, N) = 1$. Moreover, their adjoints are given by*

$$\langle m\rangle^* = \langle \overline{m}\rangle \quad \text{and} \quad T_p^* = \langle \overline{p}\rangle T_p.$$

Proof. The argument used in the proof of Proposition 4.5.8 holds verbatim. $\square$

Just as for holomorphic forms, all of the diamond operators on $\mathcal{C}_{k,\nu}(\Gamma_1(1))$ are the identity and therefore $T_p^* = T_p$ for all primes p. So the Hecke operators are self-adjoint (as are the diamond operators since they are the identity). We need one last operator since cusp forms have Fourier-Whittaker coefficients for all $n \neq 0$. Let $X : \mathcal{C}_{k,\nu}(\Gamma_1(N)) \to \mathcal{C}_{k,\nu}(\Gamma_1(N))$ be the linear operator defined by

$$(Xf)(z) = f(-\overline{z}).$$

As $-\overline{z} = -x + iy$, X acts as reflection with respect to x. Then define the parity **Hecke operator** T_{-1} to be the linear operator on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ given by

$$T_{-1} = X \prod_{\substack{-k<\ell<k \\ \ell \equiv k \pmod 2}} L_k.$$

We will also set

$$\delta(\nu, k) = \frac{\Gamma\left(\nu + \frac{1-k}{2}\right)}{\Gamma\left(\nu + \frac{1+k}{2}\right)}.$$

Notice that $\delta(\nu, 0) = 1$. The parity Hecke operator acts as an involution on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ and more as the following proposition shows (see [DFI02] for a proof):

Proposition 5.6.9. *T_{-1} is an involution on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$. In particular, T_{-1} is an involution on $\mathcal{C}_{k,\nu}(N,\chi)$ as well. If $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ has Fourier-Whittaker coefficients $a_n(f)$, then*

$$(T_{-1}f)(z) = \sum_{n\geq 1} a_n(f)\delta(\nu,k)^{-1}W_{-\frac{k}{2},\nu}(4\pi|n|y)e^{-2\pi inx} + a_{-n}(f)\delta(\nu,k)W_{\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx},$$

is the Fourier-Whittaker series of $T_{-1}f$. Moreover, T_{-1} commutes with the diamond operators $\langle m\rangle$ and Hecke operators T_m for all $m \geq 1$, is normal, and its adjoint is given by

$$T_{-1}^{*} = -T_{-1}.$$

Let $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$. As T_{-1} is an involution, the only possible eigenvalues are ± 1. Accordingly, we say that $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ is **even** if $T_{-1}f = f$ and is **odd** if $T_{-1}f = -f$. Then by Proposition 5.6.9,

$$a_{-n}(f) = \pm a_n(f)\delta(\nu,k)^{-1} = \pm a_n(f)\frac{\Gamma\left(\nu+\frac{1+k}{2}\right)}{\Gamma\left(\nu+\frac{1-k}{2}\right)},$$

for all $n \geq 1$ and with $\pm$ according to if f is even or odd. Thus the Fourier-Whittaker series of f takes the form

$$f(z) = \sum_{n\geq 1} a_n(f)\left(\delta(\nu,k)^{-1}W_{-\frac{k}{2},\nu}(4\pi|n|y)e^{-2\pi inx} \pm W_{\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx}\right),$$

with $\pm$ according to if f is even or odd. If the weight is zero, the Fourier-Whittaker series drastically simplifies via, Theorem B.7.1, the identity $\delta(\nu,0) = 1$, and the exponential identities for sine and cosine, so that we obtain

$$f(z) = a^{+}y^{\frac{1}{2}+\nu} + a^{-}y^{\frac{1}{2}-\nu} + \sum_{n\geq 1} a_n(f)\sqrt{4|n|y}K_\nu(2\pi|n|y)\,\mathrm{SC}(2\pi nx),$$

where $\mathrm{SC}(x) = \cos(x)$ if f is even and $\mathrm{SC}(x) = i\sin(x)$ if f is odd. The benefit of working with even and odd forms is that it suffices to determine non-constant Fourier-Whittaker coefficients for $n \geq 1$ instead of $n \neq 0$. Now suppose f is a non-constant cusp form. Let the eigenvalue of T_m for f be $\lambda_f(m)$. We say that the $\lambda_f(m)$ are the **Hecke eigenvalues** of f. Just as for holomorphic forms, if f is a Maass form on $\Gamma_1(N)\backslash\mathbb{H}$ that is a simultaneous eigenfunction for all diamond operators $\langle m\rangle$ and Hecke operators T_m with $(m,N) = 1$, we call f an **eigenform**. If the condition $(m,N) = 1$ can be dropped, so that f is a simultaneous eigenfunction for all diamond and Hecke operators, we say f is a **Hecke-Maass eigenform**. In particular, on $\Gamma_1(1)\backslash\mathbb{H}$ all eigenforms are Hecke-Maass eigenforms. Now let f have Fourier-Whittaker coefficients $a_n(f)$. As for holomorphic forms, if f a Hecke-Maass eigenform Proposition 5.6.6 immediately implies that the first Fourier-Whittaker coefficient of T_mf is $a_m(f)$ and so

$$a_m(f) = \lambda_f(m)a_1(f),$$

for all $m \geq 1$. Therefore we cannot have $a_1(f) = 0$ for this would mean f is constant. So we can normalize f by dividing by $a_1(f)$ which guarantees that this Fourier-Whittaker coefficient is 1. It follows that

$$a_m(f) = \lambda_f(m),$$

for all $m \geq 1$. This normalization is called the **Hecke normalization** of f. The **Petersson normalization** of f is where we normalize so that $\langle f,f\rangle = 1$. In particular, any orthonormal basis of $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ is Petersson normalized. From the spectral theorem we have an analogous corollary as for holomorphic forms:

Theorem 5.6.1. *$\mathcal{C}_{k,\nu}(\Gamma_1(N))$ admits an orthonormal basis of eigenforms.*

Proof. By Theorem 5.4.1, $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ is finite dimensional. The claim then follows from the spectral theorem along with Propositions 5.6.5 and 5.6.8. $\square$

Also, just as in the holomorphic setting, we have **Hecke relations** for Maass forms:

Proposition 5.6.10 (Hecke relations, Maass version). *Let $f \in \mathcal{C}_{k,\nu}(N,\chi)$ be a Hecke-Maass eigenform with Hecke eigenvalues $\lambda_f(m)$. Then the Hecke eigenvalues are multiplicative and satisfy*

$$\lambda_f(n)\lambda_f(m) = \sum_{d \mid (n,m)} \chi(d)d^{-1}\lambda_f\left(\frac{nm}{d^2}\right) \quad \text{and} \quad \lambda_f(nm) = \sum_{d \mid (n,m)} \mu(d)\chi(d)d^{-1}\lambda_f\left(\frac{n}{d}\right)\lambda_f\left(\frac{m}{d}\right),$$

for all $n, m \geq 1$ with $(nm, N) = 1$. Moreover,

$$\lambda_f(p^r) = \lambda_f(p)^r,$$

for all $p \mid N$ and $r \geq 2$.

Proof. The argument used in the proof of the Hecke relations for holomorphic forms holds verbatim. $\square$

As an immediate consequence of the Hecke relations, the Hecke operators satisfy analogous relations:

Corollary 5.6.1. *The Hecke operators are multiplicative and satisfy*

$$T_nT_m = \sum_{d \mid (n,m)} \chi(d)d^{-1}T_{\frac{nm}{d^2}} \quad \text{and} \quad T_{nm} = \sum_{d \mid (n,m)} \mu(d)\chi(d)d^{-1}T_{\frac{n}{d}}T_{\frac{m}{d}},$$

for all $n, m \geq 1$ with $(nm, N) = 1$.

Proof. The argument used in the proof of Corollary 4.5.1 holds verbatim. $\square$

Just as for holomorphic forms, the identities in Corollary 5.6.1 can also be established directly and the first identity can be used to show that the Hecke operators commute.

5.7 Atkin-Lehner Theory

There is also an Atkin-Lehner theory for Maass form. As with holomorphic forms, we will only deal with congruence subgroups of the form $\Gamma_1(N)$ or $\Gamma_0(N)$ and cusp forms on the corresponding modular curves. The trivial way to lift Maass forms from a smaller level to a larger level is via the natural inclusion $\mathcal{C}_{k,\nu}(\Gamma_1(M)) \subseteq \mathcal{C}_{k,\nu}(\Gamma_1(N))$ provided $M \mid N$ which follows from $\Gamma_1(N) \leq \Gamma_1(M)$. Alternatively, for any $d \mid \frac{N}{M}$, let $\alpha_d = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If $f \in \mathcal{C}_{k,\nu}(\Gamma_1(M))$, then

$$(f|_k\alpha_d)(z) = d^{-1}\varepsilon(\alpha_d, z)^{-k}f(\alpha_d z) = d^{-1}f(dz).$$

Similar to holomorphic forms, $|_k\alpha_d$ maps $\mathcal{C}_{k,\nu}(\Gamma_1(M))$ into $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ and more:

Proposition 5.7.1. *Let M and N be positive integers such that $M \mid N$. For any $d \mid \frac{N}{M}$, $|_k\alpha_d$ maps $\mathcal{C}_{k,\nu}(\Gamma_1(M))$ into $\mathcal{C}_{k,\nu}(\Gamma_1(N))$. In particular, $|_k\alpha_d$ takes $\mathcal{C}_{k,\nu}(M,\chi)$ into $\mathcal{C}_{k,\nu}(N,\chi)$.*

Proof. Arguing as in the proof of Proposition 4.6.1 with smoothness replacing holomorphy, automorphy replacing modularity, and the analogous growth condition for Maass forms, the only piece left to verify is that $f[\alpha_d]$ is an eigenfunction for Δ with eigenvalue λ if f is. This is easy since the invariance of Δ implies

$$\Delta(f|_k\alpha_d)(z) = \lambda d^{-1}f(dz) = \lambda(f|_k\alpha_d)(z).$$

Therefore $f|_k\alpha_d$ is an eigenfunction for Δ with eigenvalue λ. This completes the proof. $\square$

We can now define oldforms and newforms. For each divisor d of N, set

$$i_d : \mathcal{C}_{k,\nu}\left(\Gamma_1\left(\frac{N}{d}\right)\right) \times \mathcal{C}_{k,\nu}\left(\Gamma_1\left(\frac{N}{d}\right)\right) \to \mathcal{C}_{k,\nu}(\Gamma_1(N)) \qquad (f,g) \mapsto f + g|_k\alpha_d.$$

This map is well-defined by Proposition 5.7.1. The subspace of **oldforms** of level N is

$$\mathcal{C}^{\text{old}}_{k,\nu}(\Gamma_1(N)) = \bigoplus_{p|N} \text{Im}(i_p),$$

and the subspace of **newforms** of level N is

$$\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N)) = \mathcal{C}^{\text{old}}_{k,\nu}(\Gamma_1(N))^{\perp}.$$

An element of these subspaces is called an **oldform** or **newform** respectively. Note that there are no oldforms of level 1. Just as with holomorphic forms, we need a useful operator. Recall the matrix

$$W_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix},$$

with $\det(W_N) = N$. We define the **Atkin-Lehner operator** ω_N to be the linear operator on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ given by

$$(\omega_N f)(z) = N(f|_kW_N)(z) = \varepsilon(W_N, z)^{-k} f\left(W_N z\right) = \left(\frac{z}{|z|}\right)^{-k} f\left(-\frac{1}{Nz}\right).$$

It is not too difficult to see how ω_N acts on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$:

Proposition 5.7.2. *ω_N maps $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ into itself. In particular, ω_N takes $\mathcal{C}_{k,\nu}(N,\chi)$ into $\mathcal{C}_{k,\nu}(N,\overline{\chi})$. Moreover, ω_N is self-adjoint and*

$$\omega_N^2 f = (-1)^k f.$$

Proof. Arguing as in the proof of Proposition 4.6.2 with smoothness replacing holomorphy, automorphy replacing modularity, and the analogous growth condition for Maass forms, the only piece left to verify is that $\omega_N f$ is an eigenfunction for Δ with eigenvalue λ if f is. This is easy since the invariance of Δ implies

$$\Delta(\omega_N f)(z) = \lambda\left(\frac{z}{|z|}\right)^{-k} f\left(W_N z\right) = \lambda(\omega_N f)(z).$$

Thus $\omega_N f$ is an eigenfunction for Δ with eigenvalue λ. This completes the proof. $\square$

Proposition 5.7.2 shows that ω_N is an involution if k is even and is at most of order 4. As with holomorphic forms, we need to understand how the Atkin-Lehner operator interacts with the diamond and Hecke operators:

Proposition 5.7.3. *On $\mathcal{C}_{k,\nu}(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m satisfy the following adjoint formulas for all $m \ge 1$:*

$$\langle m\rangle^* = \omega_N \langle m\rangle \omega_N^{-1} \quad \text{and} \quad T_m^* = \omega_N T_m \omega_N^{-1}.$$

Proof. The argument used in the proof of Proposition 4.6.3 holds verbatim. □

It turns out that the spaces of oldforms and newforms are invariant under the diamond and Hecke operators (argue as in the proof of Proposition 4.6.4):

Proposition 5.7.4. *On $\mathcal{C}_{k,\nu}(\Gamma_1(N))$, the diamond operators $\langle m\rangle$ and Hecke operators T_m preserve the subspaces of oldforms and newforms for all $m \ge 1$.*

Proof. The argument used in the proof of Proposition 4.6.4 holds verbatim. □

As a corollary, these subspaces admit orthogonal bases of eigenforms:

Corollary 5.7.1. *$\mathcal{C}^{\text{old}}_{k,\nu}(\Gamma_1(N))$ and $\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$ admit orthonormal bases of eigenforms.*

Proof. This follows immediately from Theorem 5.6.1 and Proposition 5.7.4 □

We can remove the condition $(m, N) = 1$ for eigenforms in a basis of $\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$ so that the eigenforms are eigenfunctions for all of the diamond and Hecke operators. As for holomorphic forms, we need a preliminary result (argue as in the proof of Lemma 4.6.1 as given in [DS05]):

Lemma 5.7.1. *If $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ has Fourier-Whittaker coefficients $a_n(f)$ and is such that $a_n(f) = 0$ for all $n \ge 1$ whenever $(n, N) = 1$, then*

$$f = \sum_{p \mid N} p^{-1} f_p|_k \alpha_p,$$

for some $f_p \in \mathcal{C}_{k,\nu}\left(\Gamma_1\left(\frac{N}{p}\right)\right)$.

As was the case for holomorphic forms, we observe from Lemma 5.7.1 that if $f \in \mathcal{C}_{k,\nu}(\Gamma_1(N))$ is such that its positive n-th Fourier-Whittaker coefficients vanish when n is relatively prime to the level, then f must be an oldform. The main theorem about $\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$ can now be proved. We say that f is a **primitive Hecke-Maass eigenform** if it is a nonzero Hecke normalized Hecke-Maass eigenform in $\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$. We can now prove the main result about newforms which is that Hecke-Maass eigenforms exist:

Theorem 5.7.1. *Let $f \in \mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$ be an eigenform. Then the following hold:*

(i) f is a Hecke-Maass eigenform.

(ii) If g is any cusp form with the same Hecke eigenvalues at all primes, then $g = cf$ for some nonzero $c \in \mathbb{C}$.

Moreover, the primitive Hecke-Maass eigenforms in $\mathcal{C}^{\text{new}}_{k,\nu}(\Gamma_1(N))$ form an orthogonal basis and each such eigenform lies in an eigenspace $\mathcal{C}_{k,\nu}(N, \chi)$.

Proof. The argument used in the proof of Theorem 4.6.1 holds verbatim. □

Statement (i) in Theorem 5.7.1 implies that primitive Hecke-Maass eigenforms satisfy the Hecke relations for all $n, m \geq 1$. Statement (ii) is referred to as **multiplicity one** for Maass forms. As is the case for holomorphic forms, $\mathcal{C}_{k,\nu}^{\text{new}}(\Gamma_1(N))$ contains one element per set of eigenvalues for the Hecke operators. As a consequence of multiplicity one, all primitive Hecke-Maass eigenforms are either even or odd:

Proposition 5.7.5. *If $f \in \mathcal{C}_{k,\nu}(N, \chi)$ is a primitive Hecke-Maass eigenform, then f is either even or odd.*

Proof. By Proposition 5.6.9, the parity Hecke operator T_{-1} commutes with all the Hecke operators. Therefore if $f \in \mathcal{C}_{k,\nu}(N, \chi)$ is a primitive Hecke-Maass eigenform with Hecke eigenvalues $\lambda_f(m)$, then $T_{-1}f$ is a Hecke eigenform with the same Hecke eigenvalues. Then multiplicity one gives

$$T_{-1}f = cf,$$

for some nonzero $c \in \mathbb{C}$. But as T_{-1} is an involution by Proposition 5.6.9, we must have $c = \pm 1$. $\square$

We now discuss conjugate cusp forms. For any cusp form $f \in \mathcal{C}_{k,\nu}(N, \chi)$, we define the **conjugate** $\overline{f}$ of f by

$$\overline{f}(z) = \overline{f(-\overline{z})}.$$

Note that if f has Fourier-Whittaker coefficients $a_n(f)$, then $\overline{f}$ has Fourier-Whittaker coefficients $\overline{a_n(f)}$ by the conjugate symmetry of the Whittaker function (see Appendix B.7). It turns out that $\overline{f}$ is indeed a cusp form:

Proposition 5.7.6. *If $f \in \mathcal{C}_{k,\nu}(N, \chi)$, then $\overline{f} \in \mathcal{C}_{k,\nu}(N, \overline{\chi})$. Moreover,*

$$T_m\overline{f} = \overline{T_m f},$$

for all $m \geq 1$ with $(m, N) = 1$. In particular, if f is an eigenform with Hecke eigenvalues $\lambda_f(m)$ then f is too but with Hecke eigenvalues $\overline{\lambda_f(m)}$.

Proof. Argue as in the proof of Proposition 4.6.5. $\square$

Just as with holomorphic forms, Theorem 5.7.1 and Proposition 5.7.6 together imply that the primitive Hecke-Maass eigenforms in $\mathcal{C}_{k,\nu}^{\text{new}}(\Gamma_1(N))$ are conjugate invariant and if $f \in \mathcal{C}_{k,\nu}(N, \chi)$ is such an eigenform then $\overline{f} \in \mathcal{C}_{k,\nu}(N, \overline{\chi})$ is as well. The crucial fact we need is how $\omega_N f$ is related to $\overline{f}$ when f is a primitive Hecke-Maass eigenform:

Proposition 5.7.7. *If $f \in \mathcal{C}_{k,\nu}(N, \chi)$ is a primitive Hecke-Maass eigenform, then*

$$\omega_N f = \omega_N(f)\overline{f},$$

where $\overline{f} \in \mathcal{C}_{k,\nu}(N, \overline{\chi})$ is a primitive Hecke-Maass eigenform and $\omega_N(f) \in \mathbb{C}$ is nonzero with $|\omega_N(f)| = 1$.

Proof. The argument used in the proof of Proposition 4.6.6 holds verbatim. $\square$

5.8 The Ramanujan-Petersson Conjecture

As for the size of the Fourier-Whittaker coefficients of Maass form, much is currently unknown. But there is an analogous conjecture to the one for holomorphic forms. To state it, suppose $f \in \mathcal{C}_{k,\nu}(N,\chi)$ is a primitive Hecke-Maass eigenform with Hecke eigenvalues $\lambda_f(m)$. For each prime p, consider the polynomial

$$1 - \lambda_f(p)p^{-s} + \chi(p)p^{-2s}.$$

We call this the p-th **Hecke polynomial** of f. Let $\alpha_1(p)$ and $\alpha_2(p)$ denote the roots. Then

$$\alpha_1(p) + \alpha_2(p) = \lambda_f(p) \quad \text{and} \quad \alpha_1(p)\alpha_2(p) = \chi(p).$$

The **Ramanujan-Petersson conjecture** for Maass forms is following statement:

Conjecture 5.8.1 (Ramanujan-Petersson conjecture, Maass version). *Suppose $f \in \mathcal{C}_{k,\nu}(N,\chi)$ is a primitive Hecke-Maass eigenform with Hecke eigenvalues $\lambda_f(m)$. Let $\alpha_1(p)$ and $\alpha_2(p)$ be the roots of the p-th Hecke polynomial. Then for all primes p,*

$$|\lambda_f(p)| \le 2p^{-\frac{1}{2}}.$$

Moreover, if $p \nmid N$, then

$$|\alpha_1(p)| = |\alpha_2(p)| = 1.$$

The Ramanujan-Petersson conjecture has not yet been proved, but there has been partial progress toward the conjecture. The current best bound is $|\lambda_f(p)| \le 2p^{\frac{7}{64}-\frac{1}{2}}$ due to Kim and Sarnak (see [KRS03] for the proof). Under the Ramanujan-Petersson conjecture, the Hecke relations give the improved bound $\lambda_f(m) \ll \sigma_0(m)m^{-\frac{1}{2}} \ll_\varepsilon m^{\varepsilon-\frac{1}{2}}$ (recall Proposition A.3.1). It turns out that the Ramanujan-Petersson conjecture is tightly connected to another conjecture of Selberg about the smallest possible eigenvalue of Maass form on $\Gamma\backslash\mathbb{H}$. Note that the possible eigenvalues are discrete by Theorem 5.4.1 and so there exists a smallest eigenvalue. To state it, recall that if f is a Maass form with eigenvalue λ and spectral parameter r on $\Gamma\backslash\mathbb{H}$, then $\lambda = \frac{1}{4} + r^2$ with either $r \in \mathbb{R}$ or $ir \in [0, \frac{1}{2})$. The **Selberg conjecture** claims that the second case never occurs:

Conjecture 5.8.2 (Selberg conjecture). *If λ is the smallest eigenvalue for Maass forms on $\Gamma\backslash\mathbb{H}$, then*

$$\lambda \ge \frac{1}{4}.$$

Selberg was able to achieve a remarkable lower bound using the analytic continuation of a certain Dirichlet series and the Weil bound for Kloosterman sums (see [Iwa02] for a proof):

Theorem 5.8.1. *If λ is the smallest eigenvalue for Maass forms on $\Gamma\backslash\mathbb{H}$, then*

$$\lambda \ge \frac{3}{16}.$$

In the language of automorphic representations, these two conjectures are a consequence of a much larger conjecture (see [BB13] for details).

5.9 Twists of Maass Forms

We can also twist Maass forms by Dirichlet characters. Let $f \in \mathcal{C}_{k,\nu}(N,\chi)$ have Fourier-Whittaker series

$$f(z) = a^+(f)y^{\frac{1}{2}+\nu} + a^-(f)y^{\frac{1}{2}-\nu} + \sum_{n\neq 0} a_n(f)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx}.$$

and let ψ be a primitive Dirichlet character modulo M. We define the **twisted Maass form** $f\otimes\psi$ of f twisted by ψ by the Fourier-Whittaker series

$$(f\otimes\psi)(z) = a^+(f)\psi(0)y^{\frac{1}{2}+\nu} + a^-(f)\psi(0)y^{\frac{1}{2}-\nu} + \sum_{n\neq 0} a_n(f)\psi(n)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx}.$$

In order for $f\otimes\psi$ to be well-defined, we need to prove that it is a Maass form. The following proposition proves this and more when ψ is primitive:

Proposition 5.9.1. *Suppose $f \in \mathcal{C}_{k,\nu}(N,\chi)$ and ψ is a primitive Dirichlet character of conductor q. Then $f\otimes\psi \in \mathcal{C}_{k,\nu}(Nq^2,\chi\psi^2)$.*

Proof. Arguing as in the proof of Proposition 4.8.2 with smoothness replacing holomorphy, automorphy replacing modularity, and the analogous growth condition for Maass forms, the only piece left to verify is that $f\otimes\psi$ is an eigenfunction for Δ with eigenvalue λ if f is in the case ψ is primitive. This is easy since the invariance of Δ implies

$$\Delta(f\otimes\chi)(z) = \frac{1}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)\Delta(f)\left(z+\frac{r}{q}\right) = \frac{\lambda}{\tau(\overline{\psi})}\sum_{r \pmod q}\overline{\psi}(r)f\left(z+\frac{r}{q}\right) = \lambda(f\otimes\chi)(z).$$ □

The generalization of Proposition 5.9.1 to all characters is slightly more involved. Define operators U_p and V_p on $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ to be the linear operators given by

$$(U_pf)(z) = \sum_{n\neq 0} a_{np}(f)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx},$$

and

$$(V_pf)(z) = \sum_{n\neq 0} a_n(f)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|py)e^{2\pi inpx},$$

if f has Fourier-Whittaker series

$$f(z) = \sum_{n\neq 0} a_n(f)W_{\mathrm{sgn}(n)\frac{k}{2},\nu}(4\pi|n|y)e^{2\pi inx}.$$

We will show that both U_p and V_p map $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ into $\mathcal{C}_{k,\nu}(\Gamma_1(Np))$ and more:

Lemma 5.9.1. *For any prime p, U_p and V_p map $\mathcal{C}_{k,\nu}(\Gamma_1(N))$ into $\mathcal{C}_{k,\nu}(\Gamma_1(Np))$. In particular, U_p and V_p map $\mathcal{C}_{k,\nu}(N,\chi)$ into $\mathcal{C}_{k,\nu}(Np,\chi\chi_{p,0})$.*

Proof. The argument used in the proof of Lemma 4.8.1 holds verbatim. □

We can now generalize Proposition 5.9.1 to all characters:

Proposition 5.9.2. *Suppose $f \in \mathcal{C}_{k,\nu}(N, \chi)$ and ψ is a Dirichlet character modulo M. Then $f \otimes \psi \in \mathcal{C}_{k,\nu}(NM^2, \chi\psi^2)$.*

Proof. Arguing as in the proof of Proposition 4.8.2, it remains to show that $f \otimes \psi_{p,0}$ is an eigenfunction with eigenvalue λ if f is. As $U_p = T_p$ is the p-th Hecke operator on $\mathcal{C}_{k,\nu}(\Gamma_1(Np))$, U_p commutes with Δ. It is also clear that V_p commutes with Δ. These facts together with

$$f \otimes \psi_{p,0} = f - V_pU_pf,$$

show that $f \otimes \psi_{p,0}$ is an eigenfunction with eigenvalue λ if f is. $\square$

In particular, Proposition 5.9.2 shows that $f \otimes \psi$ is well-defined for any Dirichlet character ψ.

Chapter 6

Trace Formulas

There are various types of formulas that relate the Fourier coefficients of automorphic forms. One of the most important such formulas is the Petersson trace formulas.

6.1 The Petersson Trace Formula

From Theorem 4.6.1, $\mathcal{S}_k(N,\chi)$ admits an orthonormal basis of Hecke eigenforms. In particular, $\mathcal{S}_k(N,\chi)$ admits a merely orthogonal basis. Denote this basis by $\{u_j\}_{1\le j\le r}$ where r is the dimension of $\mathcal{S}_k(N,\chi)$. Each of these forms admits a Fourier series at the $\mathfrak{a}$ cusp given by

$$(u_j|\sigma_{\mathfrak{a}})(z) = \sum_{n\ge 1} a_{j,\mathfrak{a}}(n)e^{2\pi i n z}.$$

The Petersson trace formula is an equation relating the Fourier coefficients $a_{j,\mathfrak{a}}(n)$ and $a_{j,\mathfrak{b}}(n)$ of the basis $\{u_j\}_{1\le j\le r}$ for two cusps $\mathfrak{a}$ and $\mathfrak{b}$ of $\Gamma_0(N)\backslash\mathbb{H}$ to a sum of J-Bessel functions and Salié sums. To prove the Petersson trace formula we compute the inner product of two Poincaré series $P_{n,k,\chi,\mathfrak{a}}(z)$ and $P_{m,k,\chi,\mathfrak{b}}(z)$ in two different ways. One way is geometric in nature while the other is spectral. Since Theorem 4.3.2 says that $\langle P_{n,k,\chi,\mathfrak{a}}, P_{m,k,\chi,\mathfrak{b}}\rangle$ extracts the m-th Fourier coefficient of $P_{n,k,\chi,\mathfrak{a}}$ up to a constant, the Petersson trace formula amounts to computing the m-th Fourier coefficient of $P_{n,k,\chi,\mathfrak{a}}$ in two different ways. We will begin with the geometric method first. This is easy as we have already computed the Fourier series of the Poincaré series. Applying Theorem 4.3.2 to the Fourier series in Proposition 4.2.1 gives

$$\langle P_{n,k,\chi,\mathfrak{a}}, P_{m,k,\chi,\mathfrak{b}}\rangle = \frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi m)^{k-1}}\left(\delta_{\mathfrak{a},\mathfrak{b}}\delta_{n,m} + \left(\frac{\sqrt{m}}{\sqrt{n}}\right)^{k-1}\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{nm}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(n,m,c)\right).$$

This is the first half of the Petersson trace formula. To obtain the second half, we use the fact that $\{u_j\}_{1\le j\le r}$ is an orthogonal basis for $\mathcal{S}_k(N,\chi)$ and Theorem 4.3.2 to write

$$P_{n,k,\chi,\mathfrak{a}}(z) = \sum_{1\le j\le r}\frac{\langle P_{n,k,\chi,\mathfrak{a}}, u_j\rangle}{\langle u_j,u_j\rangle}u_j(z) = \sum_{1\le j\le r}\frac{\overline{\langle u_j, P_{n,k,\chi,\mathfrak{a}}\rangle}}{\langle u_j,u_j\rangle}u_j(z) = \frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi n)^{k-1}}\sum_{1\le j\le r}\frac{\overline{a_{j,\mathfrak{a}}(n)}}{\langle u_j,u_j\rangle}u_j(z).$$

This last expression is the spectral decomposition of $P_{n,k,\chi,\mathfrak{a}}(z)$ in terms of the basis $\{u_j\}_{1\le j\le r}$. So if we apply Theorem 4.3.2 to this last expression, we obtain

$$\langle P_{n,k,\chi,\mathfrak{a}}, P_{m,k,\chi,\mathfrak{b}}\rangle = \left(\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi\sqrt{nm})^{k-1}}\right)^2\sum_{1\le j\le r}\frac{\overline{a_{j,\mathfrak{a}}(n)}a_{j,\mathfrak{b}}(m)}{\langle u_j,u_j\rangle},$$

which is the second half of the Petersson trace formula. Equating the first and second halves and canceling the common $\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi m)^{k-1}}$ factor gives

$$\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi n)^{k-1}}\sum_{1\le j\le r}\frac{\overline{a_{j,\mathfrak{a}}(n)}a_{j,\mathfrak{b}}(m)}{\langle u_j,u_j\rangle}=\delta_{\mathfrak{a},\mathfrak{b}}\delta_{n,m}+\left(\frac{\sqrt{m}}{\sqrt{n}}\right)^{k-1}\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{nm}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(n,m,c).$$

Since $\left(\frac{\sqrt{m}}{\sqrt{n}}\right)^{k-1}=1$ when $n=m$, we can factor this term out of the entire right-hand side and cancel it resulting in the **Petersson trace formula** relative to the $\mathfrak{a}$ and $\mathfrak{b}$ cusps:

$$\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi\sqrt{nm})^{k-1}}\sum_{1\le j\le r}\frac{\overline{a_{j,\mathfrak{a}}(n)}a_{j,\mathfrak{b}}(m)}{\langle u_j,u_j\rangle}=\delta_{\mathfrak{a},\mathfrak{b}}\delta_{n,m}+\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{nm}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(n,m,c).$$

We refer to the left-hand side side as the **spectral side** and the right-hand side as the **geometric side**. We collect our work as a theorem:

Theorem 6.1.1 (Petersson trace formula). *Let $\{u_j\}_{1\le j\le r}$ be an orthogonal basis of Hecke eigenforms for $\mathcal{S}_k(N,\chi)$ with Fourier coefficients $a_{j,\mathfrak{a}}(n)$ at the $\mathfrak{a}$ cusp. Then for any positive integers $n,m\ge 1$ and any two cusps $\mathfrak{a}$ and $\mathfrak{b}$, we have*

$$\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi\sqrt{nm})^{k-1}}\sum_{1\le j\le r}\frac{\overline{a_{j,\mathfrak{a}}(n)}a_{j,\mathfrak{b}}(m)}{\langle u_j,u_j\rangle}=\delta_{\mathfrak{a},\mathfrak{b}}\delta_{n,m}+\sum_{c\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{nm}}{c}\right)S_{\chi,\mathfrak{a},\mathfrak{b}}(n,m,c).$$

Note that if we take $\{u_j\}_{1\le j\le r}$ to be an orthonormal basis, equivalently the basis elements are Hecke normalized, then $\langle u_j,u_j\rangle=1$. Regardless, a particularly important case of the Petersson trace formula is when $\mathfrak{a}=\mathfrak{b}=\infty$. For then $\mathcal{C}_{\infty,\infty}=\{c\ge 1 : c\equiv 0 \pmod N\}$, the Salié sum reduces to the usual one, and the Petersson trace formula takes the form

$$\frac{\Gamma(k-1)}{V_{\Gamma_0(N)}(4\pi\sqrt{nm})^{k-1}}\sum_{1\le j\le r}\frac{\overline{a_j(n)}a_j(m)}{\langle u_j,u_j\rangle}=\delta_{n,m}+\sum_{\substack{c\ge 1\\ c\equiv 0 \pmod N}}\frac{2\pi i^{-k}}{c}J_{k-1}\left(\frac{4\pi\sqrt{nm}}{c}\right)S_{\chi}(n,m,c).$$

Part IV

An Introduction to L-functions

Chapter 7

The Theory of L-functions

We start our discussion of L-functions with Dirichlet series. Dirichlet series are essential tools in analytic number theory because they are a way of analytically encoding arithmetic information. If the Dirichlet series possesses special properties we call it an L-function. From the analytic properties of L-functions we can extract number theoretic results. After discussing Dirichlet series we will define L-functions and their associated data. The material following the data of L-functions consists of many important discussions about the analytic properties of L-functions: the approximate functional equation, the Riemann hypothesis and Lindelöf hypothesis, the central value, logarithmic derivatives, zero density, zero-free regions, and explicit formulas.

7.1 Dirichlet Series

A **Dirichlet series** $D(s)$ is a sum of the form

$$D(s) = \sum_{n\geq 1} \frac{a(n)}{n^s},$$

with $a(n) \in \mathbb{C}$. We exclude the case $a(n) = 0$ for all $n \geq 1$ so that $D(s)$ is not identically zero. We would first like to understand where this series converges. It does not take much for $D(s)$ to converge uniformly in a sector:

Theorem 7.1.1. *Suppose $D(s)$ is a Dirichlet series that converges at $s_0 = \sigma_0 + it_0$. Then for any $H > 0$, $D(s)$ converges uniformly in the sector*

$$\{s \in \mathbb{C} : \sigma \geq \sigma_0 \textit{ and } |t - t_0| \leq H(\sigma - \sigma_0)\}.$$

Proof. Set $R(u) = \sum_{n\geq u} \frac{a(n)}{n^{-s_0}}$ so that $a(n) = (R(n) - R(n+1))n^{s_0}$. Then for any two positive integers N and M with $1 \leq M < N$, partial summation (see Appendix B.3) implies

$$\sum_{M\leq n\leq N} \frac{a(n)}{n^s} = R(M)M^{s_0-s} - R(N)N^{s_0-s} - \sum_{M+1\leq n\leq N} R(n)((n-1)^{s_0-s} - n^{s_0-s}). \tag{7.1}$$

We will now express the sum on the right-hand side as an integral. To do this, observe that

$$(n-1)^{s_0-s} - n^{s_0-s} = -(s_0 - s)\int_{n-1}^{n} u^{s_0-s-1}\, du.$$

Therefore

$$\begin{aligned}\sum_{M+1\le n\le N} R(n)((n-1)^{s_0-s}-n^{s_0-s}) &= -(s_0-s)\sum_{M+1\le n\le N} R(n)\int_{n-1}^{n} u^{s_0-s-1}\,du \\ &= -(s_0-s)\sum_{M+1\le n\le N}\int_{n-1}^{n} R(u)u^{s_0-s-1}\,du \\ &= -(s_0-s)\int_{M}^{N} R(u)u^{s_0-s-1}\,du,\end{aligned} \tag{7.2}$$

where the second to last line follows because $R(u)$ is constant on the interval $[u, u+1)$. Combining Equations (7.1) and (7.2) gives

$$\sum_{M\le n\le N}\frac{a(n)}{n^s} = R(M)M^{s_0-s} - R(N)N^{s_0-s} + (s_0-s)\int_{M}^{N} R(u)u^{s_0-s-1}\,du. \tag{7.3}$$

Now there exists an M such that $|R(u)| < \varepsilon$ for all $u \ge M$ because $D(s)$ is convergent at s_0. In particular, $|R(u)u^{s_0-s}| < \varepsilon$ for all $u \ge M$ because $\sigma \ge \sigma_0$. Moreover for s in the prescribed sector,

$$|s-s_0| \le (\sigma-\sigma_0) + |t-t_0| \le (H+1)(\sigma-\sigma_0).$$

These estimates and Equation (7.3) together imply

$$\left|\sum_{M\le n\le N}\frac{a(n)}{n^s}\right| \le 2\varepsilon + \varepsilon|s-s_0|\int_{M}^{N} u^{\sigma_0-\sigma-1}\,du \le 2\varepsilon + \varepsilon(H+1)(\sigma-\sigma_0)\int_{M}^{N} u^{\sigma_0-\sigma-1}\,du.$$

Since the integral is finite, $\sum_{M\le n\le N}\frac{a(n)}{n^s}$ can be made arbitrarily small uniformly for s in the desired sector. The claim now follows by the uniform version of Cauchy's criterion. □

By taking $H \to \infty$ in Theorem 7.1.1 we see that $D(s)$ converges in the region $\sigma > \sigma_0$. Let σ_c be the infimum of all σ for which $D(s)$ converges. We call σ_c the **abscissa of convergence** of $D(s)$. Similarly, let σ_a be the infimum of all σ for which $D(s)$ converges absolutely. Since the terms of $D(s)$ are holomorphic, the convergence is locally absolutely uniform (actually uniform in sectors) for $\sigma > \sigma_a$. It follows that $D(s)$ is holomorphic in the region $\sigma > \sigma_a$. We call σ_a the **abscissa of absolute convergence** of $D(s)$. One should think of σ_c and σ_a as the boundaries of convergence and absolute convergence respectively. Of course, anything can happen at $\sigma = \sigma_c$ and $\sigma = \sigma_a$, but to the right of these lines we have convergence and absolute convergence of $D(s)$ respectively. It turns out that σ_a is never far from σ_c provided σ_c is finite:

Theorem 7.1.2. *If $D(s)$ is a Dirichlet series with finite abscissa of convergence σ_c, then*

$$\sigma_c \le \sigma_a \le \sigma_c + 1.$$

Proof. The first inequality is trivial since absolute convergence implies convergence. For the second inequality, since $D(s)$ converges at $\sigma_c + \varepsilon$, the terms $a(n)n^{-(\sigma_c+\varepsilon)}$ tend to zero as $n \to \infty$. Therefore $a(n) \ll_\varepsilon n^{\sigma_c+\varepsilon}$ where the implicit constant is independent of n. But then $a(n)n^{-(\sigma_c+\varepsilon)} \ll_\varepsilon 1$ which implies $\sum_{n\ge1} a(n)n^{-(\sigma_c+1+2\varepsilon)}$ is absolutely convergent by the comparison test with respect to $\sum_{n\ge1} n^{-(1+\varepsilon)}$. In terms of $D(s)$, this means $\sigma_a \le \sigma_c + 1 + 2\varepsilon$ and taking $\varepsilon \to 0$ gives the second inequality. □

We will now introduce several convergence theorems for Dirichlet series. It will be useful to setup some notation first. If $D(s)$ is a Dirichlet series with coefficients $a(n)$, then for $X > 0$, we set

$$A(X) = \sum_{n \le X} a(n) \quad \text{and} \quad |A|(X) = \sum_{n \le X} |a(n)|.$$

These are the partial sums of the coefficients $a(n)$ and $|a(n)|$ up to X respectively. Our first convergence theorem relates boundedness of $A(X)$ to the value of σ_c:

Proposition 7.1.1. *Suppose $D(s)$ is a Dirichlet series and that $A(X) \ll 1$. Then $\sigma_c \le 0$.*

Proof. Let s be such that $\sigma > 0$. Since $A(X) \ll 1$, $A(X)X^{-s} \to 0$ as $X \to \infty$. Abel's summation formula (see Appendix B.3) then implies

$$D(s) = s\int_1^\infty A(u)u^{-(s+1)}\,du.$$

But because $A(u) \ll 1$, we have

$$s\int_1^\infty A(u)u^{-(s+1)}\,du \ll s\int_1^\infty u^{-(\sigma+1)}\,du = -\frac{s}{\sigma}u^{-\sigma}\Big|_1^\infty = \frac{s}{\sigma},$$

so that the integral exists for $\sigma > 0$. Thus $D(s)$ converges for $\sigma > 0$ and so $\sigma_c \le 0$. □

Our next theorem states that if the coefficients of $D(s)$ are of polynomial growth, we can obtain an upper bound for the abscissa of absolute convergence:

Proposition 7.1.2. *Suppose $D(s)$ is a Dirichlet series whose coefficients satisfy $a(n) \ll_\alpha n^\alpha$ for some real α. Then the abscissa of absolute convergence satisfies $\sigma_a \le 1 + \alpha$.*

Proof. It suffices to show that $D(s)$ is absolutely convergent in the region $\sigma > 1+\alpha$. For s is in this region, the polynomial bound gives

$$D(s) \ll \sum_{n \ge 1} \left|\frac{a(n)}{n^s}\right| \ll_\alpha \sum_{n \ge 1} \frac{1}{n^{\sigma-\alpha}}.$$

The latter series converges by the integral test because $\sigma - \alpha > 1$. Therefore $D(s)$ is absolutely convergent. □

Obtaining polynomial bounds on coefficients of Dirichlet series are, in most cases, not hard to establish. So the assumption in Proposition 7.1.2 is mild. Actually, there is a partial converse to Proposition 7.1.2 which gives an approximate size to $A(X)$:

Proposition 7.1.3. *Suppose $D(s)$ is a Dirichlet series with finite and nonnegative abscissa of absolute convergence σ_a. Then*

$$A(X) \ll_\varepsilon X^{\sigma_a+\varepsilon}.$$

Proof. By Abel's summation formula (see Appendix B.3),

$$\sum_{n \le X} \frac{a(n)}{n^{\sigma_a+\varepsilon}} = A(X)X^{-(\sigma_a+\varepsilon)} + (\sigma_a+\varepsilon)\int_0^X A(u)u^{-(\sigma_a+\varepsilon+1)}\,du. \tag{7.4}$$

If we set $R(u) = \sum_{n \ge u} \frac{a(n)}{n^{\sigma_a+\varepsilon}}$, then $a(n) = (R(n) - R(n+1))n^{\sigma_a+\varepsilon}$ and it follows that

$$A(u) = \sum_{n \le u} (R(n) - R(n+1))n^{\sigma_a+\varepsilon}.$$

Substituting this into Equation (7.4), we obtain

$$\int_0^X \sum_{n\le u}(R(n)-R(n+1))n^{\sigma_a+\varepsilon}u^{-(\sigma_a+\varepsilon+1)}\,du.$$

As $R(n)$ is constant on the interval $[n,n+1)$, linearity of the integral implies

$$\int_0^X \sum_{n\le u}(R(n)-R(n+1))n^{\sigma_a+\varepsilon}u^{-(\sigma_a+\varepsilon+1)}\,du = \sum_{0\le n\le X}(R(n)-R(n+1))n^{\sigma_a+\varepsilon}\int_n^{n+1} u^{-(\sigma_a+\varepsilon+1)} + O_\varepsilon(1),$$

where the O-estimate is present since X may not be an integer. Now $R(n)\ll_\varepsilon 1$ since it is the tail of $D(\sigma_a+\varepsilon)$ and moreover,

$$\int_n^{n+1} u^{-(\sigma_a+\varepsilon+1)} = -\frac{u^{-(\sigma_a+\varepsilon)}}{\sigma_a+\varepsilon}\bigg|_n^{n+1} = \frac{n^{-(\sigma_a+\varepsilon)}}{\sigma_a+\varepsilon} - \frac{(n+1)^{-(\sigma_a+\varepsilon)}}{\sigma_a+\varepsilon} \ll_\varepsilon 1,$$

because $\sigma_a+\varepsilon>0$. So

$$\int_0^X A(u)u^{-(\sigma_a+\varepsilon+1)}\,du = \int_0^X \sum_{n\le u}(R(n)-R(n+1))n^{\sigma_a+\varepsilon}u^{-(\sigma_a+\varepsilon+1)}\,du \ll_\varepsilon 1.$$

Also, $\sum_{n\le X}\frac{a(n)}{n^{\sigma_a+\varepsilon}}\ll_\varepsilon 1$ because $D(\sigma_a+\varepsilon)$ converges. We conclude

$$A(X)X^{-(\sigma_a+\varepsilon)} = \sum_{n\le X}\frac{a(n)}{n^{\sigma_a+\varepsilon}} - (\sigma_a+\varepsilon)\int_0^X A(u)u^{-(\sigma_a+\varepsilon+1)}\,du. \ll_\varepsilon 1,$$

which is equivalent to the desired estimate. □

A way to think about Proposition 7.1.3 is that if the abscissa of absolute convergence is $\sigma_a\ge 0$ then the size of the coefficients $a(n)$ is at most $n^{\sigma_a+\varepsilon}$ on average. Of corse, if $a(n)\ll_\alpha n^\alpha$ then Proposition 7.1.2 implies that $\sigma_a\le 1+\alpha$ and so Proposition 7.1.3 gives the significantly weaker estimate $A(X)\ll_\varepsilon X^{1+\alpha+\varepsilon}$. However, if we have a bound of the form $|A|(X)\ll_\alpha X^\alpha$ we can still obtain an upper estimate for the abscissa of absolute convergence:

Proposition 7.1.4. *Suppose $D(s)$ is a Dirichlet series such that $|A|(X)\ll_\alpha X^\alpha$ for some real $\alpha\ge 0$. Then the abscissa of absolute convergence satisfies $\sigma_a\le\alpha$.*

Proof. It suffices to show that $D(s)$ is absolutely convergent in the region $\sigma>\alpha$. Let s be in this region. Then

$$D(s)\ll \sum_{n\ge 1}\left|\frac{a(n)}{n^s}\right| = \sum_{n\ge 1}\frac{|a(n)|}{n^\sigma}.$$

By Abel's summation formula (see Appendix B.3),

$$\sum_{n\le N}\frac{|a(n)|}{n^\sigma} = |a(N)|N^{-\sigma} - |a(1)| + \sigma\int_1^N |A|(u)u^{-(\sigma+1)}\,du,$$

By assumption $|A|(u)\ll_\alpha u^\alpha$ and so $a(N)\ll_\alpha N^\alpha$. We then estimate as follows:

$$\sum_{n\le N}\frac{|a(n)|}{n^\sigma} = |a(N)|N^{-\sigma} - |a(1)| + \sigma\int_1^N |A|(u)u^{-(\sigma+1)}\,du \ll_\alpha |a(N)|N^{-\sigma} + |a(1)| + \sigma\int_1^N u^{\alpha-(\sigma+1)}\,du.$$

As $N \to \infty$, the left-hand side tends towards $\sum_{n\geq 1} \frac{|a(n)|}{n^\sigma}$. As for the right-hand side, the first term tends to zero since $\sigma > \alpha$ and the second term remains bounded as they are independent of N. For the third term, we compute

$$\int_1^N u^{\alpha-(\sigma+1)}\,du = \frac{u^{\alpha-\sigma}}{\alpha-\sigma}\bigg|_1^N = \frac{N^{\alpha-\sigma}}{\alpha-\sigma} - \frac{1}{\alpha-\sigma},$$

which is also bounded as $N \to \infty$. This finishes the proof. □

Do not be fooled; Proposition 7.1.4 is in general weaker than Proposition 7.1.2. For example, from our comments following Proposition 7.1.3, if $D(s)$ is a Dirichlet series with coefficients $a(n)$ and we have the estimate $|A|(X) \ll_\beta X^\beta$ for some real β then Proposition 7.1.4 only says that $\sigma_a \leq \beta$. If α is very small compared to β, this is a significantly worse upper bound for the abscissa of absolute convergence than what Proposition 7.1.2 would imply if $a(n) \ll_\alpha n^\alpha$. Actually, the question of sharp polynomial bounds for these coefficients can be very deep. However, if the coefficients $a(n)$ are always nonnegative, then **Landau's theorem** provides a way of obtaining a lower bound for their growth as well as describing a singularity of $D(s)$:

Theorem 7.1.3 (Landau's theorem). *Suppose $D(s)$ is a Dirichlet series with nonnegative coefficients $a(n)$ and finite abscissa of absolute convergence σ_a. Then σ_a is a singularity of $D(s)$.*

Proof. If we replace $a(n)$ by $a(n)n^{-\sigma_a}$ then we may assume $\sigma_a = 0$. Now suppose to the contrary that $D(s)$ was holomorphic at $s = 0$. Therefore for some $\delta > 0$, $D(s)$ is holomorphic in the domain

$$\mathcal{D} = \{s \in \mathbb{C} : \sigma_a > 0\} \cup \{s \in \mathbb{C} : |s| < \delta\}.$$

Write $D(s)$ as a power series at $s = 1$:

$$P(s) = \sum_{k\geq 0} c_k (s-1)^k,$$

where

$$c_k = \frac{D^{(k)}(1)}{k!} = \frac{1}{k!}\sum_{n\geq 1} \frac{a(n)(-\log(n))^k}{n},$$

because $D(s)$ is holomorphic and so we can differentiate termwise. The radius of convergence of $P(s)$ is the distance from $s = 1$ to the nearest singularity of $P(s)$. Since $P(s)$ is holomorphic on $\mathcal{D}$, the closest points are $\pm i\delta$. Therefore, the radius of convergence is at least $|1 \pm \delta| = \sqrt{1+\delta^2}$. We can write $\sqrt{1+\delta^2} = 1+\delta'$ for some $\delta' > 0$. Then for $|s-1| < 1+\delta'$, write $P(s)$ as

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

If s is real with $s < 1$, then this last double sum is a sum of positive terms because $a(n) \geq 0$. Moreover, since $P(s)$ is absolutely convergent the two sums can be interchanged by Fubini's theorem. Interchanging sums we see that

$$P(s) = \sum_{n\geq 1} \frac{a(n)}{n} \sum_{k\geq 0} \frac{(1-s)^k(\log(n))^k}{k!} = \sum_{n\geq 1} \frac{a(n)}{n} e^{(1-s)\log(n)} = \sum_{n\geq 1} \frac{a(n)}{n^s} = D(s),$$

for $-\delta' < s < 1$. As $\delta' > 0$, this implies that $D(s)$ converges absolutely (since $a(n)$ is nonnegative) for some $s < 0$ (say $s = -\frac{\delta'}{2}$) which contradicts $\sigma_a = 0$. □

Landau's theorem is very useful as it implies that if $D(s)$ is a Dirichlet series with nonnegative coefficients then $a(n) \not\ll_{\varepsilon} n^{\sigma_a-(1+\varepsilon)}$ because otherwise Proposition 7.1.2 implies $\sigma_a \le \sigma_a - \varepsilon$. Actually, Landau's theorem also implies $|A|(X) \not\ll_{\varepsilon} X^{\sigma_a-\varepsilon}$ for otherwise Proposition 7.1.4 would similarly imply $\sigma_a \le \sigma_\alpha - \varepsilon$. When we come across Dirichlet series whose coefficients are of polynomial growth or of polynomial growth on average, we will invoke these results without mention, except for Landau's theorem, as this is also common practice in the literature. Generally speaking, if the coefficients $a(n)$ are chosen at random, $D(s)$ will not possess any good properties outside of convergence in some region (it might not even possess that). However, the Dirichlet series we will encounter, and most of interest in the wild, will have multiplicative coefficients. In this case, the Dirichlet series admits an infinite product expression:

Proposition 7.1.5. *Suppose the coefficients $a(n)$ of a Dirichlet series $D(s)$ are multiplicative and satisfy $a(n) \ll_\alpha n^\alpha$ for some real $\alpha \ge 0$. Then*

$$D(s) = \prod_p \left(\sum_{k\ge0} \frac{a(p^k)}{p^{ks}} \right),$$

for $\sigma > 1+\alpha$. Conversely, suppose that there are coefficients $a(n)$ such that

$$\prod_p \left(\sum_{k\ge0} \left| \frac{a(p^k)}{p^{ks}} \right| \right),$$

converges for $\sigma > 1+\alpha$. Then the equality above defines a Dirichlet series $D(s)$ that converges absolutely in this region too. Moreover, if the coefficients $a(n)$ are completely multiplicative, then

$$D(s) = \prod_p (1 - a(p)p^{-s})^{-1},$$

for $\sigma > 1+\alpha$.

Proof. Since $a(n) \ll_\alpha n^\alpha$, Proposition 7.1.2 implies that $D(s)$ converges absolutely for $\sigma > 1+\alpha$. Let s be such that $\sigma > 1+\alpha$. Since

$$\sum_{k\ge0} \left| \frac{a(p^k)}{p^{ks}} \right| < \sum_{n\ge1} \left| \frac{a(n)}{n^s} \right|,$$

the infinite series on the left converges because the right does by the absolute convergence of $D(s)$. Now let $N > 0$ be an integer. Then by the fundamental theorem of arithmetic

$$\prod_{p<N} \left(\sum_{k\ge0} \frac{a(p^k)}{p^{ks}} \right) = \sum_{n<N} \frac{a(n)}{n^s} + \sideset{}{^*}\sum_{n\ge N} \frac{a(n)}{n^s}, \tag{7.5}$$

where the $*$ denotes that we are summing over only those additional terms $\frac{a(n)}{n^s}$ that appear in the expanded product on the left-hand side with $n \ge N$. As $N \to \infty$, the first sum on the right-hand side tends to $D(s)$ and the second sum tends to zero because it is part of the tail of $D(s)$ (which tends to zero by convergence). This proves that the product converges, and is equal to $D(s)$. Equation (7.5) also holds absolutely in the sense that

$$\prod_{p<N} \left(\sum_{k\ge0} \left| \frac{a(p^k)}{p^{ks}} \right| \right) = \sum_{n<N} \left| \frac{a(n)}{n^s} \right| + \sideset{}{^*}\sum_{n\ge N} \left| \frac{a(n)}{n^s} \right|, \tag{7.6}$$

since $D(s)$ converges absolutely. For the converse statement, since the product

$$\prod_p \left(\sum_{k \geq 0} \left| \frac{a(p^k)}{p^{ks}} \right| \right),$$

converges for $\sigma > 1 + \alpha$ each factor is necessarily finite. That is, for each prime p the series $\sum_{k\geq 0} \frac{a(p^k)}{p^{ks}}$ converges absolutely in this region. Now fix an integer $N > 0$. Then Equation (7.6) holds. Taking $N \to \infty$ in Equation (7.6), the left-hand side converges by assumption. Therefore the right-hand sides does too. But the first sum on the right-hand side tends to

$$\sum_{n \geq 1} \left| \frac{a(n)}{n^s} \right|,$$

and the second sum is part of its tail. So the first sum must convergence hence defining an absolutely convergent Dirichlet series in $\sigma > 1 + \alpha$, and the second sum must tend to zero. Lastly, if the $a(n)$ are completely multiplicative the formula for a geometric series gives

$$\prod_p \left(\sum_{k \geq 0} \frac{a(p^k)}{p^{ks}} \right) = \prod_p \left(\sum_{k \geq 0} \left(\frac{a(p)}{p^s} \right)^k \right) = \prod_p (1 - a(p)p^{-s})^{-1}.$$

$\square$

Note that in Proposition 7.1.5, the requirement for a product to define an absolutely convergent Dirichlet series is that the series defining the factors in the product must be absolutely convergent. Thankfully, this is always the case for geometric series. Now suppose $D(s)$ is a Dirichlet series that has the product expression

$$D(s) = \prod_p (1 - \alpha_1(p)p^{-s})^{-1}(1 - \alpha_2(p)p^{-s})^{-1} \cdots (1 - \alpha_d(p)p^{-s})^{-1}.$$

We call this product the **Euler product** of $D(s)$, and it is said to be of **degree** d. In Proposition 7.1.5, complete multiplicativity of the coefficients is enough to guarantee that $D(s)$ has an Euler product of degree 1, but in general $D(s)$ will admit an Euler product of degree $d > 1$ if the coefficients are only multiplicative but satisfy additional properties like a recurrence relation. When we come across Dirichlet series whose coefficients are multiplicative or we are given an Euler product we will use Proposition 7.1.5 without mention as this is common practice in the literature. Lastly, if $D(s)$ has an Euler product then for any $N \geq 1$ we let $D_{(N)}(s)$ denote the Dirichlet series with the factors $p \mid N$ in the Euler product removed. That is,

$$D_{(N)}(s) = D(s) \prod_{p \mid N} (1 - \alpha_1(p)p^{-s})(1 - \alpha_2(p)p^{-s}) \cdots (1 - \alpha_d(p)p^{-s}).$$

Dually, for any $N \geq 1$ we let $D_N(s)$ denote the Dirichlet series only consisting of the factors $p \mid N$ in the Euler product. That is,

$$D_N(s) = \prod_{p \mid N} (1 - \alpha_1(p)p^{-s})^{-1}(1 - \alpha_2(p)p^{-s})^{-1} \cdots (1 - \alpha_d(p)p^{-s})^{-1}.$$

7.2 Perron Formulas

With the Mellin inversion formula, it is not hard to prove a very useful integral expression for the sum of coefficients of a Dirichlet series. First, we setup some general notation. If $D(s)$ is a Dirichlet series with coefficients $a(n)$, then for $X > 0$, we set

$$A'(X) = \sideset{}{'}\sum_{n \le X} a(n),$$

where the $'$ indicates that the last term is multiplied by $\frac{1}{2}$ if X is an integer. We would like to relate $A'(X)$ to an integral involving the entire Dirichlet series $D(s)$. In particular, this integral is a type of inverse Mellin transform. Any formula that relates a finite sum of coefficients of a Dirichlet series to an integral involving the entire Dirichlet series is called a **Perron-type formula**. We will see several of them, the first being **(classical) Perron's formula** which is a consequence of Abel's summation formula and the Mellin inversion formula applied to Dirichlet series:

Theorem 7.2.1 (Perron's formula, classical version). *Let $D(s)$ be a Dirichlet series with coefficient $a(n)$ and finite and nonnegative abscissa of absolute convergence σ_a. Then for any $c > \sigma_a$,*

$$A'(X) = \frac{1}{2\pi i} \int_{(c)} D(s) X^s \, \frac{ds}{s}.$$

Proof. Let s be such that $\sigma > \sigma_a$. By Abel's summation formula (see Appendix B.3),

$$\sum_{n \ge 1} \frac{a(n)}{n^s} = \lim_{Y \to \infty} A'(Y) Y^{-s} + s \int_1^\infty A'(u) u^{-(s+1)} \, du.$$

Now $A'(Y) \le A(Y)$ and $A(Y) \ll_\varepsilon Y^{\sigma_a + \varepsilon}$ by Proposition 7.1.3 so that $A'(Y)Y^{-s} \ll_\varepsilon Y^{\sigma_a + \varepsilon - \sigma}$. Choosing $\varepsilon < \sigma - \sigma_a$, this latter term tends to zero as $Y \to \infty$, which implies that $A'(Y)Y^{-s} \to 0$ as $Y \to \infty$. Therefore we can write the equation above as

$$\frac{D(s)}{s} = \int_1^\infty A'(u) u^{-(s+1)} \, du = \int_0^\infty A'(u) u^{-(s+1)} \, du, \tag{7.7}$$

where the second equality follows because $A(u) = 0$ in the interval $[0, 1)$. The Mellin inversion formula immediately gives the result. □

As a corollary, we obtain an important integral representation for Dirichlet series:

Corollary 7.2.1. *Let $D(s)$ be a Dirichlet series with coefficient $a(n)$ and finite and nonnegative abscissa of absolute convergence σ_a. Then for $\sigma > \sigma_a$,*

$$D(s) = s \int_1^\infty A'(u) u^{-(s+1)} \, du.$$

Proof. The identity is an equivalent form of the first equality in Equation (7.7). □

We would like to relate this sum to an integral involving the entire Dirichlet series $D(s)$. In particular, this integral is a type of inverse Mellin transform. Any formula that resembles Perron's formula is particularly useful because it allows one examine a sum of coefficients of a Dirichlet series, a discrete object, by means of a complex integral where analytic techniques are at our disposal. There is also a truncated version of Perron's formula which is often more useful for estimates rather than abstract results. To state it, we need to setup some notation and will require a lemma. For any $c > 0$, consider the discontinuous integral (see [Dav80])

$$\delta(y) = \frac{1}{2\pi i}\int_{(c)} y^s \frac{ds}{s} = \begin{cases} 0 & \text{if } 0 < y < 1, \\ \frac{1}{2} & \text{if } y = 1, \\ 1 & \text{if } y > 1. \end{cases}$$

Also, for any $T > 0$, let

$$I(y,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} y^s \frac{ds}{s},$$

be $\delta(y)$ truncated outside of height T. The lemma we require gives an approximation for how close $I(y,T)$ is to $\delta(y)$ (see [Dav80] for a proof):

Lemma 7.2.1. *For any $c > 0$, $y > 0$, and $T > 0$, we have*

$$I(y,T) - \delta(y) = \begin{cases} O\left(y^c \min\left(1, \frac{1}{T|\log(y)|}\right)\right) & \text{if } y \neq 1, \\ O\left(\frac{c}{T}\right) & \text{if } y = 1. \end{cases}$$

We can now state and prove **(truncated) Perron's formula**:

Theorem 7.2.2 (Perron's formula, truncated version). *Let $D(s)$ be a Dirichlet series with coefficient $a(n)$ and finite and nonnegative abscissa of absolute convergence σ_a. Then for any $c > \sigma_a$ and $T > 0$,*

$$A'(X) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} D(s)X^s \frac{ds}{s} + O\left(X^c \sum_{\substack{n\geq 1 \\ n\neq X}} \frac{a(n)}{n^c} \min\left(1, \frac{1}{T\left|\log\left(\frac{X}{n}\right)\right|}\right) + \delta_X a(X)\frac{c}{T}\right),$$

where $\delta_X = 1, 0$ according to if X is an integer or not.

Proof. By Appendix E.1, we have

$$A'(X) = \sum_{n\geq 1} a(n)\delta\left(\frac{X}{n}\right).$$

Using Lemma 7.2.1, we may replace $\delta\left(\frac{X}{n}\right)$ and obtain

$$A'(X) = \sum_{n\geq 1} a(n)\frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{X^s}{n^s}\frac{ds}{s} + \sum_{\substack{n\geq 1 \\ n\neq X}} a(n) O\left(\frac{X^c}{n^c}\min\left(1, \frac{1}{T\left|\log\left(\frac{X}{n}\right)\right|}\right)\right) + \delta_X a(X) O\left(\frac{c}{T}\right).$$

Since $D(s)$ converges absolutely, the sum may be moved inside of the first O-estimate and then we may combine the resulting two O-estimates. Fubini's theorem implies we may interchange the sum and the integral, and the statement of the lemma follows. $\square$

There is a slightly weaker variant of truncated Perron's formula that follows as a corollary:

Corollary 7.2.2. *Let $D(s)$ be a Dirichlet series with coefficient $a(n)$ and finite and nonnegative abscissa of absolute convergence σ_a. Then for any $c > \sigma_a$ and $T > 0$,*

$$A'(X) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} D(s)X^s \frac{ds}{s} + O\left(\frac{X^c}{T}\right),$$

Proof. For sufficiently large X, we have

$$\min\left(1, \frac{1}{T\left|\log\left(\frac{X}{n}\right)\right|}\right) \ll \frac{X^c}{T}.$$

The statement now follows from truncated Perron's formula. □

There is also a version of Perron's formula where we add a smoothing function. For any $X > 0$, we set

$$A_\psi(X) = \sum_{n\geq 1} a(n)\psi\left(\frac{n}{X}\right),$$

where $\psi(y)$ is smooth function such that $\psi(y) \to 0$ as $y \to \infty$. This is most useful in two cases. The first is when we choose $\psi(y)$ to be a bump function. In this setting, the bump function can be chosen such that it assigns weight 1 or 0 to the coefficients $a(n)$ and we can estimate sums like

$$\sum_{\frac{X}{2}\leq n<X} a(n) \quad \text{or} \quad \sum_{X\leq n<X+Y} a(n),$$

for some X and Y with $Y < X$. Sums of this type are called **unweighted**. As an example of an unweighted sum, let $\psi(y)$ be a bump function that is identically 1 on $[0,1]$ and has compact support within the interval $\left[1, \frac{X+1}{X}\right]$. For example,

$$\psi(y) = \begin{cases} 1 & \text{if } 0 \leq y \leq 1, \\ e^{-\frac{1-y}{\frac{X+1}{X}-y}} & \text{if } 1 \leq y < \frac{X+1}{X}, \\ 0 & \text{if } y \geq \frac{X+1}{X}. \end{cases}$$

Then

$$A_\psi(X) = \sum_{n\geq 1} a(n)\psi\left(\frac{n}{X}\right) = \sum_{n\leq X} a(n).$$

In the second case, we want $\psi(y)$ to dampen the $a(n)$ with a weight other than 1 or 0. Sums of this type are called **weighted**. In any case, suppose the support of $\psi(y)$ is contained in $[0, M]$. These conditions will force the Mellin transform $\Psi(s)$ of $\psi(y)$ to exist and have nice properties. To see that $\Psi(s)$ exists, let K be a compact set in $\mathbb{C}$ and let $\alpha = \max_{s\in K}(\sigma)$ and $\beta = \min_{s\in K}\{\sigma\}$. Note that $\psi(y)$ is bounded because it is compactly supported. Then for $s \in K$,

$$\Psi(s) = \int_0^\infty \psi(y)y^s\,\frac{dy}{y} \ll \int_0^M y^{\sigma-1}\,dy \ll_{\alpha,\beta} 1.$$

Therefore $\Psi(s)$ is locally absolutely uniformly convergent for $s \in \mathbb{C}$. In particular, the Mellin inversion formula implies that $\psi(y)$ is the Mellin inverse of $\Psi(s)$. As for nice properties, $\Psi(s)$ does not grow too fast in vertical strips:

Proposition 7.2.1. *Suppose $\psi(y)$ is a bump function and let $\Psi(s)$ denote its Mellin transform. Then for any $N \geq 1$,*

$$\Psi(s) \ll (|s|+1)^{-N},$$

provided s is contained in the vertical strip $a < \sigma < b$ for any a and b with $0 < a < b$.

Proof. Fix a and b with $a < b$. Also, let the support of $\psi(y)$ be contained in $[0, M]$. Now consider

$$\Psi(s) = \int_0^\infty \psi(y) y^s \, \frac{dy}{y}.$$

Since $\psi(y)$ is compactly supported, integrating by parts yields

$$\Psi(s) = \frac{1}{s} \int_0^\infty \psi'(y) y^{s+1} \, \frac{dy}{y}.$$

Repeatedly integrating by parts $N \geq 1$ times, we arrive at

$$\Psi(s) = \frac{1}{s(s+1)\cdots(s+N-1)} \int_0^\infty \psi^{(N)}(y) y^{s+N} \, \frac{dy}{y}.$$

Therefore

$$\Psi(s) \ll (|s|+1)^{-N} \int_0^\infty \psi^{(N)}(y) y^{\sigma+N} \, \frac{dy}{y}.$$

The claim will follow if we can show that the integral is bounded. Since $\psi(y)$ is compactly supported in $[0, M]$ so is $\psi^{(N)}(y)$. In particular, $\psi^{(N)}(y)$ is bounded. Therefore

$$\int_0^\infty \psi^{(N)}(y) y^{\sigma+N} \, \frac{dy}{y} \ll \int_0^M y^{\sigma+N} \, \frac{dy}{y} = \frac{y^{\sigma+N}}{\sigma+N}\bigg|_0^M = \frac{M^{\sigma+N}}{\sigma+N} \ll \frac{M^{b+N}}{N} \ll 1,$$

where the second to last estimate follows because $a < \sigma < b$ with $0 < a < b$. So the integral is bounded and the claim follows. $\square$

The following theorem is **(smoothed) Perron's formula**:

Theorem 7.2.3 (Perron's formula, smoothed version). *Let $D(s)$ be a Dirichlet series with coefficient $a(n)$ and finite and nonnegative abscissa of absolute convergence σ_a. Let $\psi(y)$ be a bump function and denote its Mellin transform by $\Psi(s)$. Then for any $c > \sigma_a$,*

$$A_\psi(X) = \frac{1}{2\pi i} \int_{(c)} D(s) \Psi(s) X^s \, ds.$$

In particular,

$$\sum_{n \geq 1} a(n) \psi(n) = \frac{1}{2\pi i} \int_{(c)} D(s) \Psi(s) \, ds.$$

Proof. The first statement is just a computation:

$$\begin{aligned}
A_\psi(X) &= \sum_{n \geq 1} a(n) \psi\left(\frac{n}{X}\right) \\
&= \sum_{n \geq 1} \frac{a(n)}{2\pi i} \int_{(c)} \Psi(s) \left(\frac{n}{X}\right)^{-s} ds && \text{Mellin inversion formula} \\
&= \frac{1}{2\pi i} \int_{(c)} \sum_{n \geq 1} a(n) \Psi(s) \left(\frac{n}{X}\right)^{-s} ds && \text{FT} \\
&= \frac{1}{2\pi i} \int_{(c)} D(s) \Psi(s) X^s \, ds.
\end{aligned}$$

This proves the first statement. For the second statement, take $X = 1$. $\square$

Smoothed Perron's formula is useful because it is often more versatile as the convergence of the integral is improved if $\psi(y)$ is chosen appropriately.

7.3 Analytic Data of L-functions

We are now ready to discuss L-functions in some generality. In the following, we will denote an L-functions by $L(s,f)$, and for the moment, f will carry no formal meaning. It is only used to suggest that the L-function is attached to some interesting arithmetic object f. When we discuss specific L-functions, f will carry a formal meaning. An **L-series** $L(s,f)$ is a Dirichlet series

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

where the $a_f(n) \in \mathbb{C}$ are coefficients usually attached to some arithmetic object f. We call $L(s,f)$ an **L-function** if it satisfies the following properties:

(i) $L(s,f)$ is locally absolutely uniformly convergent for $\sigma > 1$ and admits a degree d_f Euler product:

$$L(s,f) = \sum_{n\geq 1} \frac{a_f(n)}{n^s} = \prod_p (1-\alpha_1(p)p^{-s})^{-1}(1-\alpha_2(p)p^{-s})^{-1}\cdots(1-\alpha_{d_f}(p)p^{-s})^{-1},$$

with $a_f(n), \alpha_j(p) \in \mathbb{C}$, $a_f(1) = 1$, and $|\alpha_j(p)| \leq p$ for all primes p. We call

$$L_p(s,f) = (1-\alpha_1(p)p^{-s})^{-1}(1-\alpha_2(p)p^{-s})^{-1}\cdots(1-\alpha_{d_f}(p)p^{-s})^{-1},$$

the **local factor** at p, and the $\alpha_j(p)$ are called the **local roots** (or **local parameters**) at p.

(ii) There exists a factor

$$\gamma(s,f) = \pi^{-\frac{d_f s}{2}} \prod_{1\leq j\leq d_f} \Gamma\left(\frac{s+\kappa_j}{2}\right),$$

with $\kappa_j \in \mathbb{C}$ that are either real or appear in conjugate pairs. We also require $\mathrm{Re}(\kappa_j) > -1$. The κ_j are called the **local roots** (or **local parameters**) at infinity.

(iii) There exists an integer $q(f) \geq 1$ called the **conductor** such that $\alpha_j(p) \neq 0$ for all prime p such that $p \nmid q(f)$. If $p \mid q(f)$, then we say p is **ramified** and is **unramified** otherwise. The **analytic conductor** $\mathfrak{q}(s,f)$ is defined to be

$$\mathfrak{q}(s,f) = q(f)\mathfrak{q}_\infty(s,f),$$

where we set

$$\mathfrak{q}_\infty(s,f) = \prod_{1\leq j\leq d_f} (|s+\kappa_j|+3).$$

For simplicity, we will also suppress the dependence upon s if $s=0$. That is, we write

$$\mathfrak{q}(f) = \mathfrak{q}(0,f) \quad \text{and} \quad \mathfrak{q}_\infty(f) = \mathfrak{q}_\infty(0,f).$$

(iv) The **completed L-function**

$$\Lambda(s,f) = q(f)^{\frac{s}{2}}\gamma(s,f)L(s,f),$$

must admit meromorphic continuation to $\mathbb{C}$ with at most poles at $s=0$ and $s=1$. Moreover, it must satisfy the **functional equation** given by

$$\Lambda(s,f) = \varepsilon(f)\Lambda(1-s,\overline{f}),$$

where $\varepsilon(f)$ is a complex number with $|\varepsilon(f)| = 1$ called the **root number** of $L(s,f)$, and $\overline{f}$ is an object associated to f called the **dual** of f such that $L(s,\overline{f})$ satisfies $a_{\overline{f}}(n) = \overline{a_f(n)}$, $\gamma(s,\overline{f}) = \overline{\gamma(s,f)}$, $q(\overline{f}) = q(f)$, and $\varepsilon(\overline{f}) = \overline{\varepsilon(f)}$. We call $L(s,\overline{f})$ the **dual** of $L(s,f)$. If $\overline{f} = f$, we say $L(s,f)$ is **self-dual**.

(v) $L(s,f)$ admits meromorphic continuation to $\mathbb{C}$ with at most a pole at $s = 1$ of order $r_f \in \mathbb{Z}$, and must be of order 1 (see Appendix B.4) after clearing the possible polar divisor.

Property (ii) ensures that $\gamma(s,f)$ is holomorphic and nonzero for $\sigma \geq 1$. Then as $\gamma(1,f)$ is nonzero, r_f is also the order of possible poles of $\Lambda(s,f)$ at $s = 0$ and $s = 1$ which are equal by the functional equation. It follows that $r_{\overline{f}} = -r_f$. As $L(s,f)$ admits meromorphic continuation by property (v), we denote the continuation by $L(s,f)$ as well. Also note that the product of L-functions is an L-function but the sum of L-functions need not be.

Remark 7.3.1. *The only L-function of degree 0 is $L(s,\mathbf{1})$ defined by setting $a_{\mathbf{1}}(1) = 1$ and $a_{\mathbf{1}}(n) = 0$ for all $n > 1$ so that $L(s,\mathbf{1}) \equiv 1$. Moreover, $\gamma(s,\mathbf{1}) \equiv 1$, $q(\mathbf{1}) = 1$, $\varepsilon(\mathbf{1}) = 1$, and $L(s,\mathbf{1})$ is self-dual.*

We say that an L-function $L(s,f)$ belongs to the **Selberg class** if the additional adjustments to (i), (ii), and (v) hold:

(i) The local roots at p satisfy $|\alpha_j(p)| = 1$ if $p \nmid q(f)$ and $|\alpha_j(p)| \leq 1$ if $p \mid q(f)$.

(ii) The local roots at infinity satisfy $\mathrm{Re}(\kappa_j) \geq 0$.

(v) $L(s,f)$ has at most a simple pole at $s = 1$. That is, $r_f \leq 1$.

The adjustment for (i) is called the **(generalized) Ramanujan-Petersson conjecture** and it forces $a_f(n) \ll \sigma_0(n) \ll_\varepsilon n^\varepsilon$ (recall Proposition A.3.1). The adjustment for (ii) is called the **(generalized) Selberg conjecture** and it ensures that $\gamma(s,f)$ is holomorphic and nonzero for $\sigma > 0$. As for the adjustment for (v), it is expected that $L(s,f)$ is entire unless $\alpha_i(p) \geq 0$. The Selberg class constitutes a very special class of L-functions. Note that the product of two Selberg class L-functions is also Selberg class. Accordingly, we say that a Selberg class L-function is **primitive** if it cannot be factored into a product of two Selberg class L-functions of strictly smaller positive degree (positive degree is necessary so that one of the factors cannot be $L(s,\mathbf{1})$). Clearly, any Selberg class L-function of degree 1 is primitive.

Remark 7.3.2. *In general, it is a difficult problem to determine when an L-function $L(s,f)$ of degree $d_f \geq 2$ is primitive.*

Suppose we are given two L-functions $L(s,f)$ and $L(s,g)$ with local roots $\alpha_j(p)$ and $\beta_\ell(p)$ at p and local roots κ_j and ν_ℓ at infinity respectively. We say that an L-function $L(s,f\otimes g)$ of degree $d_{f\otimes g} = d_f d_g$ is the **Rankin-Selberg convolution** of $L(s,f)$ and $L(s,g)$ (or **Rankin-Selberg square** if $f = g$) if it satisfies the following adjustments:

(i) The Euler product of $L(s,f\otimes g)$ takes the form

$$L(s,f\otimes g) = \prod_{p\nmid q(f)q(g)} L_p(s,f\otimes g) \prod_{p\mid q(f)q(g)} H_p(p^{-s}),$$

with

$$L_p(s,f\otimes g) = \prod_{\substack{1\leq j\leq d_f\\ 1\leq \ell\leq d_g}} \left(1 - \alpha_j(p)\overline{\beta_\ell(p)}p^{-s}\right)^{-1} \quad \text{and} \quad H_p(p^{-s}) = \prod_{1\leq j\leq d_{f\otimes g}} (1 - \gamma_j(p)p^{-s}),$$

for some $\gamma_j(p) \in \mathbb{C}$ with $|\gamma_j(p)| \leq p$.

(ii) The gamma factor $\gamma(s, f \otimes g)$ takes the form

$$\gamma(s, f \otimes g) = \pi^{-\frac{d_{f\otimes g}s}{2}} \prod_{\substack{1 \le j \le d_f \\ 1 \le \ell \le d_g}} \Gamma\left(\frac{s+\mu_{j,\ell}}{2}\right),$$

with the local roots at infinity satisfying the additional bounds $\mathrm{Re}(\mu_{j,\ell}) \le \mathrm{Re}(\kappa_j) + \mathrm{Re}(\nu_\ell)$ and $|\mu_{j,\ell}| \le |\kappa_j| + |\nu_\ell|$.

(iii) The root number $q(f\otimes g)$ satisfies $q(f\otimes g) \mid q(f)^{d_f} q(g)^{d_g}$. If $q(f\otimes g)$ is a proper divisor of $q(f)^{d_f} q(g)^{d_g}$, we say that **conductor dropping** occurs.

(v) $L(s, f \otimes g)$ has a pole of order $r_{f\otimes g} \ge 1$ at $s = 1$ if $g = \overline{f}$.

We now introduce some important concepts associated to L-functions. The **critical strip** is the vertical strip left invariant by the transformation $s \to 1 - s$, that is, the region defined by

$$\left\{ s \in \mathbb{C} : \left|\sigma - \frac{1}{2}\right| \le \frac{1}{2} \right\}.$$

Moreover, it is also the region where we cannot determine the value of an L-function from its representation as a Dirichlet series or using the functional equation. It turns out that much of the important information about L-functions are contained inside of the critical strip. The **critical line** is the line left invariant by the transformation $s \to 1 - s$ which is the line defined by $\sigma = \frac{1}{2}$. It is also the line that bisects the critical strip vertically. The **central point** is the fixed point of the transformation $s \to 1 - s$, in other words, the point $s = \frac{1}{2}$. Clearly the central point is also the center of the critical line. The critical strip, critical line, and central point are all displayed in Figure 7.1:

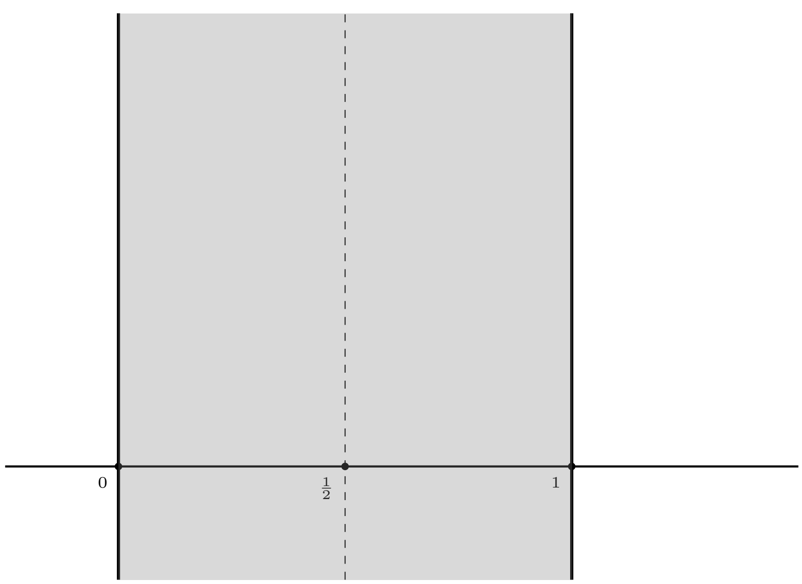

Figure 7.1: The critical strip, critical line, and central point.

Lastly, we provide some bounds about gamma functions, gamma factors, and their logarithmic derivatives that will be extremely useful in the study of L-functions. Suppose σ is bounded and $|t| > 1$. Then s is bounded away from zero and by Corollary 1.7.2 we have the asymptotic

$$\Gamma(s) \sim \sqrt{2\pi} t^{\sigma - \frac{1}{2}} e^{-\frac{\pi}{2}|t|}.$$

This gives the weaker estimates

$$\Gamma(s) \ll t^{\sigma-\frac{1}{2}}e^{-\frac{\pi}{2}|t|} \quad \text{and} \quad \frac{1}{\Gamma(s)} \ll t^{\frac{1}{2}-\sigma}e^{\frac{\pi}{2}|t|},$$

for $|t| > 1$. It is not hard to obtain estimates that holds in vertical strips. Suppose s is in the vertical strip $a \le \sigma \le b$ and is distance ε away from the poles of $\Gamma(s)$. Then $\Gamma(s)$ is bounded on the compact region $a \le \sigma \le b$ with $|t| \le 1$ provided s is distance ε away from the poles of $\Gamma(s)$. It follows that the estimates

$$\Gamma(s) \ll_\varepsilon (|t|+1)^{\sigma-\frac{1}{2}}e^{-\frac{\pi}{2}|t|} \quad \text{and} \quad \frac{1}{\Gamma(s)} \ll (|t|+1)^{\frac{1}{2}-\sigma}e^{\frac{\pi}{2}|t|}, \tag{7.8}$$

are valid in the vertical strip $a \le \sigma \le b$ provided s is distance ε away from the poles of $\Gamma(s)$ in the former case. We immediately obtain the useful estimate

$$\frac{\Gamma(1-s)}{\Gamma(s)} \ll_\varepsilon (|t|+1)^{1-2\sigma},$$

valid in the vertical strip $a \le \sigma \le b$ provided s is distance ε away from the poles of $\Gamma(1-s)$. Moreover, it easily follows from the definition of $\gamma(s,f)$ that the estimates

$$\frac{\gamma(1-s,f)}{\gamma(s,f)} \ll_\varepsilon \mathfrak{q}_\infty(s,f)^{\frac{1}{2}-\sigma} \quad \text{and} \quad q(f)^{\frac{1}{2}-s}\frac{\gamma(1-s,f)}{\gamma(s,f)} \ll_\varepsilon \mathfrak{q}(s,f)^{\frac{1}{2}-\sigma}, \tag{7.9}$$

are valid in the vertical strip $a \le \sigma \le b$ provided s is distance ε away from the poles of $\gamma(1-s,f)$. There are analogous useful estimates for the digamma function as well. To derive them, suppose σ is bounded and $|t| > 1$. Then s is bounded away from zero, so making use of the formula $\frac{\Gamma'}{\Gamma}(s+1) = \frac{\Gamma'}{\Gamma}(s) + \frac{1}{s}$ if $\sigma < 0$, Proposition 1.7.3 gives the estimate

$$\frac{\Gamma'}{\Gamma}(s) \ll \log(s).$$

We can also obtain estimates that holds in vertical strips. Suppose s is in the vertical strip $a \le \sigma \le b$ and is distance ε away from the poles of $\Gamma(s)$. Then $\frac{\Gamma'}{\Gamma}(s)$ is bounded on the compact region $a \le \sigma \le b$ with $|t| \le 1$ provided s is distance ε away from the poles of $\Gamma(s)$. It follows that the estimate

$$\frac{\Gamma'}{\Gamma}(s) \ll_\varepsilon \log(|s|+1),$$

is valid in the vertical strip $a \le \sigma \le b$ provided s is distance ε away from the poles of $\Gamma(s)$. It follows from the definition of $\gamma(s,f)$ that the estimates

$$\frac{\gamma'}{\gamma}(s,f) \ll_\varepsilon \log \mathfrak{q}_\infty(s,f) \quad \text{and} \quad \log(q(f)) + \frac{\gamma'}{\gamma}(s,f) \ll_\varepsilon \log \mathfrak{q}(s,f), \tag{7.10}$$

are valid in the vertical strip $a \le \sigma \le b$ provided s is distance ε away from the poles of $\gamma(s,f)$.

7.4 The Approximate Functional Equation

If $L(s,f)$ is an L-function, then there is a formula which acts as a compromise between the functional equation for $L(s,f)$ and expressing $L(s,f)$ as a Dirichlet series. This formula is known as the approximate functional equation and it is important because it is valid inside of the critical strip and therefore can be used to obtain data about $L(s,f)$ in that region. First we use Equation (7.9) to show that $L(s,f)$ has polynomial growth in the t-aspect in vertical strips:

Proposition 7.4.1. *For any L-function $L(s,f)$ and $a<b$, $L(s,f)$ is of polynomial growth in the t-aspect in the vertical half-strip $a \le \sigma \le b$ with $|t| \ge 1$.*

Proof. On the one hand, for $\sigma > \max(1,b)$ we have $L(s,f) \ll 1$ on the line $\sigma = 1$ with $t \ge 1$. On the other hand, the functional equation and Equation (7.9) together imply

$$L(s,f) \ll \mathfrak{q}(s,f)^{\frac{1}{2}-\sigma} L(1-s,f).$$

Clearly $\mathfrak{q}(s,f)^{\frac{1}{2}-\sigma}$ is of polynomial growth in the t-aspect provided σ is bounded. As $|\sigma| < \max(a,b)$, we see that $L(s,f)$ is also of polynomial growth in the t-aspect for $|t| \ge 1$. Moreover, this estimate also implies $L(s,f)$ is bounded on the line $t=1$ since σ is bounded. As $L(s,f)$ is holomorphic for $|t| \ge 1$ and of order 1, we can apply the Phragmén-Lindelöf convexity principle in this region (see Appendix B.5) so that $L(s,f)$ is of polynomial growth in the t-aspect in the vertical half-strip $a \le \sigma \le b$ with $|t| \ge 1$. □

Proposition 7.4.1 is a very important property that L-functions possess. It is usually used to estimate Perron type formulas. We can also use it to deduce the **approximate function equation**:

Theorem 7.4.1 (Approximate functional equation). *Let $L(s,f)$ be an L-function, $\Phi(u)$ be an even holomorphic function bounded in the vertical strip $|\tau| < a+1$ for any $a>1$ such that $\Phi(0)=1$, and let $X>0$. Then for s in the critical strip, we have*

$$L(s,f) = \sum_{n \ge 1} \frac{a_f(n)}{n^s} V_s\left(\frac{n}{\sqrt{q(f)}X}\right) + \varepsilon(s,f) \sum_{n \ge 1} \frac{\overline{a_f(n)}}{n^{1-s}} V_{1-s}\left(\frac{nX}{\sqrt{q(f)}}\right) + \frac{R}{q(f)^{\frac{s}{2}}\gamma(s,f)},$$

where $V_s(y)$ is the inverse Mellin transform defined by

$$V_s(y) = \frac{1}{2\pi i}\int_{(a)} \frac{\gamma(s+u,f)}{\gamma(s,f)}\Phi(u) y^{-u}\frac{du}{u},$$

and

$$\varepsilon(s,f) = \varepsilon(f) q(f)^{\frac{1}{2}-s}\frac{\gamma(1-s,f)}{\gamma(s,f)}.$$

Moreover, R is zero if $\Lambda(s,f)$ is entire, and otherwise

$$R = \left(\operatorname*{Res}_{u=1-s} + \operatorname*{Res}_{u=-s}\right) \frac{\Lambda(s+u,f)\Phi(u)X^u}{u}.$$

Proof. Let

$$I(X,s,f) = \frac{1}{2\pi i}\int_{(a)} \Lambda(s+u,f)\Phi(u)X^u\frac{du}{u}.$$

$L(s,f)$ has polynomial growth in the t-aspect by Proposition 7.4.1. From Equation (7.8) we see that $\gamma(s+u,f)$ exhibits rapid decay. Since $\Phi(u)$ is bounded, it follows that the integrand exhibits rapid decay in a vertical strip containing $|\tau| \le a$. Therefore the integral is locally absolutely uniformly convergent. Moreover, we may shift the line of integration to $(-a)$. In doing so, we pass by a simple pole at $u=0$ and possible poles at $u=1-s$ and $u=-s$, giving

$$I(X,s,f) = \frac{1}{2\pi i}\int_{(-a)} \Lambda(s+u,f)\Phi(u)X^u\frac{du}{u} + \Lambda(s,f) + R.$$

Applying the functional equation to $\Lambda(s+u,f)$ and performing the change of variables $u \to -u$, we obtain

$$I(X,s,f) = -\varepsilon(f)I(X^{-1},1-s,\overline{f}) + \Lambda(s,f) - R,$$

since $\Phi(u)$ is even. This equation is equivalent to

$$\Lambda(s,f) = I(X,s,f) + \varepsilon(f)I(X^{-1},1-s,\overline{f}) + R.$$

Since $\mathrm{Re}(s+u) > 1$, we can expand the L-function $L(s,f)$ inside of $I(X,s,f)$ as a Dirichlet series:

$$\begin{aligned}
I(X,s,f) &= \frac{1}{2\pi i}\int_{(a)} \Lambda(s+u,f)\Phi(u)X^u\,\frac{du}{u} \\
&= \frac{1}{2\pi i}\int_{(a)} q(f)^{\frac{s+u}{2}}\gamma(s+u,f)L(s+u,f)\Phi(u)X^u\,\frac{du}{u} \\
&= \frac{1}{2\pi i}\int_{(a)} \sum_{n\ge1}\frac{a_f(n)}{n^{s+u}}q(f)^{\frac{s+u}{2}}\gamma(s+u,f)\Phi(u)X^u\,\frac{du}{u} \\
&= \sum_{n\ge1}\frac{1}{2\pi i}\int_{(a)} \frac{a_f(n)}{n^{s+u}}q(f)^{\frac{s+u}{2}}\gamma(s+u,f)\Phi(u)X^u\,\frac{du}{u} && \text{FT} \\
&= q(f)^{\frac{s}{2}}\gamma(s,f)\sum_{n\ge1}\frac{a_f(n)}{n^s}\frac{1}{2\pi i}\int_{(a)}\frac{\gamma(s+u,f)}{\gamma(s,f)}\Phi(u)\left(\frac{\sqrt{q(f)}X}{n}\right)^u\frac{du}{u} \\
&= q(f)^{\frac{s}{2}}\gamma(s,f)\sum_{n\ge1}\frac{a_f(n)}{n^s}V_s\left(\frac{n}{\sqrt{q(f)}X}\right).
\end{aligned}$$

Performing the same computation for $I(X^{-1},1-s,\overline{f})$ and substituting in the results, we arrive at

$$\Lambda(s,f) = q(f)^{\frac{s}{2}}\gamma(s,f)\sum_{n\ge1}\frac{a_f(n)}{n^s}V_s\left(\frac{n}{\sqrt{q(f)}X}\right) + \varepsilon(f)q(f)^{\frac{1-s}{2}}\gamma(1-s,f)\sum_{n\ge1}\frac{a_f(n)}{n^{1-s}}V_{1-s}\left(\frac{nX}{\sqrt{q(f)}}\right) + R.$$

Diving by $q(f)^{\frac{s}{2}}\gamma(s,f)$ completes the proof. $\square$

The approximate functional equation was first developed by Hardy and Littlewood in the series [HL21, HL23, HL29]. The function $V_s(y)$ has the effect of smoothing out the two sums on the right-hand side of the approximate functional equation. In most cases, we will take

$$\Phi(u) = \cos^{-4d_fM}\left(\frac{\pi u}{4M}\right),$$

for an integer $M \ge 1$. Clearly $\Phi(u)$ holomorphic in the vertical strip $|\tau| < (2M-2)+1$, even, and satisfies $\Phi(0) = 1$. To see that it is bounded in this vertical strip, using the formula $\cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2}$, we have

$$\cos^{-4d_fM}\left(\frac{\pi u}{4M}\right) = \left(\frac{e^{i\frac{\pi u}{4M}} + e^{-i\frac{\pi u}{4M}}}{2}\right)^{-4d_fM} \ll e^{-d_f\pi|r|}, \tag{7.11}$$

where in the estimate we have used the reverse triangle equality. It follows that $\Phi(u)$ exhibits rapid decay. With this choice of $\Phi(u)$, we can prove a useful bound for $V_s(y)$:

Proposition 7.4.2. *Let $L(s,f)$ be an L-function, set $\Phi(u) = \cos^{-4d_f M}\left(\frac{\pi u}{4M}\right)$ for some integer $M \geq 1$, and let $V_s(y)$ be the inverse Mellin transform defined by*

$$V_s(y) = \frac{1}{2\pi i}\int_{(2M-2)} \frac{\gamma(s+u,f)}{\gamma(s,f)}\Phi(u)y^{-u}\frac{du}{u}.$$

Then for s in the critical strip, $V_s(y)$ satisfies the estimate

$$V_s(y) \ll \left(1+\frac{y}{\sqrt{\mathfrak{q}_\infty(s,f)}}\right)^{-M}.$$

Proof. Suppose $0 \leq \sigma \leq \frac{1}{2}$ so that $\sigma - \frac{1}{2} \leq 0$. Then from Equation (7.8) and the reverse triangle inequality, we deduce that

$$\frac{\Gamma(s+u)}{\Gamma(s)} \ll \frac{(|t+r|+1)^{\sigma+\tau-\frac{1}{2}}}{(|t|+1)^{\sigma-\frac{1}{2}}}e^{\frac{\pi}{2}(|t|-|t+r|)} \ll (|t+r|+1)^{\tau}e^{\frac{\pi}{2}|r|}.$$

From the definition of $\mathfrak{q}(s,f)$, the above estimate implies

$$\frac{\gamma(s+u,f)}{\gamma(s,f)} \ll \mathfrak{q}_\infty(s,f)^{\frac{\tau}{2}}e^{d_f\frac{\pi}{2}|r|},$$

for s in the critical strip. By Equation (7.11), the integral defining $V_s(y)$ is locally absolutely uniformly convergent and we may shift the line of integration to (M). In doing so, we do not pass by any poles and obtain

$$V_s(y) = \frac{1}{2\pi i}\int_{(M)} \frac{\gamma(s+u,f)}{\gamma(s,f)}\Phi(u)y^{-u}\frac{du}{u}$$

The fact that u is bounded away from zero, Equation (7.11), and the estimate for the ratio of gamma factors, gives the first estimate in the following chain:

$$\begin{aligned}
V_s(y) &= \frac{1}{2\pi i}\int_{(M)} \frac{\gamma(s+u,f)}{\gamma(s,f)}\Phi(u)y^{-u}\frac{du}{u} \\
&\ll \int_{\mathbb{R}} \mathfrak{q}_\infty(s,f)^{\frac{M}{2}}e^{-d_f\frac{\pi}{2}|r|}y^{-M}\,dr \\
&\ll \int_{\mathbb{R}} \mathfrak{q}_\infty(s,f)^{\frac{M}{2}}e^{-d_f\frac{\pi}{2}|r|}(1+y)^{-M}\,dr \\
&\ll \int_{\mathbb{R}} e^{-\frac{\pi}{2}d_f|r|}\left(1+\frac{y}{\sqrt{\mathfrak{q}_\infty(s,f)}}\right)^{-M}\,dr \\
&\ll \left(1+\frac{y}{\sqrt{\mathfrak{q}_\infty(s,f)}}\right)^{-M}\int_{\mathbb{R}} e^{-\frac{\pi}{2}d_f|r|}\,dr \\
&\ll \left(1+\frac{y}{\sqrt{\mathfrak{q}_\infty(s,f)}}\right)^{-M},
\end{aligned}$$

where in the third and fourth lines we have used that $cy \ll (1+cy)$ for all $y \geq 0$ and any c and the last line holds since the integrand exhibits rapid decay. This completes the proof. □

From Proposition 7.4.2 we see that $V_s(y)$ is bounded for $y \ll_\varepsilon \mathfrak{q}_\infty(s,f)^{\frac{1}{2}+\varepsilon}$ and then starts to exhibit polynomial decay that can be taken arbitrarily large. In a similar spirit to the approximate functional equation, a useful summation formula can be derived from the functional equation of each L-function:

Theorem 7.4.2. *Let $\psi(y)$ be a bump function and let $\Psi(s)$ denote its Mellin transform. Then for any L-function $L(s,f)$, we have*

$$\sum_{n\geq 1} a_f(n)\psi(n) = \frac{\varepsilon(f)}{\sqrt{q(f)}}\sum_{n\geq 1} a_{\overline{f}}(n)\phi(n) + R\Psi(1),$$

where $\phi(y)$ is the inverse Mellin transform defined by

$$\phi(y) = \frac{1}{2\pi i}\int_{(a)} q(f)^s \frac{\gamma(s,f)}{\gamma(1-s,f)} y^{-s}\Psi(1-s)\,ds,$$

for any $a>1$. Moreover, R is zero if $L(s,f)$ is entire, and otherwise

$$R = \operatorname*{Res}_{s=1} L(s,f).$$

Proof. By smoothed Perron's formula,

$$\sum_{n\geq 1} a_f(n)\psi(n) = \frac{1}{2\pi i}\int_{(a)} L(s,f)\Psi(s)\,ds.$$

By Propositions 7.2.1 and 7.4.1, the integrand has polynomial decay of arbitrarily larger order and therefore is locally absolutely uniformly convergent. Shifting the line of integration to $(1-a)$, we pass by a potential pole at $s=1$ from $L(s,f)$ and obtain

$$\sum_{n\geq 1} a_f(n)\psi(n) = \frac{1}{2\pi i}\int_{(1-a)} L(s,f)\Psi(s)\,ds + R\Psi(1).$$

Applying the functional equation, we further have

$$\sum_{n\geq 1} a_f(n)\psi(n) = \frac{1}{2\pi i}\int_{(1-a)} \varepsilon(f)q(f)^{\frac{1}{2}-s}\frac{\gamma(1-s,f)}{\gamma(s,f)}L(1-s,\overline{f})\Psi(s)\,ds + R\Psi(1).$$

Performing the change of variables $s\to 1-s$ in this latter integral gives

$$\sum_{n\geq 1} a_f(n)\psi(n) = \frac{1}{2\pi i}\int_{(a)} \varepsilon(f)q(f)^{s-\frac{1}{2}}\frac{\gamma(s,f)}{\gamma(1-s,f)}L(s,\overline{f})\Psi(1-s)\,ds + R\Psi(1).$$

The proof is complete upon interchanging the sum over the Dirichlet series and the integral by Fubini's theorem and factoring out $\frac{\varepsilon(f)}{\sqrt{q(f)}}$. $\square$

7.5 The Riemann Hypothesis & Nontrivial Zeros

The zeros of an L-function $L(s,f)$ have interesting behavior. Recall that

$$L(s,f) = \prod_p (1-\alpha_1(p)p^{-s})^{-1}(1-\alpha_2(p)p^{-s})^{-1}\cdots(1-\alpha_{d_f}(p)p^{-s})^{-1},$$

for $\sigma>1$. This product vanishes if and only if one of its factors are zero. As $\sigma>1$, this is impossible so that $L(s,f)$ has no zeros in this region. The functional equation will allow us to understand more about the zeros of $L(s,f)$. Rewrite the functional equation for $L(s,f)$ as

$$L(s,f) = \varepsilon(f)q(f)^{\frac{1}{2}-s}\frac{\gamma(1-s,f)}{\gamma(s,f)}L(1-s,\overline{f}). \tag{7.12}$$

If $\sigma < 0$ then $L(1-s,\overline{f})$ is nonzero by our previous comments. Moreover, $\gamma(1-s,f)$ is holomorphic and nonzero in this region because $\mathrm{Re}(\kappa_j) > -1$. We conclude that poles of $\gamma(s,f)$ are zeros of $L(s,f)$ for $\sigma < 0$. Such a zero is called a **trivial zero**. From the definition of $\gamma(s,f)$, they are all simple and of the form $s = -(\kappa_j + 2n)$ for some local root at infinity κ_j and some integer $n \geq 0$. Any other zero of $L(s,f)$ is called a **nontrivial zero** and it lies inside of the critical strip (it may also be a pole of $\gamma(s,f)$). Now let ρ be a nontrivial zero of $L(s,f)$. Note that $L(\overline{s},\overline{f}) = \overline{L(s,f)}$ for $\sigma > 1$ where $L(s,f)$ is defined by a Dirichlet series and thus for all s by the identity theorem. It follows that $\overline{\rho}$ is a nontrivial zero of $L(s,\overline{f})$ an from the functional equation $1-\overline{\rho}$ is also a nontrivial zero of $L(s,f)$. In short, the nontrivial zeros occur in pairs:

$$\rho \quad \text{and} \quad 1-\overline{\rho}.$$

We can sometimes say more. If $L(s,f)$ takes real values for $s > 1$, the Schwarz reflection principle implies $L(\overline{s},f) = \overline{L(s,f)}$ and that $L(s,f)$ takes real values on the entire real axis save for the possible poles at $s = 0$ and $s = 1$. We find that $\overline{\rho}$ and $1-\overline{\rho}$ are nontrivial zeros too and therefore the nontrivial zeros of $L(s,f)$ come in sets of four and are displayed in Figure 7.2:

$$\rho, \quad \overline{\rho}, \quad 1-\rho, \quad \text{and} \quad 1-\overline{\rho}.$$

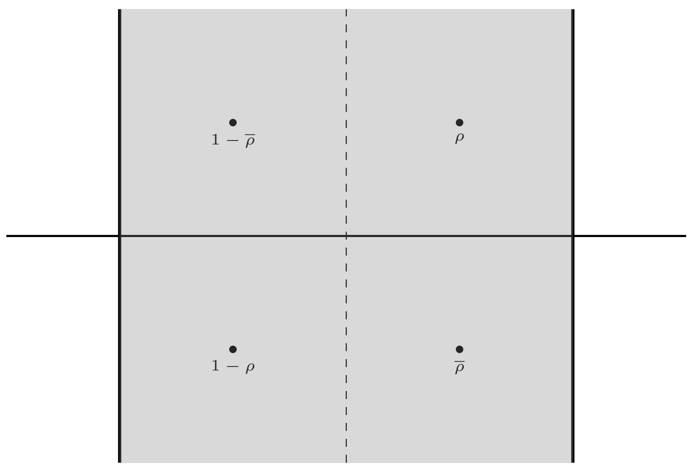

Figure 7.2: Symmetric nontrivial zeros.

The **Riemann hypothesis** for $L(s,f)$ says that this symmetry should be as simple as possible:

Conjecture 7.5.1 (Riemann hypothesis, $L(s,f)$ version). *For the L-function $L(s,f)$, all of the nontrivial zeros lie on the line $\sigma = \frac{1}{2}$.*

Somewhat confusingly, do not expect the Riemann hypothesis to hold for just any L-function but we expect it to hold for many L-functions. In particular, the **grand Riemann hypothesis** says that this symmetry should hold for any L-function in the Selberg class:

Conjecture 7.5.2 (Grand Riemann hypothesis). *For any Selberg class L-function $L(s,f)$, all of the nontrivial zeros lie on the line $\sigma = \frac{1}{2}$.*

So far, the Riemann hypothesis remains completely out of reach for any L-function and thus the grand Riemann hypothesis does as well.

7.6 The Lindelöf Hypothesis & Convexity Bounds

Instead of asking about the zeros of an L-function $L(s,f)$ on the critical line, we can ask about the growth of $L(s,f)$, and more generally its derivatives, on the critical line. More precisely, we want to derive an upper bound for $L(s,f)$, or one of its derivatives, on the critical line using the Phragmén-Lindelöf convexity principle. The argument we will describe for $L(s,f)$ is essentially a refinement of the proof of Proposition 7.4.1. Let

$$p_{r_f}(s) = \left(\frac{s-1}{s+1}\right)^{r_f}.$$

Note that $p_{r_f}(s) \sim 1$. The first step is to guarantee the Phragmén-Lindelöf convexity principle for $p_{r_f}(s)L(s,f)$ in a region containing the critical strip. As $L(s,f)$ is of order 1, this is assured (see Appendix B.5). Therefore, we are reduced to estimating the growth of $p_{r_f}(s)L(s,f)$ for σ to the left of 0 and to the right of 1. That is, just outside the edges of the critical strip. The right edge is easily estimated by setting $\sigma = 1+\varepsilon$ so that

$$p_{r_f}(1+\varepsilon+it)L(1+\varepsilon+it,f) \ll_\varepsilon 1,$$

which holds since $L(s,f)$ is defined by a locally absolutely uniformly convergent Dirichlet series for $\sigma > 1$. The left edge is only slightly more difficult. Upon isolating $L(s,f)$ in the functional equation, we have

$$L(s,f) = \varepsilon(f)q(f)^{\frac{1}{2}-s}\frac{\gamma(1-s,f)}{\gamma(s,f)}L(1-s,f).$$

Applying Equation (7.9) result in the bound

$$L(s,f) \ll_\varepsilon \mathfrak{q}(s,f)^{\frac{1}{2}-\sigma}L(1-s,f),$$

for s in any vertical strip with distance ε away from the poles of $\gamma(1-s,f)$. Multiplying both sides by $p_{r_f}(s)$ and taking $\sigma = -\varepsilon$, it follows that

$$p_{r_f}(-\varepsilon+it)L(-\varepsilon+it,f) \ll_\varepsilon \mathfrak{q}(s,f)^{\frac{1}{2}+\varepsilon},$$

which holds since $L(1-s,f)$ is defined by a locally absolutely uniformly convergent Dirichlet series for $\sigma < 0$. As $p_{r_f}(s)L(s,f)$ is holomorphic in a region containing the vertical strip $-\varepsilon \le \sigma \le 1+\varepsilon$ and $p_{r_f}(s) \sim 1$, the Phragmén-Lindelöf convexity principle gives

$$L(s,f) \ll_\varepsilon \mathfrak{q}(s,f)^{\frac{1-\sigma}{2}+\varepsilon}, \tag{7.13}$$

in the vertical strip $-\varepsilon \le \sigma \le 1+\varepsilon$ provided s is distance ε away from the pole of $L(s,f)$ at $s=1$ if it exists. At the critical line Equation (7.13) gives the following **convexity bound**:

$$L\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\frac{1}{4}+\varepsilon}. \tag{7.14}$$

The **Lindelöf hypothesis** for $L(s,f)$ says that the exponent can be reduced to ε:

Conjecture 7.6.1 (Lindelöf hypothesis, $L(s,f)$ version). *For the L-function $L(s,f)$, we have*

$$L\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\varepsilon}.$$

Just as for the Riemann hypothesis, we do not expect the Lindelöf hypothesis to hold for just any L-function. Accordingly, the **grand Lindelöf hypothesis** says that the exponent can be reduced to ε for any L-function in the Selberg class and we expect this to hold:

Conjecture 7.6.2 (Grand Lindelöf hypothesis). *For any Selberg class L-function $L(s,f)$, we have*

$$L\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\varepsilon}.$$

Like the Riemann hypothesis, we have been unable to prove the Lindelöf hypothesis for any L-function. However, the Lindelöf hypothesis seems to be much more tractable. Generally speaking, any improvement upon the exponent in the convexity bound in any aspect of the analytic conductor is called a **subconvexity estimate** (or a **convexity breaking bound**). In other words, we would want a bound of the form

$$L\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\delta+\varepsilon},$$

for some $0 \le \delta \le \frac{1}{4}$. The convexity bound says that we may take $\delta = \frac{1}{4}$ while the Lindelöf hypothesis for $L(s,f)$ implies that we may take $\delta = 0$.

Remark 7.6.1. *Some subconvexity bounds are deserving of names. The cases $\delta = \frac{3}{16}$ and $\delta = \frac{1}{6}$ are referred to as the **Burgess bound** and **Weyl bound** respectively.*

With a little more work, we can obtain a a similar bound for $L^{(n)}(s,f)$ for any $n \ge 1$. First observe that Equation (7.13) implies

$$L(s,f) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\frac{1}{4}+\varepsilon},$$

in the vertical strip $\left|\sigma - \frac{1}{2}\right| \le \frac{\varepsilon}{2}$. This is just slightly stronger than the convexity bound. Now Cauchy's integral formula gives

$$L^{(n)}\left(\frac{1}{2}+it,f\right) = \frac{n!}{2\pi i}\oint_C \frac{L(s,f)}{(s-\frac{1}{2}-it)^{n+1}}\,ds \ll_\varepsilon \oint_C |L(s,f)|ds,$$

where C is a circle about $\frac{1}{2}+it$ of radius $\frac{\varepsilon}{2}$. These two estimates, and that the disk bounded by C is compact, together imply the following **convexity bound**:

$$L^{(n)}\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\frac{1}{4}+\varepsilon}.$$

7.7 Estimating the Central Value

The Lindelöf hypothesis is concerned with the growth of the L-function $L(s,f)$ along the critical line, but sometimes we are only concerned with the size of $L(s,f)$ at the central point. The value of $L(s,f)$ at the central point is called the **central value** of $L(s,f)$. Many important properties about $L(s,f)$ can be connected to its central value. Any argument used to estimate the central value of an L-function is called a **central value estimate**. We will prove central value estimate which gives a very useful upper bound in the $q(f)$-aspect. To state it, let $\psi(y)$ be a bump function with compact support in $\left[\frac{1}{2},2\right]$. For example,

$$\psi(y) = \begin{cases} e^{-\frac{1}{9-(4y-5)^2}} & \text{if } |4y-5| < 3, \\ 0 & \text{if } |4y-5| \ge 3. \end{cases}$$

The theorem is the following:

Theorem 7.7.1. *Let $L(s,f)$ be an L-function and let $\psi(y)$ be a bump function with compact support in $\left[\frac{1}{2},2\right]$. Then we have*

$$L\left(\frac{1}{2},f\right) \ll_\varepsilon \max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left|\frac{A_\psi(X)}{q(f)^{\frac{1}{4}}}\right|\right) + \left|\frac{S}{q(f)^{\frac{1}{4}}}\right|,$$

where S is zero if $\Lambda(s,f)$ is entire, and otherwise

$$S = \left(\operatorname*{Res}_{u=\frac{1}{2}} + \operatorname*{Res}_{u=-\frac{1}{2}}\right) \Lambda\left(\frac{1}{2}+u,f\right).$$

Proof. Taking $s=\frac{1}{2}$, $X=1$, and $\Phi(u)=\cos^{-4dM}\left(\frac{\pi u}{4M}\right)$ with $M \gg 1$ in the approximate functional equation gives

$$L\left(\frac{1}{2},f\right) = \sum_{n\ge1}\frac{a_f(n)}{\sqrt{n}}V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) + \varepsilon(f)\sum_{n\ge1}\frac{\overline{a_f(n)}}{\sqrt{n}}V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) + \frac{R}{q(f)^{\frac{1}{4}}\gamma\left(\frac{1}{2},f\right)}.$$

This implies the bound

$$L\left(\frac{1}{2},f\right) \ll \left|\sum_{n\ge1}\frac{a_f(n)}{\sqrt{n}}V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right)\right| + \left|\frac{S}{q(f)^{\frac{1}{4}}}\right|.$$

Now consider the set of functions $\left\{\psi\left(\frac{y}{2^k}\right)\right\}_{k\in\mathbb{Z}}$. Since $\psi\left(\frac{y}{2^k}\right)$ has support in $[2^{k-1},2^{k+1}]$, the sum $\sigma(y)=\sum_{k\in\mathbb{Z}}\psi\left(\frac{y}{2^k}\right)$, defined for $y>0$, is finite since at most finitely many terms are nonzero for every y. It is also bounded away from zero since for any $y>0$ there is some $k\in\mathbb{Z}$ for which $2^k \le y \le 3\cdot 2^{k-1}$ so that $\frac{y}{2^k}$ is at least distance $\frac{1}{2}$ from the endpoints of $\left[\frac{1}{2},2\right]$. Defining $\psi_k(y)=\psi\left(\frac{y}{2^k}\right)\sigma(y)^{-1}$, it follows that $\{\psi_k(y)\}_{k\in\mathbb{Z}}$ satisfies

$$\sum_{k\in\mathbb{Z}}\psi_k(y)=1,$$

for any $y>0$. Then we can write

$$V_s(y)=\sum_{k\in\mathbb{Z}}\psi_k(y)V_s(y).$$

It follows that

$$\begin{aligned}
\sum_{n\ge1}\frac{a_f(n)}{\sqrt{n}}V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) &= \sum_{n\ge1}\frac{a_f(n)}{\sqrt{n}}\sum_{k\in\mathbb{Z}}\psi_k\left(\frac{n}{\sqrt{q(f)}}\right)V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) \\
&= \sum_{k\in\mathbb{Z}}\sum_{n\ge1}\frac{a_f(n)}{\sqrt{n}}\psi_k\left(\frac{n}{\sqrt{q(f)}}\right)V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) \qquad \text{FT} \\
&\ll_\varepsilon \sum_{k\in\mathbb{Z}}\left|\sum_{\substack{n\ll\mathfrak{q}(f)^{\frac{1}{2}+\varepsilon} \\ 2^{k-1}\sqrt{q(f)}\le n\le 2^{k+1}\sqrt{q(f)}}}\frac{a_f(n)}{\sqrt{n}}\psi_k\left(\frac{n}{\sqrt{q(f)}}\right)\right| \\
&\ll_\varepsilon \sum_{k\ll\log(q(f)^{\frac{\varepsilon}{2}})}\left|\sum_{\substack{n\ll\mathfrak{q}(f)^{\frac{1}{2}+\varepsilon} \\ 2^{k-1}\sqrt{q(f)}\le n\le 2^{k+1}\sqrt{q(f)}}}\frac{a_f(n)}{\sqrt{n}}\psi_k\left(\frac{n}{\sqrt{q(f)}}\right)\right|,
\end{aligned}$$

where in the second to last line we have used that $V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right)$ is bounded for $n \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2},f\right)^{\frac{1}{2}+\varepsilon} \ll_\varepsilon \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}$ and then exhibits polynomial decay thereafter by Proposition 7.4.2 and in the last line we have used that $\psi_k(y)$ has compact support in $[2^{k-1},2^{k+1}]$ (recall $k \in \mathbb{Z}$). Since $\sigma(y)$ is bounded away from zero and $\log(y) \ll y$, we obtain the crude bound

$$\sum_{k \ll \log(q(f)^{\frac{\varepsilon}{2}})} \left| \sum_{\substack{n \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon} \\ 2^{k-1}\sqrt{q(f)} \le n \le 2^{k+1}\sqrt{q(f)}}} \frac{a_f(n)}{\sqrt{n}} \psi_k\left(\frac{n}{\sqrt{q(f)}}\right) \right| \ll_\varepsilon q(f)^{\frac{\varepsilon}{2}} \max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left| \sum_{\frac{X}{2} \le n \le 2X} \frac{a_f(n)}{\sqrt{n}} \psi\left(\frac{n}{X}\right) \right| \right).$$

We will estimate this latter sum. Abel's summation formula (see Appendix B.3) gives

$$\sum_{\frac{X}{2} \le n \le 2X} \frac{a_f(n)}{\sqrt{n}} \psi\left(\frac{n}{X}\right) = \frac{A_\psi(2X)}{\sqrt{2X}} - \frac{A_\psi\left(\frac{X}{2}\right)}{\sqrt{\frac{X}{2}}} + \frac{1}{2}\int_{\frac{X}{2}}^{2X} A_\psi(u) u^{-\frac{3}{2}}\, du.$$

But as

$$\frac{1}{2}\int_{\frac{X}{2}}^{2X} A_\psi(u) u^{-\frac{3}{2}}\, du \ll X \max_{\frac{X}{2} \le u \le 2X} \left(\left| A_\psi(u) u^{-\frac{3}{2}} \right| \right) \ll \max_{\frac{X}{2} \le u \le 2X} \left(\left| \frac{A_\psi(u)}{\sqrt{u}} \right| \right),$$

we obtain the bound

$$\max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left| \sum_{\frac{X}{2} \le n \le 2X} \frac{a_f(n)}{\sqrt{n}} \psi\left(\frac{n}{X}\right) \right| \right) \ll \max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left| \frac{A_\psi(X)}{\sqrt{X}} \right| \right).$$

Putting everything together gives

$$\sum_{n \ge 1} \frac{a_f(n)}{\sqrt{n}} V_{\frac{1}{2}}\left(\frac{n}{\sqrt{q(f)}}\right) \ll_\varepsilon q(f)^{\frac{\varepsilon}{2}} \max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left| \frac{A_\psi(X)}{\sqrt{X}} \right| \right) \ll_\varepsilon \max_{X \ll \mathfrak{q}(f)^{\frac{1}{2}+\varepsilon}} \left(\left| \frac{A_\psi(X)}{q(f)^{\frac{1}{4}}} \right| \right),$$

where in the last estimate we may replace X with $q(f)^{\frac{1}{2}+\varepsilon}$ because $X \ge 1$ so that X is bounded away from zero. $\square$

7.8 Logarithmic Derivatives

There is an incredibly useful formula for the logarithmic derivative of any L-function which is often the starting point for deeper analytic investigations. To deduce it, we will need a more complete understanding of $\Lambda(s,f)$. First observe that the zeros ρ of $\Lambda(s,f)$ are contained inside of the critical strip. Indeed, we have already remarked that $L(s,f)$ has no zeros for $\sigma > 0$ and clearly $\gamma(s,f)$ does not have zeros in this region as well. Therefore $\Lambda(s,f)$ is nonzero for $\sigma > 1$. By the functional equation, $\Lambda(s,f)$ is also nonzero for $\sigma < 0$ too. In other words, the zeros of $\Lambda(s,f)$ are the nontrivial zeros of $L(s,f)$. Before we state our result, we setup some notation. For an L-function $L(s,f)$ we define

$$\xi(s,f) = (s(1-s))^{r_f}\Lambda(s,f).$$

Note that $\xi(s,f)$ is essentially just $\Lambda(s,f)$ with the potential poles at $s=0$ and $s=1$ removed. From the functional equation, we also have

$$\xi(s,f) = \varepsilon(f)\xi(1-s,\overline{f}).$$

We now state our desired result:

Proposition 7.8.1. *For any L-function $L(s,f)$, there exist constants $A(f)$ and $B(f)$ such that*

$$\xi(s,f) = e^{A(f)+B(f)s} \prod_{\rho \neq 0,1} \left(1 - \frac{s}{\rho}\right) e^{\frac{s}{\rho}},$$

and hence the sum

$$\sum_{\rho \neq 0,1} \frac{1}{|\rho|^{1+\varepsilon}},$$

is convergent provided the product and sum are both counted with multiplicity and ordered with respect to the size of the ordinate. Moreover,

$$-\frac{L'}{L}(s,f) = \frac{r_f}{s} + \frac{r_f}{s-1} + \frac{1}{2}\log q(f) + \frac{\gamma'}{\gamma}(s,f) - B(f) - \sum_{\rho \neq 0,1} \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right).$$

Proof. For the first statement, observe that $\xi(s,f)$ is entire since the only possible poles of $\Lambda(s,f)$ are at $s=0$ and $s=1$ and are of order r_f. We also claim that $\xi(s,f)$ is of order 1. By the functional equation, it suffices to show this for $\sigma \geq \frac{1}{2}$. This follows from $L(s,f)$ being of order 1 and Equation (7.8) (within ε of the poles of $\gamma(s,f)$ we know $\xi(s,f)$ is bounded because it is entire). By the Hadamard factorization theorem (see Appendix B.4),

$$\xi(s,f) = e^{A(f)+B(f)s} \prod_{\rho \neq 0,1} \left(1 - \frac{s}{\rho}\right) e^{\frac{s}{\rho}},$$

for some constants $A(f)$ and $B(f)$ and the desired sum converges. This proves the first statement. For the second, taking the logarithmic derivative of the definition of $\xi(s,f)$ yields

$$\frac{\xi'}{\xi}(s,f) = \frac{r_f}{s} + \frac{r_f}{s-1} + \frac{1}{2}\log q(f) + \frac{\gamma'}{\gamma}(s,f) + \frac{L'}{L}(s,f). \tag{7.15}$$

On the other hand, taking the logarithmic derivative of the Hadamard factorization gives

$$\frac{\xi'}{\xi}(s,f) = B(f) + \sum_{\rho \neq 0,1} \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right). \tag{7.16}$$

Equating Equations (7.15) and (7.16), we arrive at

$$B(f) + \sum_{\rho \neq 0,1} \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right) = \frac{r_f}{s} + \frac{r_f}{s-1} + \frac{1}{2}\log q(f) + \frac{\gamma'}{\gamma}(s,f) + \frac{L'}{L}(s,f).$$

Isolating $-\frac{L'}{L}(s,f)$ completes the proof. □

We now need to make a few comments. Our first is regarding the constants $A(f)$ and $B(f)$. The explicit evaluation of these constants can be challenging and heavily depends upon the arithmetic object f. However, useful estimates are not too difficult to obtain. We also claim $A(\overline{f}) = \overline{A(f)}$ and $B(\overline{f}) = \overline{B(f)}$. To see this, recall that $L(\overline{s},\overline{f}) = \overline{L(s,f)}$. Then $\xi(\overline{s},\overline{f}) = \overline{\xi(s,f)}$ because $\gamma(s,\overline{f}) = \gamma(s,f)$, the κ_j are real or occur in conjugate pairs, and $\Gamma(\overline{s}) = \overline{\Gamma(s)}$. But then the functional equation and Proposition 7.8.1 together imply

$$e^{A(\overline{f})+B(\overline{f})s} = \frac{\xi'}{\xi}(0,\overline{f}) = \overline{\frac{\xi'}{\xi}(0,f)} = e^{\overline{A(f)}+\overline{B(f)}s},$$

and the claim follows. Our second comment concerns the negative logarithmic derivative of $L(s,f)$. In general, this function attracts much attention for analytic investigations. As $L(s,f)$ is holomorphic for $\sigma>1$ and admits an Euler product there, we can take the logarithm of the Euler product (turning it into a sum) and differentiate termwise to obtain

$$-\frac{L'}{L}(s,f) = -\sum_{p}\sum_{1\le j\le d_f}\frac{d}{ds}\log(1-\alpha_j(p)p^{-s}) = \sum_{p}\sum_{1\le j\le d_f}\frac{\alpha_j(p)\log(p)}{(1-\alpha_j(p)p^{-s})p^s}. \tag{7.17}$$

From the Taylor series of $\frac{1}{1-s}$, it follows that $-\frac{L'}{L}(s,f)$ is a locally absolutely uniformly convergent Dirichlet series of the form

$$-\frac{L'}{L}(s,f) = \sum_{n\ge1}\frac{\Lambda_f(n)}{n^s},$$

for $\sigma>1$, where

$$\Lambda_f(n) = \begin{cases}\sum_{1\le j\le d_f}\alpha_j(p)^k\log(p) & \text{if } n=p^k \text{ for some } k\ge1,\\ 0 & \text{otherwise.}\end{cases}$$

It is worth noting that $\Lambda_{\overline{f}}(n)=\overline{\Lambda_f(n)}$.

7.9 Zero Density

The deepest subject of the theory of L-functions is arguably the distribution of the zeros of L-functions. Here we introduce a method of counting zeros of L-functions and gaining a very simple understanding of their density as a result. We first require an immensely useful lemma:

Lemma 7.9.1. *Let $L(s,f)$ be an L-function. The following statements hold:*

(i) The constant $B(f)$ satisfies

$$\mathrm{Re}(B(f)) = -\sum_{\rho\ne0,1}\mathrm{Re}\left(\frac{1}{\rho}\right),$$

where the sum is counted with multiplicity and ordered with respect to the size of the ordinate.

(ii) For any $T\ge0$, the number of nontrivial zeros $\rho=\beta+i\gamma$ with $\rho\ne0,1$ and such that $|T-\gamma|\le1$ is $O(\log\mathfrak{q}(iT,f))$.

(iii) For $\sigma>1$, we have

$$\mathrm{Re}\left(\frac{1}{s-\rho}\right)>0 \quad \text{and} \quad \mathrm{Re}\left(\frac{1}{s+\kappa_j}\right)>0.$$

(iv) We have

$$-\frac{L'}{L}(s,f) = \frac{r_f}{s}+\frac{r_f}{s-1}-\sum_{|s+\kappa_j|<1}\frac{1}{s+\kappa_j}-\sum_{\substack{|s-\rho|<1\\ \rho\ne0,1}}\frac{1}{s-\rho}+O(\log\mathfrak{q}(s,f)),$$

for any s in the vertical strip $-\frac{1}{2}\le\sigma\le2$. Moreover, in this vertical strip we also have

$$\mathrm{Re}\left(-\frac{L'}{L}(s,f)\right)\le\mathrm{Re}\left(\frac{r_f}{s}\right)+\mathrm{Re}\left(\frac{r_f}{s-1}\right)-\sum_{|s+\kappa_j|<1}\mathrm{Re}\left(\frac{1}{s+\kappa_j}\right)-\sum_{\substack{|s-\rho|<1\\ \rho\ne0,1}}\mathrm{Re}\left(\frac{1}{s-\rho}\right)+O(\log\mathfrak{q}(s,f)),$$

and we may discard any term in either sum if $1<\sigma\le2$.

Proof. We will prove each statement separately.

(i) To prove (i), first recall that $B(\overline{f}) = \overline{B(f)}$. Then Equation (7.16) and the functional equation together imply

$$2\mathrm{Re}(B(f)) = B(f) + B(\overline{f}) = -\sum_{\rho \neq 0,1} \left(\frac{1}{s-\rho} + \frac{1}{1-s-\overline{\rho}} + \frac{1}{\rho} + \frac{1}{\overline{\rho}} \right),$$

where we have made use of the fact that the nontrivial zeros occur is pairs ρ and $1-\overline{\rho}$ where the latter is also a nontrivial zero of $L(1-s,\overline{f})$. Now fix s such that it does not coincide with the ordinate of a nontrivial zero. Then s is bounded away from all of the nontrivial zeros and it follows that $\frac{1}{(s-\rho)} + \frac{1}{(1-s-\overline{\rho})} \ll \frac{1}{\rho^2}$ and $\frac{1}{\rho} + \frac{1}{\overline{\rho}} \ll \frac{1}{\rho^2}$. Therefore the sums

$$\sum_{\rho \neq 0,1} \left(\frac{1}{s-\rho} + \frac{1}{1-s-\overline{\rho}} \right) \quad \text{and} \quad \sum_{\rho \neq 0,1} \left(\frac{1}{\rho} + \frac{1}{\overline{\rho}} \right),$$

converge absolutely by Proposition 7.8.1 and so we can sum them separately. The first sum vanishes by again using the fact that the nontrivial zeros occur is pairs ρ and $1-\overline{\rho}$. Thus

$$2\mathrm{Re}(B(f)) = \sum_{\rho \neq 0,1} \left(\frac{1}{\rho} + \frac{1}{\overline{\rho}} \right) = -\sum_{\rho \neq 0,1} \mathrm{Re}\left(\frac{1}{\rho} \right),$$

which gives (i).

(ii) For (ii), we first bound two important quantities. For the first quantity, the definition of $\Lambda_f(n)$ and that $|\alpha_j(p)| \le p$ together imply the weak bound $|\Lambda_f(n)| \le d_f n \log(n)$. Then

$$\frac{L'}{L}(s,f) \ll d_f \zeta'(s-1) \ll \log \mathfrak{q}(f), \tag{7.18}$$

provided $\sigma > 2$. For the second quantity, Equation (7.10) implies

$$\frac{1}{2}\log q(f) + \frac{\gamma'}{\gamma}(s,f) \ll \log \mathfrak{q}(s,f), \tag{7.19}$$

for $\sigma > 0$. Now fix $T \ge 0$ and let $s = 3 + iT$. Taking the real part of the formula for the negative logarithmic derivative in Proposition 7.8.1 and combining Equations (7.18) and (7.19) with (i) results in

$$\sum_{\rho \neq 0,1} \mathrm{Re}\left(\frac{1}{s-\rho} \right) \ll \log \mathfrak{q}(iT,f).$$

But as

$$\frac{2}{9+(T-\gamma)^2} \le \mathrm{Re}\left(\frac{1}{s-\rho} \right) \le \frac{3}{4+(T-\gamma)^2},$$

we obtain

$$\sum_{\rho \neq 0,1} \frac{1}{1+(T-\gamma)^2} \ll \log \mathfrak{q}(iT,f), \tag{7.20}$$

which is stronger than the first statement of (ii) since all of the terms in the sum are positive. The second statement is also clear.

(iii) For (iii), just observe that

$$\operatorname{Re}\left(\frac{1}{s-\rho}\right) = \frac{\sigma-\beta}{(\sigma-\beta)^2+(t-g)^2} > 0 \quad \text{and} \quad \operatorname{Re}\left(\frac{1}{s+\kappa_j}\right) = \frac{\sigma+\operatorname{Re}(\kappa_j)}{(\sigma+\operatorname{Re}(\kappa_j))^2+(t+\operatorname{Im}(\kappa_j))^2} > 0,$$

where the first bound holds because $\beta \le 1$ and the second bound holds because $\operatorname{Re}(\kappa_j) > -1$.

(iv) To deduce (iv), let s be such that $-\frac{1}{2} \le \sigma \le 2$. Using Equation (7.18), we can write

$$-\frac{L'}{L}(s,f) = -\frac{L'}{L}(s,f) + \frac{L'}{L}(3+it,f) + O(\log \mathfrak{q}(s,f)).$$

Applying the formula for the negative logarithmic derivative in Proposition 7.8.1 to the two terms on the right-hand side and using Equation (7.10) for $\sigma > 0$, we get

$$-\frac{L'}{L}(s,f) = \frac{r_f}{s} + \frac{r_f}{s-1} + \frac{\gamma'}{\gamma}(s,f) - \sum_{\rho \ne 0,1}\left(\frac{1}{s-\rho} - \frac{1}{3+it-\rho}\right) + O(\log \mathfrak{q}(s,f)).$$

We now estimate the remaining sum. Retain the first part of the terms for which $|s-\rho| < 1$. The contribution from the second part of these terms is $O(\log \mathfrak{q}(it,f))$ by (ii). For those terms with $|s-\rho| \ge 1$, we have

$$\left|\frac{1}{s-\rho} - \frac{1}{3+it-\rho}\right| \le \frac{3-\sigma}{(3-\beta)^2+(t-\gamma)^2} \le \frac{3}{1+(t-\gamma)^2}.$$

Therefore from Equation (7.20), the contribution of these terms is $O(\log \mathfrak{q}(it,f))$ too. It follows that

$$-\frac{L'}{L}(s,f) = \frac{r_f}{s} + \frac{r_f}{s-1} + \frac{\gamma'}{\gamma}(s,f) - \sum_{\substack{|s-\rho|<1 \\ \rho \ne 0,1}} \frac{1}{s-\rho} + O(\log \mathfrak{q}(s,f)).$$

Applying $\frac{\Gamma'}{\Gamma}(s+1) = \frac{\Gamma'}{\Gamma}(s) + \frac{1}{s}$ to $\frac{\gamma'}{\gamma}(s,f)$ and using Equation (7.10) gives

$$\frac{\gamma'}{\gamma}(s,f) = -\sum_{|s+\kappa_j|<1} \frac{1}{s+\kappa_j} + O(\log \mathfrak{q}_\infty(s,f)).$$

Then we obtain

$$-\frac{L'}{L}(s,f) = \frac{r_f}{s} + \frac{r_f}{s-1} - \sum_{|s+\kappa_j|<1} \frac{1}{s+\kappa_j} - \sum_{\substack{|s-\rho|<1 \\ \rho \ne 0,1}} \frac{1}{s-\rho} + O(\log \mathfrak{q}(s,f)),$$

which is the first statement of (iv). For second statement, tak the real part of this estimate and write it as

$$\sum_{|s+\kappa_j|<1} \operatorname{Re}\left(\frac{1}{s+\kappa_j}\right) + \sum_{\substack{|s-\rho|<1 \\ \rho \ne 0,1}} \operatorname{Re}\left(\frac{1}{s-\rho}\right) \le \operatorname{Re}\left(-\frac{L'}{L}(s,f)\right) + \operatorname{Re}\left(\frac{r_f}{s}\right) + \operatorname{Re}\left(\frac{r_f}{s-1}\right) + O(\log \mathfrak{q}(s,f)).$$

Now observe that we can discard any of the terms in either sum provided $1 < \sigma \le 2$ by (iii). $\square$

With Lemma 7.9.1 in hand, we can deduce a result which estimates the number of nontrivial zeros in a box. Accordingly, for any $T \geq 0$ we define

$$N(T,f) = |\{\rho = \beta + i\gamma \in \mathbb{C} : L(\rho,f) = 0 \text{ with } 0 \leq \beta \leq 1 \text{ and } |\gamma| \leq T\}|.$$

In other words, $N(T,f)$ is the number of nontrivial zeros of $L(s,f)$ with ordinate in $[-T,T]$. We will prove the following:

Theorem 7.9.1. *For any L-function $L(s,f)$ and $T \geq 1$,*

$$N(T,f) = \frac{T}{\pi}\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right) + O(\log \mathfrak{q}(iT,f)).$$

Proof. Let $T \geq 1$ and set

$$N'(T,f) = |\{\rho = \beta + i\gamma \in \mathbb{C} : L(\rho,f) = 0 \text{ with } 0 \leq \beta \leq 1 \text{ and } 0 < \gamma \leq T\}|.$$

As the nontrivial zeros occur in pairs ρ and $1-\overline{\rho}$ where the latter is also a nontrivial zero of $L(s,\overline{f})$, it follows that

$$N(T,f) = N'(T,f) + N'(T,\overline{f}) + O(\log \mathfrak{q}(f)),$$

where $O(\log \mathfrak{q}(f))$ accounts for the possible real nontrivial zeros. There are finitely many such nontrivial zeros because the interval $0 \leq s \leq 1$ is compact. We will estimate $N'(T,f)$ and in doing so we may assume $L(s,f)$ does not vanish on the line $t = T$ by varying T by a sufficiently small constant, if necessary, and observing that $N(T,f)$ is modified by a quantity of size $O(\log \mathfrak{q}(iT,f))$ by Lemma 7.9.1 (i). Since the nontrivial zeros are isolated, let $\delta > 0$ be small enough such that $\Lambda(s,f)$ has no nontrivial zeros for $-\delta \leq t < 0$. Then by our previous comments and the argument principle,

$$N'(T,f) = \frac{1}{2\pi i}\int_{\eta}\frac{\xi'}{\xi}(s,f)\,ds + O(\log \mathfrak{q}(iT,f)),$$

where $\eta = \sum_{1\leq i\leq 6}\eta_i$ is the contour in Figure 7.3:
Since $\log(s) = \log|s| + i\arg(s)$, we have

$$\frac{1}{2\pi i}\int_{\eta}\frac{\xi'}{\xi}(s,f)\,ds = \frac{1}{2\pi i}\int_{\eta}\frac{d}{ds}\log|\xi(s,f)|\,ds + \frac{1}{2\pi}\int_{\eta}\frac{d}{ds}\arg\xi(s,f)\,ds = \frac{1}{2\pi}\Delta_{\eta}\arg(\xi(s,f)),$$

where the last equality holds by parameterizing the curve η and noting that η is closed so that the first integral vanishes. For convenience, set $\eta_L = \eta_1+\eta_2+\eta_3$ and $\eta_R = \eta_4+\eta_5+\eta_6$. Recall that we have already shown $\xi(\overline{s},\overline{f}) = \overline{(\xi(s,f))}$. Using this fact along with the functional equation and that $-\arg(s) = \arg(\overline{s})$, we compute

$$\begin{aligned}
\Delta_{\eta_L}\arg(\xi(s,f)) &= \Delta_{\eta_L}\arg\left(\varepsilon(f)\xi(1-s,\overline{f})\right)\\
&= -\Delta_{\overline{\eta_R}}\arg\left(\varepsilon(f)\xi(s,\overline{f})\right)\\
&= \Delta_{\overline{\eta_R}}\arg\left(\overline{\varepsilon(f)}\xi(\overline{s},f)\right)\\
&= \Delta_{\eta_R}\arg\left(\overline{\varepsilon(f)}\xi(s,f)\right)\\
&= \Delta_{\eta_R}\arg\left(\overline{\varepsilon(f)}\right) + \Delta_{\eta_R}\arg(\xi(s,f))\\
&= \Delta_{\eta_R}\arg(\xi(s,f)).
\end{aligned}$$

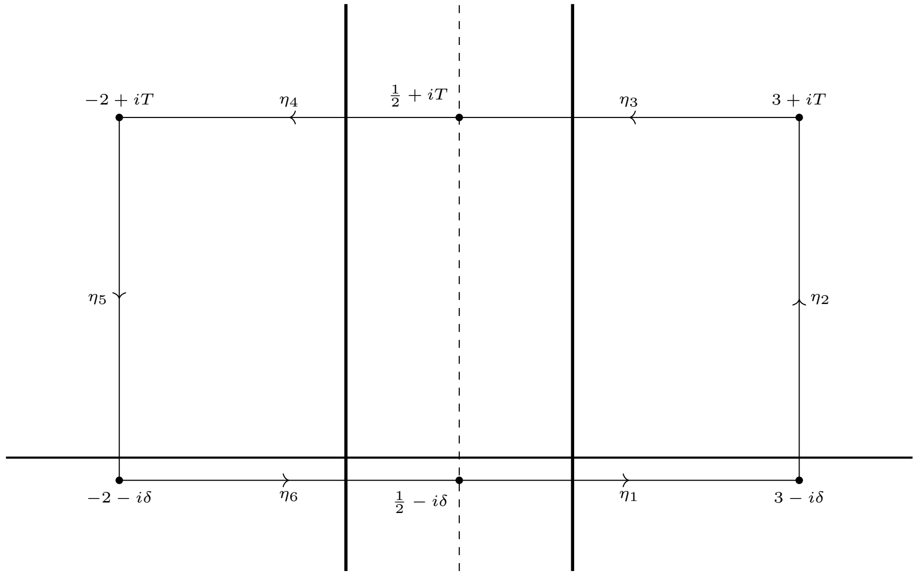

Figure 7.3: A zero counting contour.

In other words, the change in argument along η_L is equal to the change in argument along η_R and so

$$\frac{1}{2\pi i}\int_{\eta}\frac{\xi'}{\xi}(s,f)\,ds=\frac{1}{\pi}\Delta_{\eta_L}\arg\xi(s,f).$$

Thus to estimate the integral, we estimate the change in argument along η_L of each factor in

$$\xi(s,f)=(s(1-s))^{r_f}q(f)^{\frac{s}{2}}\pi^{-\frac{d_f s}{2}}\prod_{1\le j\le d_f}\Gamma\left(\frac{s+\kappa_j}{2}\right)L(s,f).$$

For the factor $(s(1-s))^{r_f}$, we first have

$$\Delta_{\eta_L}\arg(s)=\arg(s)\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT}=\arg\left(\frac{1}{2}+iT\right)-\arg\left(\frac{1}{2}-i\delta\right)=O\left(\frac{1}{T}\right),$$

and

$$\Delta_{\eta_L}\arg(1-s)=\arg(1-s)\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT}=\arg\left(\frac{1}{2}-iT\right)-\arg\left(\frac{1}{2}+i\delta\right)=O\left(\frac{1}{T}\right),$$

where in both computations we have used $\arg(s)=\tan^{-1}\left(\frac{t}{\sigma}\right)=\frac{\pi}{2}+O\left(\frac{1}{t}\right)$, provided $\sigma>0$, which holds by the Laurent series of the inverse tangent. Combining these two bounds, we obtain

$$\Delta_{\eta_L}(s(1-s))^{r_f}=O\left(\frac{1}{T}\right). \tag{7.21}$$

For the factor $q(f)^{\frac{s}{2}}$, we use that $\arg(s)=\mathrm{Im}(\log(s))$ and compute

$$\Delta_{\eta_L}\arg q(f)^{\frac{s}{2}}=\mathrm{Im}(\log q(f)^{\frac{s}{2}})\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT}=\log q(f)\left(\frac{T}{2}+\frac{\delta}{2}\right)=\frac{T}{2}\log q(f)+O(1). \tag{7.22}$$

For the factor $\pi^{-\frac{d_f s}{2}}$, we use that $\arg(s) = \mathrm{Im}(\log(s))$ and compute

$$\Delta_{\eta_L} \arg(\pi^{-\frac{d_f s}{2}}) = \mathrm{Im}(\log(\pi^{-\frac{d_f s}{2}}))\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT} = \log\left(\frac{1}{\pi^{d_f}}\right)\left(\frac{T}{2}+\frac{\delta}{2}\right) = \frac{T}{2}\log\left(\frac{1}{\pi^{d_f}}\right) + O(1). \tag{7.23}$$

For the factor $\prod_{1\le j\le d_f}\Gamma\left(\frac{s+\kappa_j}{2}\right)$, we first use Equation (1.5) (valid since $T \geq 1$) and that $\arg(s) = \mathrm{Im}(\log(s))$ to obtain

$$\begin{aligned}
\Delta_{\eta_L} \arg\Gamma(s) &= \mathrm{Im}(\log\Gamma(s))\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT} \\
&= T\log\left|\frac{1}{2}+iT\right| - T + \delta\log\left|\frac{1}{2}+i\delta\right| - \delta + O(1) \\
&= T\log(T) - T + O(1) \\
&= T\log\left(\frac{T}{e}\right) + O(1).
\end{aligned}$$

It follows that

$$\Delta_{\eta_L} \arg\left(\prod_{1\le j\le d_f}\Gamma\left(\frac{s+\kappa_j}{2}\right)\right) = \frac{T}{2}\log\left(\frac{T^{d_f}}{(2e)^{d_f}}\right) + O(\log\mathfrak{q}(f)). \tag{7.24}$$

For the factor $L(s,f)$, note that

$$\Delta_{\eta_L} \arg(L(s,f)) = \mathrm{Im}(\log L(s,f))\Big|_{\frac{1}{2}-i\delta}^{\frac{1}{2}+iT} = \mathrm{Im}\left(\int_{\eta_L}\frac{L'}{L}(s,f)\,ds\right).$$

By Equation (7.18), the integral is $O(\log q(f))$ on η_2. On η_1 and η_3, Lemma 7.9.1 (ii) and (iv) together imply that the integral is $O(\log\mathfrak{q}(iT,f))$. It follows that

$$\Delta_{\eta_L} \arg(L(s,f)) = O(\log\mathfrak{q}(iT,f)). \tag{7.25}$$

Combining Equations (7.21) to (7.25) results in

$$\frac{1}{\pi}\Delta_{\eta_L} \arg\xi(s,f) = \frac{T}{2\pi}\log\left(\frac{T^{d_f}}{(2\pi e)^{d_f}}\right) + O(\log\mathfrak{q}(iT,f)),$$

and therefore

$$N'(T,f) = \frac{T}{2\pi}\log\left(\frac{T^{d_f}}{(2\pi e)^{d_f}}\right) + O(\log\mathfrak{q}(iT,f)),$$

The claim follows immediately from this estimate and that the same exact estimate holds for $N'(T,\overline{f})$ by using the L-function $L(s,\overline{f})$. $\square$

It is worth noting that the main term in the proof of Theorem 7.9.1 comes from the change in argument of $q(f)^{\frac{s}{2}}\gamma(s,f)$ along the vertical segment η_3 (equivalently η_4). Moreover, the contribution from $L(s,f)$ is only to the error term. This is a good example of how analytic information of an L-function is intrinsically connected to its gamma factor. Also, with Theorem 7.9.1 we can now derive our zero density estimate:

Corollary 7.9.1. *For and L-function $L(s,f)$ and $T \geq 1$,*

$$\frac{N(T,f)}{T} \sim \frac{1}{\pi}\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right).$$

Proof. From Theorem 7.9.1,

$$\frac{N(T,f)}{T} = \frac{1}{\pi}\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right)\left(1+O\left(\frac{\log \mathfrak{q}(iT,f)}{\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right)T}\right)\right) = \frac{1}{\pi}\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right)\left(1+O\left(\frac{1}{T}\right)\right).$$

Since $O\left(\frac{1}{T}\right) = o(1)$, the result follows. □

Corollary 7.9.1 can be interpreted as saying that for large T the density of $N(T,f)$ is approximately $\frac{1}{\pi}\log\left(\frac{q(f)T^{d_f}}{(2\pi e)^{d_f}}\right)$. Since this grows as $T \to \infty$, we see that the nontrivial zeros tend to accumulate farther up the critical strip with logarithmic growth. We can dispense with this accumulation. If $\rho = \beta + i\gamma$ is a nontrivial zero of $L(s,f)$, then we call $\rho_{\text{unf}} = \beta + i\omega$ the **unfolded nontrivial zero** corresponding to ρ where

$$\omega = \frac{\gamma}{\pi}\log\left(\frac{q(f)|\gamma|^{d_f}}{(2\pi e)^{d_f}}\right).$$

Now for any $W \geq 0$, define

$$N_{\text{unf}}(W,f) = |\{\rho_{\text{unf}} = \beta + i\omega \in \mathbb{C} : L(\rho,f) = 0 \text{ with } 0 \leq \beta \leq 1 \text{ and } |\omega| \leq W\}|.$$

In other words, $N_{\text{unf}}(T,f)$ is the number of unfolded nontrivial zeros of $L(s,f)$ with ordinate in $[-W,W]$. We then have the following well-known result:

Proposition 7.9.1. *For any L-function $L(s,f)$ and $W \geq \frac{1}{\pi}\log\left(\frac{q(f)}{(2\pi e)^{d_f}}\right)$,*

$$\frac{N_{\mathit{unf}}(W,f)}{W} \sim 1.$$

Proof. Consider the function $f(t)$ defined by

$$f(t) = \frac{t}{\pi}\log\left(\frac{q(f)|t|^{d_f}}{(2\pi e)^{d_f}}\right),$$

for $t \in \mathbb{R}$. Sine $f(t)$ is a strictly increasing continuous function, it has an inverse $g(w)$ for $w \in \mathbb{R}$. It follows that $|\omega| \leq W$ if and only if $|\rho| \leq g(W)$ and so $N_{\text{unf}}(W,f) = N(g(W),f)$. But by Corollary 7.9.1 and that $g(w)$ is the inverse of $f(t)$, we have $N(g(W),f) \sim W$. It follows that

$$N_{\text{unf}}(W,f) \sim W,$$

which is equivalent to the claim. □

We interpret Proposition 7.9.1 as saying that the unfolded nontrivial zeros are evenly spaced opposed to Corollary 7.9.1 which says that they tend to accumulate up the critical strip.

7.10 A Zero-free Region

Although the Riemann hypothesis remains out of reach, some progress has been made to understand regions inside of the critical strip for which L-functions are nonzero except for possibly one real exception. Such regions are known as **zero-free regions** and there is great interest in improving the breadth of such regions. We will derive a standard zero-free region for any L-function under some mild assumptions. First a useful lemma:

Lemma 7.10.1. *Let $L(s,f)$ be an L-function such that $\mathrm{Re}(\Lambda_f(n)) \ge 0$ provided $(n,q(f)) = 1$. Also suppose that $|\alpha_j(p)| \le \frac{p}{2}$ for ramified primes p. Then $L(1,f) \ne 0$ and hence $r_f \ge 0$. Moreover, there exists a constant $c > 0$ such that $L(s,f)$ has at most r_f real zeros in the region*

$$\sigma \ge 1 - \frac{c}{d_f(r_f+1)\log \mathfrak{q}(f)}.$$

Proof. Let β_j be a real nontrivial zero with $\frac{1}{2} \le \beta_j \le 1$. There are finitely many β_j since they belong to the compact interval $\frac{1}{2} \le s \le 1$ so we have $1 \le j \le n$ for some $n \ge 1$. Letting $1 < \sigma \le 2$, and applying Lemma 7.9.1 (iv) while discarding all the terms except those corresponding to the nontrivial zeros β_j, we obtain the inequality

$$\sum_{1\le j\le n} \frac{1}{\sigma-\beta_j} < \frac{r_f}{\sigma-1} + \mathrm{Re}\left(\frac{L'}{L}(\sigma,f)\right) + O(\log \mathfrak{q}(f)).$$

To estimate $\mathrm{Re}\left(\frac{L'}{L}(\sigma,f)\right)$, we first note that as $\mathrm{Re}(\Lambda_f(n)) \ge 0$ provided $(n,q(f)) = 1$ by assumption, the Dirichlet series of $\frac{L'^{(q(f))}}{L^{(q(f))}}(s,f)$ shows that

$$\mathrm{Re}\left(\frac{L^{(q(f))\prime}}{L^{(q(f))}}(\sigma,f)\right) \le 0.$$

This gives an estimate for the contribution of the local factors of $L(s,f)$ corresponding to unramified primes. For the contribution of the local factors corresponding to ramified primes, we use Equation (7.17) to compute

$$\mathrm{Re}\left(\frac{L'_{q(f)}}{L_{q(f)}}(\sigma,f)\right) \le \left|\frac{L'_{q(f)}}{L_{q(f)}}(\sigma,f)\right| = \left|\sum_{p|q(f)}\sum_{1\le j\le d_f} \frac{\alpha_j(p)\log(p)}{(1-\alpha_j(p)p^{-\sigma})p^{\sigma}}\right| \le d_f \sum_{p|q(f)} \log(p) \le d_f \log(q(f)),$$

where in the second inequality we have made use of the assumption $|\alpha_j(p)| \le \frac{p}{2}$ for ramified primes p to conclude that $\left|\frac{\alpha_j(p)}{(1-\alpha_j(p)p^{-\sigma})p^{\sigma}}\right| \le 1$. These estimates together imply

$$\sum_{1\le j\le n} \frac{1}{\sigma-\beta_j} < \frac{r_f}{\sigma-1} + O(\log d_f \mathfrak{q}(f)).$$

From this inequality we see that $\beta_j \ne 1$. For if some $\beta_j = 1$, $r_f < 0$ and the right-hand side is negative for σ sufficiently close to 1 contradicting the positivity of the left-hand side. Thus $L(1,f) \ne 0$ and hence $r_f \ge 0$. As there are finitely many β_j, there exists a $c > 0$ such that the β_j satisfy

$$\beta_j \ge 1 - \frac{c}{d_f(r_f+1)\log \mathfrak{q}(f)}.$$

Setting $\sigma = 1 + \frac{2c}{d_f \log \mathfrak{q}(f)}$ and choosing c smaller, if necessary, we guarantee $1 < \sigma \le 2$. Then the two inequalities above together imply

$$\frac{n d_f \log \mathfrak{q}(f)}{2c + \frac{c}{(r_f+1)}} < \left(\frac{r_f}{2c} + O(1)\right) d_f \log \mathfrak{q}(f).$$

Isolating n, we see that

$$n < r_f + \frac{r_f}{2(r_f+1)} + O(c),$$

and taking c smaller, if necessary, we have $n \le r_f$. As $L(s,f)$ is nonzero for $\sigma > 1$, it follows that there are at most r_f real zeros satisfying

$$\sigma \ge 1 - \frac{c}{d_f(r_f+1)\log \mathfrak{q}(f)}.$$

This completes the proof. □

We can now prove our zero-free region result:

Theorem 7.10.1. *Let $L(s,f)$ be an L-function with at most a simple pole at $s=1$, $\mathrm{Re}(\Lambda_f(n)) \ge 0$ provided $(n,q(f))=1$, and $|\alpha_j(p)| \le \frac{p}{2}$ for ramified primes p. Then there exists a constant $c>0$ such that $L(s,f)$ has no zeros in the region*

$$\sigma \ge 1 - \frac{c}{d_f^2 \log(\mathfrak{q}(f)(|t|+3))},$$

except for possibly one simple real zero β_f with $\beta_f < 1$ in the case $L(s,f)$ has a simple pole at $s=1$.

Proof. For $t \in \mathbb{R}$, let $L(s,g)$ be the L-function defined by

$$L(s,g) = L^3(s,f)L^3(s,\overline{f})L^4(s+it,f)L^4(s+it,\overline{f})L(s+2it,f)L(s+2it,\overline{f}).$$

Clearly $d_g = 16d_f$ and so $\mathfrak{q}(g)$ satisfies

$$\mathfrak{q}(g) \le \mathfrak{q}(f)^6\mathfrak{q}(it,f)^8\mathfrak{q}(2it,f)^2 \le \mathfrak{q}(f)^{16d_f}(|t|+3)^{10d_f} < (\mathfrak{q}(f)(|t|+3))^{16d_f}.$$

We claim that $\mathrm{Re}(\Lambda_g(n)) \ge 0$ for $(n,q(f))=1$. To see this, let p be an unramified prime. The local roots of $L(s,g)$ at p are $\alpha_j(p)$ and $\overline{\alpha_j(p)}$ both with multiplicity three, $\alpha_j(p)p^{-it}$ and $\overline{\alpha_j(p)}p^{-it}$ both with multiplicity four, and $\alpha_j(p)p^{-2it}$ and $\overline{\alpha_j(p)}p^{-2it}$ both with multiplicity one. So for any $k \ge 1$, the sum of k-th powers of these local roots is

$$\sum_{1\le j\le d_f}\left(6\mathrm{Re}(\alpha_j(p)^k) + 8\mathrm{Re}(\alpha_j(p)^k)p^{-kit} + 2\mathrm{Re}(\alpha_j(p)^k)p^{-2kit}\right).$$

The real part of this expression is

$$(6+8\cos\log(p^{kt})+2\cos\log(p^{2kt}))\mathrm{Re}(\Lambda_f(p^k)) = 4(1+\cos\log(p^{kt}))^2\mathrm{Re}(\Lambda_f(p^k)) \ge 0.$$

where we have made use of the identity $3+4\cos(\theta)+\cos(2\theta) = 2(1+\cos(\theta))^2$. It follows that $\mathrm{Re}(\Lambda_g(n)) \ge 0$ for $(n,q(f))=1$. Therefore the conditions of Lemma 7.10.1 are satisfied for $L(s,g)$. Now let $\rho = \beta + i\gamma$ be a complex nontrivial zero of $L(s,f)$. Setting $t=\gamma$, $L(s,g)$ has a real nontrivial zero at $s=\beta$ of order at least 8 and a pole at $s=1$ of order 6. That is, $r_g = 6$. But Lemma 7.10.1 implies that $L(s,g)$ can have at most 6 real nontrivial zeros in the given region. Letting the constant for the region in Lemma 7.10.1 be c', it follows that β must satisfy

$$\beta < 1 - \frac{c'}{d_g(r_g+1)\log\mathfrak{q}(g)} < 1 - \frac{c'}{1792 d_f^2 \log(\mathfrak{q}(f)(|\gamma|+3))},$$

Take $c = \frac{c'}{1792}$. Now let β be a real nontrivial zero of $L(s,f)$. Since $\mathrm{Re}(\Lambda_f(n)) \ge 0$ and $L(s,f)$ has at most a simple pole at $s=1$, Lemma 7.10.1 implies, upon shrinking c if necessary, that there is at most one simple real zero β_f in the desired region and it can only occur if $L(s,f)$ has a simple pole at $s=1$. Note that if b_f exists, then $\beta_f < 1$ because $L(s,f)$ is nonzero for $\sigma > 1$ and has a simple pole at $s=1$. This completes the proof. □

Some comments are in order. Since the zeros of $L(s,f)$ occur in pairs ρ and $1-\overline{\rho}$, the zero-free region in Theorem 7.10.1 implies a symmetric zero-free region about the critical line with at most one real zero in each half as displayed in Figure 7.4. The possible simple zero b_f in Theorem 7.10.1 is referred to as a **Siegel zero** (or **exceptional zero**). If β_f is a Siegel zero of $L(s,\chi)$, then $1-\beta_f$ is also a real nontrivial zero of $L(s,\chi)$ since the nontrivial zeros occur in pairs (or sets of four in some cases we have discussed). This immediately implies that $\beta_f > \frac{1}{2}$ for we cannot have $\beta_f = \frac{1}{2}$ since the zero is simple and either β_f or $1-\beta_f$ must be at least $\frac{1}{2}$. Such zeros are conjectured to not exist. In full generality, Theorem 7.10.1 is the best result that one can hope for. It is possible to obtain better zero-free regions in certain cases but this heavily depends upon the particular L-function of interest and hence the arithmetic object f attached to $L(s,f)$. In many cases it is possible to augment the proof of Theorem 7.10.1 to make the constant c effective and such results have important applications. If $L(s,f)$ is of Selberg class, the conclusion of Theorem 7.10.1 is expected to hold without the possibly of a Siegel zero since $L(s,f)$ is expected to satisfy the Riemann hypothesis. Nevertheless, we can often satisfy the assumptions of Theorem 7.10.1. Indeed, if $\alpha_i(p) \geq 0$ then the assumptions are satisfied immediately. If not, since $|\alpha_i(p)| \leq 1$ the assumptions will hold for $\zeta(s)L(s,f)$ provided $L(s,f)$ does not have a pole at $s=1$. This is often possible to prove in practice since $L(s,f)$ is expected to be entire unless $\alpha_i(p) \geq 0$.

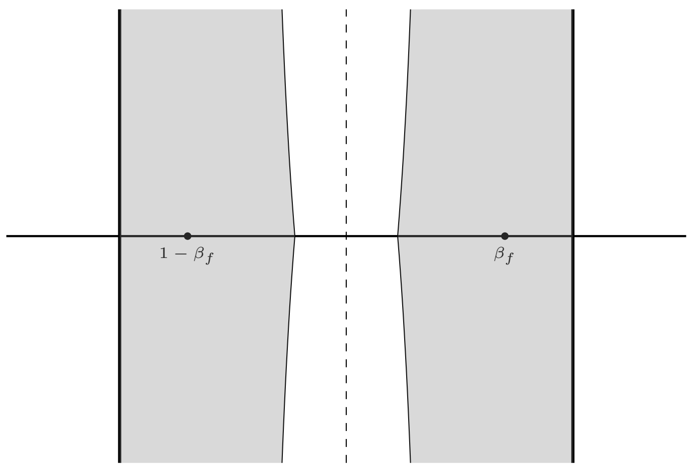

Figure 7.4: The symmetric zero-free region in Theorem 7.10.1.

7.11 The Explicit Formula

A formula somewhat analogous to the approximate functional equation can be derived for the negative logarithmic derivative of any L-function. This formula is called the the **explicit formula**:

Theorem 7.11.1 (Explict Formula). *Let $\psi(y)$ be a bump function with compact support and let $\Psi(s)$ denote its Mellin transform. Set $\phi(y) = y^{-1}\psi(y^{-1})$ so that its Mellin transform satisfies $\Phi(s) = \Psi(1-s)$. Then for any L-function $L(s,f)$, we have*

$$\sum_{n\geq 1}\big(\Lambda_f(n)\psi(n) + \Lambda_{\overline{f}}(n)\phi(n)\big) = \psi(1)\log q(f) + r_f\Psi(1) + \frac{1}{2\pi i}\int_{(\frac{1}{2})}\left(\frac{\gamma'}{\gamma}(s,f) + \frac{\gamma'}{\gamma}(1-s,\overline{f})\right)\Psi(s)\,ds - \sum_{\rho}\Psi(\rho),$$

where the sum is counted with multiplicity and ordered with respect to the size of the ordinate.

Proof. By smoothed Perron's formula, we have

$$\sum_{n\geq 1}\Lambda_f(n)\psi(n)=\frac{1}{2\pi i}\int_{(c)}-\frac{L'}{L}(s,f)\Phi(s)\,ds,$$

for any $c>1$. Since the trivial zeros are isolated, let $\delta>0$ be such that there are no trivial zeros in the vertical strip $-2\delta\leq\sigma<0$. By Propositions 7.2.1 and 7.4.1, the integrand has polynomial decay of arbitrarily large order and therefore is locally absolutely uniformly convergent. Shifting the line of integration to $(-\delta)$, we pass by simple poles at $s=1$ and $s=\rho$ for every nontrivial zero ρ obtaining

$$\sum_{n\geq 1}\Lambda_f(n)\psi(n)=r_f\Psi(1)-\sum_{\rho}\Psi(\rho)+\frac{1}{2\pi i}\int_{(-\delta)}-\frac{L'}{L}(s,f)\Psi(s)\,ds.$$

For the latter integral, the functional equation and that $q(\overline{f})=q(f)$ together imply

$$-\frac{L'}{L}(s,f)=\log q(f)+\frac{\gamma'}{\gamma}(s,f)+\frac{\gamma'}{\gamma}(1-s,\overline{f})+\frac{L'}{L}(1-s,\overline{f}).$$

Thus

$$\frac{1}{2\pi i}\int_{(-\delta)}-\frac{L'}{L}(s,f)\Phi(s)\,ds=\frac{1}{2\pi i}\int_{(-\delta)}\left(\log q(f)+\frac{\gamma'}{\gamma}(s,f)+\frac{\gamma'}{\gamma}(1-s,\overline{f})+\frac{L'}{L}(1-s,\overline{f})\right)\Psi(s)\,ds.$$

The integral over the first term on the right-hand side is $\psi(1)\log q(f)$ by the Mellin inversion formula. As for the last term, smoothed Perron's formula and that $\Phi(s)=\Psi(1-s)$ together give

$$\frac{1}{2\pi i}\int_{(-\delta)}-\frac{L'}{L}(1-s,\overline{f})\Psi(s)\,ds=\sum_{n\geq 1}\Lambda_{\overline{f}}(n)\phi(n).$$

So altogether, we have

$$\begin{aligned}\sum_{n\geq 1}\big(\Lambda_f(n)\psi(n)+\Lambda_{\overline{f}}(n)\phi(n)\big)&=\psi(1)\log q(f)+r_f\Psi(1)\\&\quad+\frac{1}{2\pi i}\int_{(-\delta)}\left(\frac{\gamma'}{\gamma}(s,f)+\frac{\gamma'}{\gamma}(1-s,\overline{f})\right)\Psi(s)\,ds-\sum_{\rho}\Psi(\rho).\end{aligned}$$

Lastly, we note that by Propositions 1.7.3 and 7.2.1, the remaining integrand has polynomial decay of arbitrarily large order and therefore is locally absolutely uniformly convergent. Shifting the line of integration to $\left(\frac{1}{2}\right)$, we pass over no poles because the residues of $\frac{\gamma'}{\gamma}(1-s,\overline{f})$ are negative of those of $\frac{\gamma'}{\gamma}(s,f)$ for $\sigma\geq-\delta$ (the nontrivial zeros of $L(s,f)$ occur in pairs ρ and $1-\overline{\rho}$). This completes the proof. $\square$

The explicit formula is a very useful tool for analytic investigations. Since $\Lambda_f(n)$ is essentially a weighted sum over prime powers, the explicit formula can be thought of as expressing a smoothed weighted sum over primes for f in terms of the zeros of $L(s,f)$.

Chapter 8

Types of L-functions

We discuss a variety of L-functions: the Riemann zeta function, L-functions attached to Dirichlet characters, and Hecke L-functions. In the case of Hecke L-functions, we also describe a method of Rankin and Selberg for constructing new L-functions from old ones.

8.1 The Riemann Zeta Function

The Definition & Euler Product

The **Riemann zeta function** $\zeta(s)$ is defined by the following Dirichlet series:

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

This is the prototypical example of a Dirichlet series as all the coefficients are 1. We will see that $\zeta(s)$ is a primitive Selberg class L-function. As the coefficients are uniformly constant, $\zeta(s)$ is locally absolutely uniformly convergent for $\sigma > 1$. Also note that $\zeta(s)$ is necessarily nonzero in this region. Determining the Euler product is also an easy matter. As the coefficients are obviously completely multiplicative, we have the degree 1 Euler product:

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1},$$

in this region as well. The local factor at p is

$$\zeta_p(s) = (1 - p^{-s})^{-1},$$

with local root 1. As the degree of $\zeta(s)$ is 1, it will necessarily be primitive once we show it is Selberg class.

The Integral Representation: Part I

We want to find an integral representation for $\zeta(s)$. To do this, consider the function

$$\omega(z) = \sum_{n\geq 1} e^{\pi i n^2 z},$$

defined for $z \in \mathbb{H}$. It is locally absolutely uniformly convergent in this region by the ratio and Weierstrass M-tests. Moreover, the Taylor series of $\frac{e^y}{1-e^y}$ gives

$$\omega(z) = O\left(\sum_{n \geq 1} e^{-\pi n^2 y}\right) = O\left(\sum_{n \geq 1} e^{-\pi n y}\right) = O(e^{-\pi y}),$$

and so $\omega(z)$ exhibits rapid decay. Now consider the following Mellin transform:

$$\int_0^\infty \omega(iy) y^{\frac{s}{2}} \frac{dy}{y}.$$

By the rapid decay of $\omega(z)$, this integral exists and defines an analytic function for $\sigma > 1$. Then we compute

$$\begin{aligned}
\int_0^\infty \omega(iy) y^{\frac{s}{2}} \frac{dy}{y} &= \int_0^\infty \sum_{n \geq 1} e^{-\pi n^2 y} y^{\frac{s}{2}} \frac{dy}{y} \\
&= \sum_{n \geq 1} \int_0^\infty e^{-\pi n^2 y} y^{\frac{s}{2}} \frac{dy}{y} && \text{FT} \\
&= \sum_{n \geq 1} \frac{1}{\pi^{\frac{s}{2}} n^s} \int_0^\infty e^{-y} y^{\frac{s}{2}} \frac{dy}{y} && y \to \frac{y}{\pi n^2} \\
&= \frac{\Gamma\left(\frac{s}{2}\right)}{\pi^{\frac{s}{2}}} \sum_{n \geq 1} \frac{1}{n^s} \\
&= \frac{\Gamma\left(\frac{s}{2}\right)}{\pi^{\frac{s}{2}}} \zeta(s).
\end{aligned}$$

Therefore we have an integral representation

$$\zeta(s) = \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \int_0^\infty \omega(iy) y^{\frac{s}{2}} \frac{dy}{y}. \tag{8.1}$$

Unfortunately, we cannot proceed until we obtain a functional equation for $\omega(z)$. So we will make a slight detour and come back to the integral representation after.

The Jacobi Theta Function

The **Jacobi theta function** $\vartheta(z)$ is defined by

$$\vartheta(z) = \sum_{n \in \mathbb{Z}} e^{2\pi i n^2 z},$$

for $z \in \mathbb{H}$. Its relation to $\omega(z)$ is given by the identity

$$\omega(z) = \frac{\vartheta\left(\frac{z}{2}\right) - 1}{2}, \tag{8.2}$$

and so $\vartheta(z)$ is locally absolutely uniformly convergent in this region and exhibits rapid decay. The essential fact we will need is a transformation law for the Jacobi theta function:

Theorem 8.1.1. *For $z \in \mathbb{H}$,*

$$\vartheta(z) = \frac{1}{\sqrt{-2iz}}\vartheta\left(-\frac{1}{4z}\right).$$

Proof. By the identity theorem it suffices to verify this for $z = iy$ with $y > 0$. So set $f(x) = e^{-2\pi x^2 y}$. Then $f(x)$ is of Schwarz class. We compute its Fourier transform:

$$(\mathcal{F}f)(t) = \int_{\mathbb{R}} f(x)e^{-2\pi itx}\,dx = \int_{\mathbb{R}} e^{-2\pi x^2 y}e^{-2\pi itx}\,dx = \int_{\mathbb{R}} e^{-2\pi(x^2y+itx)}\,dx.$$

Making the change of variables $x \to \frac{x}{\sqrt{y}}$, the last integral above becomes

$$\frac{1}{\sqrt{y}}\int_{\mathbb{R}} e^{-2\pi\left(x^2+\frac{itx}{\sqrt{y}}\right)}.$$

Complete the square in the exponent by noticing

$$-2\pi\left(x^2+\frac{itx}{\sqrt{y}}\right) = -2\pi\left(\left(x+\frac{it}{2\sqrt{y}}\right)^2+\frac{t^2}{4y}\right).$$

Taking exponentials, this implies that the previous integral is equal to

$$\frac{e^{-\frac{\pi t^2}{2y}}}{\sqrt{y}}\int_{\mathbb{R}} e^{-2\pi\left(x+\frac{it}{2\sqrt{y}}\right)^2}\,dx.$$

The change of variables $x \to \frac{x}{\sqrt{2}} - \frac{it}{2\sqrt{y}}$ is permitted without affecting the line of integration by viewing the integral as a complex integral, noting that the integrand is entire as a complex function, and shifting the line of integration. This gives

$$\frac{e^{-\frac{\pi t^2}{2y}}}{\sqrt{2y}}\int_{\mathbb{R}} e^{-\pi x^2}\,dx = \frac{e^{-\frac{\pi t^2}{2y}}}{\sqrt{2y}},$$

where the last equality follows because the last integral above is 1 since it is the Gaussian integral (see Appendix E.1). Thus

$$(\mathcal{F}f)(t) = \frac{e^{-\frac{\pi t^2}{2y}}}{\sqrt{2y}}.$$

By the Poisson summation formula, we have

$$\vartheta(iy) = \sum_{t\in\mathbb{Z}} \frac{e^{-\frac{\pi t^2}{2y}}}{\sqrt{2y}} = \frac{1}{\sqrt{2y}}\sum_{t\in\mathbb{Z}} e^{-\frac{\pi t^2}{2y}} = \frac{1}{\sqrt{2y}}\vartheta\left(-\frac{1}{4iy}\right),$$

and the identity theorem finishes the proof. □

We will use Theorem 8.1.1 to analytically continue $\zeta(s)$.

The Integral Representation: Part II

Returning to the Riemann zeta function, we split the integral in Equation (8.1) into two pieces

$$\int_0^\infty \omega(iy)y^{\frac{s}{2}}\,\frac{dy}{y} = \int_0^1 \omega(iy)y^{\frac{s}{2}}\,\frac{dy}{y} + \int_1^\infty \omega(iy)y^{\frac{s}{2}}\,\frac{dy}{y}. \tag{8.3}$$

The idea now is to rewrite the first piece in the same form and symmetrize the result as much as possible. We being by performing a change of variables $y \to \frac{1}{y}$ to the first piece to obtain

$$\int_1^\infty \omega\left(\frac{i}{y}\right) y^{-\frac{s}{2}}\,\frac{dy}{y}$$

Now Equation (8.2) and Theorem 8.1.1 together imply

$$\begin{aligned}
\omega\left(\frac{i}{y}\right) &= \omega\left(-\frac{1}{iy}\right) \\
&= \frac{\vartheta\left(-\frac{1}{2iy}\right)-1}{2} \\
&= \frac{\sqrt{y}\vartheta\left(\frac{iy}{2}\right)-1}{2} \\
&= \frac{\sqrt{y}(2\omega(iy)+1)-1}{2} \\
&= \sqrt{y}\omega(iy) + \frac{\sqrt{y}}{2} - \frac{1}{2}.
\end{aligned}$$

This relation gives the first equality in the following chain:

$$\begin{aligned}
\int_1^\infty \omega\left(\frac{1}{y}\right) y^{-\frac{s}{2}}\,\frac{dy}{y} &= \int_1^\infty \left(\sqrt{y}\omega(iy) + \frac{\sqrt{y}}{2} - \frac{1}{2}\right) y^{-\frac{s}{2}}\,\frac{dy}{y} \\
&= \int_1^\infty \omega(iy)y^{\frac{1-s}{2}}\,\frac{dy}{y} + \int_1^\infty \frac{x^{\frac{1-s}{2}}}{2}\,\frac{dy}{y} - \int_1^\infty \frac{y^{-\frac{s}{2}}}{2}\,\frac{dy}{y} \\
&= \int_1^\infty \omega(iy)y^{\frac{1-s}{2}}\,\frac{dy}{y} + \frac{1}{1-s} - \frac{1}{s} \\
&= \int_1^\infty \omega(iy)y^{\frac{1-s}{2}}\,\frac{dy}{y} - \frac{1}{s(1-s)}.
\end{aligned}$$

Substituting this result back into Equation (8.3) with Equation (8.1) yields the integral representation

$$\zeta(s) = \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)}\left[-\frac{1}{s(1-s)} + \int_1^\infty \omega(iy)y^{\frac{1-s}{2}}\,\frac{dy}{y} + \int_1^\infty \omega(iy)y^{\frac{s}{2}}\,\frac{dy}{y}\right].$$

This integral representation will give analytic continuation. To see this, first observe that everything outside the brackets is entire. Moreover, the integrands exhibit rapid decay and therefore the integrals are locally absolutely uniformly convergent on $\mathbb{C}$. The fractional term is holomorphic except for simple poles at $s=0$ and $s=1$. The meromorphic continuation to $\mathbb{C}$ follows with possible simple poles at $s=0$ and $s=1$. There is no pole at $s=0$. Indeed, $\gamma(s,\zeta)$ has a simple pole coming from the gamma factor there and so its reciprocal has a simple zero. This cancels the corresponding simple pole of $\frac{1}{s(1-s)}$ so that $\zeta(s)$ has a removable singularity and thus is holomorphic at $s=0$. At $s=1$, $\gamma(s,\zeta)$ is nonzero, and so $\zeta(s)$ has a simple pole. Therefore $\zeta(s)$ has meromorphic continuation to all of $\mathbb{C}$ with a simple pole at $s=1$.

The Functional Equation

An immediate consequence of applying the symmetry $s \to 1-s$ to the integral representation is the following functional equation:

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

We identify the gamma factor as

$$\gamma(s,\zeta) = \pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right),$$

with $\kappa = 0$ the only local root at infinity. Clearly it satisfies the required bounds. The conductor is $q(\zeta) = 1$ so no primes ramify. The completed Riemann zeta function is

$$\Lambda(s,\zeta) = \pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s),$$

with functional equation

$$\Lambda(s,\zeta) = \Lambda(1-s,\zeta).$$

This is the functional equation of $\zeta(s)$ and in this case is just a reformulation of the previous functional equation. From it we find that the root number is $\varepsilon(\zeta) = 1$ and that $\zeta(s)$ is self-dual. We can now show that the order of $\zeta(s)$ is 1. As there is only a simple pole at $s = 1$, multiply by $(s-1)$ to clear the polar divisor. As the integrals in the integral representation are locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. Since the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s,\zeta)} \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

Thus the reciprocal of the gamma factor is also of order 1. It follows that

$$(s-1)\zeta(s) \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

This shows $(s-1)\zeta(s)$ is of order 1, and thus $\zeta(s)$ is as well after removing the polar divisor. We now compute the residue of $\zeta(s)$ at $s = 1$:

Proposition 8.1.1.

$$\operatorname*{Res}_{s=1}\zeta(s) = 1.$$

Proof. The only term in the integral representation of $\zeta(s)$ contributing to the pole is $-\frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)}\frac{1}{s(1-s)}$. Observe

$$\lim_{s\to1}\frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} = 1,$$

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

$$\operatorname*{Res}_{s=1}\zeta(s) = \operatorname*{Res}_{s=1} -\frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)}\frac{1}{s(1-s)} = \operatorname*{Res}_{s=1} -\frac{1}{s(1-s)} = \lim_{s\to1} -\frac{(s-1)}{s(1-s)} = 1. \qquad \square$$

We summarize all of our work into the following theorem:

Theorem 8.1.2. *$\zeta(s)$ is a primitive Selberg class L-function. For $\sigma > 1$, it has a degree 1 Euler product given by*

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}.$$

Moreover, it admits meromorphic continuation to $\mathbb{C}$, possesses the functional equation

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

and has a simple pole at $s = 1$ of residue 1.

Lastly, we note that by virtue of the functional equation we can also compute $\zeta(0)$. Indeed, since $\mathrm{Res}_{s=1} \zeta(s) = 1$, we have

$$\lim_{s \to 1} (s - 1)\Lambda(s, \zeta) = \operatorname*{Res}_{s=1} \zeta(s) \lim_{s \to 1} \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) = 1.$$

In other words, $\Lambda(s, \zeta)$ has a simple pole at $s = 1$ with residue 1 too. Since the completed Riemann zeta function is completely symmetric as $s \to 1 - s$, it has a simple pole at $s = 0$ with residue 1. Hence

$$1 = \lim_{s \to 1} (s - 1)\Lambda(1 - s, \zeta) = \operatorname*{Res}_{s=1} \Gamma\left(\frac{1 - s}{2}\right) \lim_{s \to 1} \pi^{-\frac{1-s}{2}} \zeta(1 - s) = -2\zeta(0),$$

because $\mathrm{Res}_{s=0} \Gamma(s) = 1$. Therefore $\zeta(0) = -\frac{1}{2}$.

8.2 Dirichlet L-functions

The Definition & Euler Product

To every Dirichlet character χ there is an associated L-function. Throughout we will let m denote the modulus and q the conductor of χ respectively. The **Dirichlet L-series** (respectively **Dirichlet L-function** if it is an L-function) $L(s, \chi)$ attached to the Dirichlet character χ is defined by the following Dirichlet series:

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

Since $\chi(n) = 0$ if $(n, m) > 1$, the above sum can be restricted to all positive integers relatively prime to m. We will see that $L(s, \chi)$ is a Selberg class L-function if χ is primitive and of conductor $q > 1$ (in the case $q = 1$, $L(s, \chi) = \zeta(s)$). From now we make this assumption about χ. As $|\chi(n)| \ll 1$, $L(s, \chi)$ is locally absolutely uniformly convergent for $\sigma > 1$. Because χ is completely multiplicative we also have the degree 1 Euler product:

$$L(s, \chi) = \prod_p (1 - \chi(p)p^{-s})^{-1} = \prod_{p \nmid q} (1 - \chi(p)p^{-s})^{-1},$$

in this region as well. The last equality holds because if $p \mid q$ we have $\chi(p) = 0$. The local factor at p is

$$L_p(s, \chi) = 1 \quad \text{or} \quad L_p(s, \chi) = (1 - \chi(p)p^{-s})^{-1},$$

with local root 0 or $\chi(p)$ respectively and according to if $p \mid q$ or not. As the degree of $L(s, \chi)$ is 1, it will necessarily be primitive once we show it is Selberg class.

The Integral Representation: Part I

The integral representation for $L(s,\chi)$ is deduced in a similar way as for $\zeta(s)$. However, it will depend on if χ is even or odd. To handle both cases simultaneously let $\mathfrak{a} = 0, 1$ according to whether χ is even or odd. In other words,

$$\mathfrak{a} = \frac{\chi(1) - \chi(-1)}{2}.$$

We also have $\chi(-1) = (-1)^{\mathfrak{a}}$. Note that $\mathfrak{a}$ takes the same value for both χ and $\overline{\chi}$. To find an integral representation for $L(s,\chi)$, consider the function

$$\omega_\chi(z) = \sum_{n\geq 1} \chi(n) n^{\mathfrak{a}} e^{\pi i n^2 z},$$

defined for $z \in \mathbb{H}$. It is locally absolutely uniformly convergent in this region by the ratio and Weierstrass M-tests. Moreover, the Taylor series of $\frac{e^y}{1-e^y}$ and its derivative together give

$$\omega_\chi(z) = O\left(\sum_{n\geq 1} n e^{-\pi n^2 y}\right) = O\left(\sum_{n\geq 1} n e^{-\pi n y}\right) = O(e^{-\pi y}),$$

and so $\omega_\chi(z)$ exhibits rapid decay. Now consider the following Mellin transform:

$$\int_0^\infty \omega_\chi(iy) y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y}.$$

By the rapid decay of w_χ, this integral exists and defines an analytic function for $\sigma > 1$. Then we compute

$$\begin{aligned}
\int_0^\infty \omega_\chi(iy) y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y} &= \int_0^\infty \sum_{n\geq 1} \chi(n) n^{\mathfrak{a}} e^{-\pi n^2 y} y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y} \\
&= \sum_{n\geq 1} \int_0^\infty \chi(n) n^{\mathfrak{a}} e^{-\pi n^2 y} y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y} && \text{FT} \\
&= \sum_{n\geq 1} \frac{\chi(n)}{\pi^{\frac{s+\mathfrak{a}}{2}} n^s} \int_0^\infty e^{-y} y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y} && y \to \frac{y}{\pi n^2} \\
&= \frac{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}{\pi^{\frac{s+\mathfrak{a}}{2}}} \sum_{n\geq 1} \frac{\chi(n)}{n^s} \\
&= \frac{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}{\pi^{\frac{s+\mathfrak{a}}{2}}} L(s,\chi).
\end{aligned}$$

Therefore we have an integral representation

$$L(s,\chi) = \frac{\pi^{\frac{s+\mathfrak{a}}{2}}}{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)} \int_0^\infty \omega_\chi(iy) y^{\frac{s+\mathfrak{a}}{2}} \frac{dy}{y}, \tag{8.4}$$

and just as for the Riemann zeta function, we need to find a functional equation for $\omega_\chi(z)$ before we can proceed.

The Dirichlet Theta Function

The **Dirichlet theta function** $\vartheta_\chi(z)$ attached to the character χ, is defined by

$$\vartheta_\chi(z) = \sum_{n\in\mathbb{Z}} \chi(n)n^{\mathfrak{a}} e^{2\pi i n^2 z},$$

for $z \in \mathbb{H}$. The relationship to $\omega_\chi(z)$ is

$$\omega_\chi(z) = \frac{\vartheta_\chi\left(\frac{z}{2}\right)}{2}, \tag{8.5}$$

and so $\vartheta_\chi(z)$ is locally absolutely uniformly convergent in this region and exhibits rapid decay.

Remark 8.2.1. *Equation* (8.5) *is a slightly less complex relationship than Equation* (8.2). *This is because assuming $q > 1$ means $\chi(0) = 0$.*

The essential fact we will need is a transformation law for the Dirichlet theta function:

Theorem 8.2.1. *Let χ be a primitive Dirichlet character of conductor $q > 1$. For $z \in \mathbb{H}$,*

$$\vartheta_\chi(z) = \frac{\varepsilon_\chi}{i^{\mathfrak{a}}(-2qiz)^{\frac{1}{2}+\mathfrak{a}}} \vartheta_{\overline{\chi}}\left(-\frac{1}{4q^2 z}\right).$$

Proof. By the identity theorem it suffices to verify this for $z = iy$ with $y > 0$. Since χ is q-periodic and $\vartheta_\chi(z)$ is absolutely convergent, we can write

$$\vartheta_\chi(iy) = \sum_{a \pmod q} \chi(a) \sum_{m\in\mathbb{Z}} (mq+a)^{\mathfrak{a}} e^{-2\pi(mq+a)^2 y}.$$

Set $f(x) = (xq+a)^{\mathfrak{a}} e^{-2\pi(xq+a)^2 y}$. Then $f(x)$ is of Schwarz class. We compute its Fourier transform:

$$(\mathcal{F}f)(t) = \int_{\mathbb{R}} f(x)e^{-2\pi i t x}\,dx = \int_{\mathbb{R}} (xq+a)^{\mathfrak{a}} e^{-2\pi(xq+a)^2 y} e^{-2\pi i t x}\,dx = \int_{\mathbb{R}} (xq+a)^{\mathfrak{a}} e^{-2\pi((xq+a)^2 y + itx)}\,dx.$$

By performing the change of variables $x \to \frac{x}{q\sqrt{y}} - \frac{a}{q}$, the last integral above becomes

$$\frac{e^{\frac{2\pi i a t}{q}}}{qy^{\frac{1+\mathfrak{a}}{2}}} \int_{\mathbb{R}} x^{\mathfrak{a}} e^{-2\pi\left(x^2 + \frac{itx}{q\sqrt{y}}\right)}\,dx.$$

Complete the square in the exponent by observing

$$-2\pi\left(x^2 + \frac{itx}{q\sqrt{y}}\right) = -2\pi\left(\left(x + \frac{it}{2q\sqrt{y}}\right)^2 + \frac{t^2}{4q^2 y}\right).$$

Taking exponentials, this implies that the previous integral is equal to

$$\frac{e^{\frac{2\pi i a t}{q} - \frac{\pi t^2}{2q^2 y}}}{qy^{\frac{1+\mathfrak{a}}{2}}} \int_{\mathbb{R}} x^{\mathfrak{a}} e^{-2\pi\left(x + \frac{it}{2q\sqrt{y}}\right)^2}\,dx.$$

The change of variables $x \to \frac{x}{\sqrt{2}} - \frac{it}{2q\sqrt{y}}$ is permitted without affecting the line of integration by viewing the integral as a complex integral, noting that the integrand is entire as a complex function, and shifting the line of integration. This gives

$$\frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\int_{\mathbb{R}}\left(\frac{x}{\sqrt{2}}-\frac{it}{2q\sqrt{y}}\right)^{\mathfrak{a}}e^{-\pi x^2}\,dx = \frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\int_{\mathbb{R}}\left(\frac{x}{\sqrt{2}}+\frac{t}{2qi\sqrt{y}}\right)^{\mathfrak{a}}e^{-\pi x^2}\,dx.$$

If $\mathfrak{a}=0$, we obtain

$$\frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\int_{\mathbb{R}}e^{-\pi x^2}\,dx = \frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}, \tag{8.6}$$

where the equality holds because the integral is 1 since it is the Gaussian integral (see Appendix E.1). If $\mathfrak{a}=1$, then by direct computation

$$\int_{\mathbb{R}}\frac{x}{\sqrt{2}}e^{-\pi x^2}\,dx = -\frac{1}{\sqrt{8\pi}}e^{-\pi x^2}\bigg|_{-\infty}^{\infty} = 0,$$

and

$$\frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\int_{\mathbb{R}}\left(\frac{t}{2qi\sqrt{y}}\right)e^{-\pi x^2}\,dx = \frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\left(\frac{t}{2qi\sqrt{y}}\right)\int_{\mathbb{R}}e^{-\pi x^2}\,dx = \frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\left(\frac{t}{2qi\sqrt{y}}\right), \tag{8.7}$$

where the last equality follows because the last integral is the Gaussian integral again. Since $\left(\frac{t}{2qi\sqrt{y}}\right)^{\mathfrak{a}}=1$ if $\mathfrak{a}=0$, Equations (8.6) and (8.7) together imply

$$(\mathcal{F}f)(t) = \frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\left(\frac{t}{2qi\sqrt{y}}\right)^{\mathfrak{a}}.$$

By the Poisson summation formula, we have

$$\begin{aligned}
\vartheta_\chi(iy) &= \sum_{a \pmod q}\chi(a)\sum_{t\in\mathbb{Z}}\frac{e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}}}{\sqrt{2}qy^{\frac{1+\mathfrak{a}}{2}}}\left(\frac{t}{2qi\sqrt{y}}\right)^{\mathfrak{a}} \\
&= \frac{1}{i^{\mathfrak{a}}q^{1+\mathfrak{a}}(2y)^{\frac{1}{2}+\mathfrak{a}}}\sum_{a \pmod q}\chi(a)\sum_{t\in\mathbb{Z}}t^{\mathfrak{a}}e^{\frac{2\pi iat}{q}-\frac{\pi t^2}{2q^2y}} \\
&= \frac{1}{i^{\mathfrak{a}}q^{1+\mathfrak{a}}(2y)^{\frac{1}{2}+\mathfrak{a}}}\sum_{t\in\mathbb{Z}}t^{\mathfrak{a}}e^{-\frac{\pi t^2}{2q^2y}}\sum_{a \pmod q}\chi(a)e^{\frac{2\pi iat}{q}} \\
&= \frac{1}{i^{\mathfrak{a}}q^{1+\mathfrak{a}}(2y)^{\frac{1}{2}+\mathfrak{a}}}\sum_{t\in\mathbb{Z}}t^{\mathfrak{a}}e^{-\frac{\pi t^2}{2q^2y}}\tau(t,\chi) && \text{definition of } \tau(t,\chi) \\
&= \frac{\tau(\chi)}{i^{\mathfrak{a}}q^{1+\mathfrak{a}}(2y)^{\frac{1}{2}+\mathfrak{a}}}\sum_{t\in\mathbb{Z}}\overline{\chi}(t)t^{\mathfrak{a}}e^{-\frac{\pi t^2}{2q^2y}} && \text{Corollary 1.4.1} \\
&= \frac{\varepsilon_\chi}{i^{\mathfrak{a}}(2qy)^{\frac{1}{2}+\mathfrak{a}}}\sum_{t\in\mathbb{Z}}\overline{\chi}(t)t^{\mathfrak{a}}e^{-\frac{\pi t^2}{2q^2y}} && \varepsilon_\chi = \frac{\tau(\chi)}{\sqrt{q}} \\
&= \frac{\varepsilon_\chi}{i^{\mathfrak{a}}(2qy)^{\frac{1}{2}+\mathfrak{a}}}\vartheta_{\overline{\chi}}\left(-\frac{1}{4q^2iy}\right),
\end{aligned}$$

and the identity theorem finishes the proof. $\square$

Notice that the functional equation relates $\vartheta_\chi(z)$ to $\vartheta_{\overline{\chi}}(z)$. Regardless, we will use Theorem 8.2.1 to analytically continue $L(s,\chi)$.

The Integral Representation: Part II

Returning to $L(s,\chi)$, split the integral in Equation (8.4) into two pieces

$$\int_0^\infty \omega_\chi(iy)y^{\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y} = \int_0^{\frac{1}{q}} \omega_\chi(iy)y^{\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y} + \int_{\frac{1}{q}}^\infty \omega_\chi(iy)y^{\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y}. \tag{8.8}$$

We now rewrite the first piece in the same form and symmetrize the result as much as possible. Start by performing a change of variables $y \to \frac{1}{q^2y}$ to the first piece to obtain

$$q^{-(s+\mathfrak{a})}\int_{\frac{1}{q}}^\infty \omega_\chi\left(\frac{i}{q^2y}\right)y^{-\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y}.$$

Now Equation (8.5) and Theorem 8.2.1 together imply

$$\begin{aligned}
\omega_\chi\left(\frac{i}{q^2y}\right) &= \omega_\chi\left(-\frac{1}{q^2iy}\right)\\
&= \frac{\vartheta_\chi\left(-\frac{1}{2q^2iy}\right)}{2}\\
&= \frac{i^{\mathfrak{a}}(qy)^{\frac{1}{2}+\mathfrak{a}}}{\varepsilon_{\overline{\chi}}}\frac{\vartheta_{\overline{\chi}}\left(\frac{iy}{2}\right)}{2}\\
&= \frac{i^{\mathfrak{a}}(qy)^{\frac{1}{2}+\mathfrak{a}}}{\varepsilon_{\overline{\chi}}}\frac{\vartheta_{\overline{\chi}}\left(\frac{iy}{2}\right)}{2}\\
&= \varepsilon_\chi(-i)^{\mathfrak{a}}(qy)^{\frac{1}{2}+\mathfrak{a}}\frac{\vartheta_{\overline{\chi}}\left(\frac{iy}{2}\right)}{2} && \text{Proposition 1.4.3 and } \chi(-1)=(-1)^{\mathfrak{a}}\\
&= \frac{\varepsilon_\chi(qy)^{\frac{1}{2}+\mathfrak{a}}}{i^{\mathfrak{a}}}\frac{\vartheta_{\overline{\chi}}\left(\frac{iy}{2}\right)}{2}\\
&= \frac{\varepsilon_\chi(qy)^{\frac{1}{2}+\mathfrak{a}}}{i^{\mathfrak{a}}}\omega_{\overline{\chi}}(iy).
\end{aligned}$$

This relation gives the first equality in the following chain:

$$\begin{aligned}
q^{-(s+\mathfrak{a})}\int_{\frac{1}{q}}^\infty \omega_\chi\left(\frac{1}{q^2y}\right)y^{-\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y} &= q^{-(s+\mathfrak{a})}\int_{\frac{1}{q}}^\infty \left(\frac{\varepsilon_\chi(qy)^{\frac{1}{2}+\mathfrak{a}}}{i^{\mathfrak{a}}}\omega_{\overline{\chi}}(iy)\right)y^{-\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y}\\
&= \frac{\varepsilon_\chi}{i^{\mathfrak{a}}}q^{\frac{1}{2}-s}\int_{\frac{1}{q}}^\infty \omega_{\overline{\chi}}(iy)y^{\frac{(1-s)+\mathfrak{a}}{2}}\,\frac{dy}{y}.
\end{aligned}$$

Substituting this last expression back into Equation (8.8) with Equation (8.4) gives the integral representation

$$L(s,\chi) = \frac{\pi^{\frac{s+\mathfrak{a}}{2}}}{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}\left[\frac{\varepsilon_\chi}{i^{\mathfrak{a}}}q^{\frac{1}{2}-s}\int_{\frac{1}{q}}^\infty \omega_{\overline{\chi}}(iy)y^{\frac{(1-s)+\mathfrak{a}}{2}}\,\frac{dy}{y} + \int_{\frac{1}{q}}^\infty \omega_\chi(iy)y^{\frac{s+\mathfrak{a}}{2}}\,\frac{dy}{y}\right].$$

This integral representation will give analytic continuation. Indeed, we know everything outside the brackets is entire. The integrands exhibit rapid decay and therefore the integrals are locally absolutely uniformly convergent on $\mathbb{C}$. This gives analytic continuation to all of $\mathbb{C}$. In particular, $L(s,\chi)$ has no poles.

The Functional Equation

An immediate consequence of applying the symmetry $s \to 1-s$ to the integral representation is the following functional equation:

$$q^{\frac{s}{2}} \frac{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}{\pi^{\frac{s+\mathfrak{a}}{2}}} L(s,\chi) = \frac{\varepsilon_\chi}{i^{\mathfrak{a}}} q^{\frac{1-s}{2}} \frac{\Gamma\left(\frac{(1-s)+\mathfrak{a}}{2}\right)}{\pi^{\frac{(1-s)+\mathfrak{a}}{2}}} L(1-s,\overline{\chi}).$$

We identify the gamma factor as

$$\gamma(s,\chi) = \pi^{-\frac{s}{2}} \Gamma\left(\frac{s+\mathfrak{a}}{2}\right),$$

with $\kappa = \mathfrak{a}$ the only local root at infinity. Clearly it satisfies the required bounds. The conductor is $q(\chi) = q$ and if p is an unramified prime then the local root is $\chi(p) \neq 0$. The completed L-function is

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

with functional equation

$$\Lambda(s,\chi) = \frac{\varepsilon_\chi}{i^{\mathfrak{a}}} \Lambda(1-s,\overline{\chi}).$$

From it we see that the root number is $\varepsilon(\chi) = \frac{\varepsilon_\chi}{i^{\mathfrak{a}}}$ and that $L(s,\chi)$ has dual $L(s,\overline{\chi})$. We now show that $L(s,\chi)$ is of order 1. Since $L(s,\chi)$ has no poles, we do not need to clear any polar divisors. As the integrals in the integral representation are locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. Since the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s,\chi)} \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So the reciprocal of the gamma factor is also of order 1. It follows that

$$L(s,\chi) \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So $L(s,\chi)$ is of order 1. We summarize all of our work into the following theorem:

Theorem 8.2.2. *For any primitive Dirichlet character χ of conductor $q > 1$, $L(s,\chi)$ is a primitive Selberg class L-function. For $\sigma > 1$, it has a degree 1 Euler product given by*

$$L(s,\chi) = \prod_p (1-\chi(p)p^{-s})^{-1} = \prod_{p \nmid q} (1-\chi(p)p^{-s})^{-1}.$$

Moreover, it admits analytic continuation to $\mathbb{C}$ and possesses the functional equation

$$q^{\frac{s}{2}} \pi^{-\frac{s}{2}} \Gamma\left(\frac{s+\mathfrak{a}}{2}\right) L(s,\chi) = \Lambda(s,\chi) = \frac{\varepsilon_\chi}{i^{\mathfrak{a}}} \Lambda(1-s,\overline{\chi}).$$

Beyond Primitivity

We can still obtain meromorphic continuation of the L-series $L(s,\chi)$ if χ is imprimitive. Indeed, if χ is induced by $\widetilde{\chi}$, then $\chi(p) = \widetilde{\chi}(p)$ if $p \nmid q$ and $\chi(p) = 0$ if $p \mid m$. Moreover, χ is completely multiplicative so we have a degree 1 Euler product:

$$L(s,\chi) = \prod_{p \nmid m} (1-\widetilde{\chi}(p)p^{-s})^{-1} = \prod_p (1-\widetilde{\chi}(p)p^{-s})^{-1} \prod_{p \mid m} (1-\widetilde{\chi}(p)p^{-s}) = L(s,\widetilde{\chi}) \prod_{p \mid m} (1-\widetilde{\chi}(p)p^{-s}). \tag{8.9}$$

From this relation, we can prove the following:

Theorem 8.2.3. *For any Dirichlet character χ modulo m of conductor $q > 1$, $L(s,\chi)$ admits meromorphic continuation to $\mathbb{C}$ and if χ is principal there is a simple pole at $s = 1$ of residue $\prod_{p|m}(1 - \widetilde{\chi}(p)p^{-1})$ where $\widetilde{\chi}$ is the primitive character inducing χ.*

Proof. This follows from Theorems 8.1.2 and 8.2.2 and Equation (8.9). □

8.3 Todo: [Dedekind Zeta Functions]

The Definition & Euler Product

We can associate and L-function to every number field. Let K be a number field of degree d and signature (r_1, r_2). The **Dedekind zeta function** $\zeta_K(s)$ is defined by the following Dirichlet series

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

We will see that $\zeta_K(s)$ is a Selberg class L-function. We have already shown that the coefficients $a_K(n)$ are polynomially bounded, for Proposition 2.6.2 implies $a_K(n) \leq \sigma_0(n)^d \ll_\varepsilon n^\varepsilon$ (recall Proposition A.3.1). Therefore $\zeta_K(s)$ is locally absolutely uniformly convergent for $\sigma > 1$. To determine the Euler product, we need to express $\zeta_K(s)$ in another form. Since $\zeta_K(s)$ converges absolutely, we can write

$$\zeta_K(s) = \sum_{\mathfrak{a}} \frac{1}{\mathrm{N}(\alpha)^s},$$

where the sum is over all integral ideals of K. Since the norm is completely multiplicative by Proposition 2.4.5, the unique factorization of integral ideals implies that we can express $\zeta_K(s)$ as the infinite product

$$\zeta_K(s) = \prod_{\mathfrak{p}} (1 - \mathrm{N}(\mathfrak{p})p^{-s})^{-1},$$

for $\sigma > 1$, and where the product is over all prime integral ideals $\mathfrak{p}$ of K. Since each such ideal is above a unique prime p and this product is absolutely convergent for $\sigma > 1$, it can be expressed as

$$\zeta_K(s) = \prod_p \prod_{\mathfrak{p} \text{ above } p} (1 - \mathrm{N}(\mathfrak{p})p^{-s})^{-1} = \prod_p \prod_{1 \leq i \leq r_p} (1 - p^{-f_p(\mathfrak{p}_i)s})^{-1},$$

where r_p is the number of prime integral ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_{r_p}$ above p. As finitely many primes ramify by Theorem 2.6.1, Proposition 2.6.1 implies that all but finitely many primes p satisfy

$$d = \sum_{1 \leq i \leq r_p} f_p(\mathfrak{p}_i).$$

In particular, at least one prime does. Therefore, letting $\omega_{p,i}$ be a primitive $f_p(\mathfrak{p}_i)$-th root of unity, we have the degree d Euler product:

$$\zeta_K(s) = \prod_p \prod_{\substack{1 \leq i \leq r_p \\ 1 \leq k \leq f_p(\mathfrak{p}_i)}} (1 - \omega_{p,i}^k p^{-s})^{-1},$$

in this region as well. The local factor at p is

$$\zeta_{K,p}(s) = \prod_{\substack{1 \leq i \leq r_p \\ 1 \leq k \leq f_p(\mathfrak{p}_i)}} (1 - \omega_{p,i}^k p^{-s})^{-1},$$

with local roots $\omega_{p,i}^k$ and $\sum_{1 \leq i \leq r_p} (e_p(\mathfrak{p}_i) - 1) f_p(\mathfrak{p}_i)$ many local roots 0.

Todo: [The Integral Representation: Part I]

Todo: [The Hecke Theta Function]

Todo: [The Integral Representation: Part II]

Todo: [The Functional Equation]

Todo: [xxx] We identify the gamma factor of $\zeta_K(s)$ to be

$$\gamma(s,\zeta_K)=\pi^{-\frac{ds}{2}}\Gamma\left(\frac{s}{2}\right)^{r_1+r_2}\Gamma\left(\frac{s+1}{2}\right)^{r_2},$$

with $\kappa_i=0$ for $1\le i\le r_1+r_2$ and $\kappa_i=1$ for $r_1+r_2+1\le i\le d$ the local roots at infinity. Clearly these satisfy the required bounds. The conductor is $q(\zeta_K)=|\Delta_K|$ and p is ramified if and only if it is ramified in K by Theorem 2.6.1. Note that if p is unramified then $\sum_{1\le i\le r_p}(e_p(\mathfrak{p}_i)-1)f_p(\mathfrak{p}_i)=0$ so that all local roots are of the form $\omega_{p,i}^k\neq 0$. The completed Dedekind zeta function is

$$\Lambda(s,\zeta_K)=|\Delta_K|^{\frac{s}{2}}\pi^{-\frac{ds}{2}}\Gamma\left(\frac{s}{2}\right)^{r_1+r_2}\Gamma\left(\frac{s+1}{2}\right)^{r_2}\zeta_K(s),$$

with functional equation

$$\Lambda(s,\zeta_K)=\Lambda(1-s,\zeta_K).$$

This is the function equation of $\zeta_K(s)$ and from it we see that the root number is $\varepsilon(\zeta_K)=1$ and that $\zeta_K(s)$ is self-dual. To see that the order is 1, multiply by $(s-1)$ to clear the polar divisor. As the integrals in the integral representation are locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. And because the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s,\zeta_K)}\ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

Therefore the reciprocal of the gamma factor is also of order 1 and it follows that

$$(s-1)\zeta_K(s)\ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

This shows $(s-1)\zeta_K(s)$ is of order 1 and hence $\zeta_K(s)$ is too after removing the polar divisor. It now remains to compute the residue of $\zeta_K(s)$ at $s=1$. This result is known as the infamous **analytic class number formula**:

Theorem 8.3.1 (Analytic class number formula). *Let K be a number field with signature (r_1,r_2). Then*

$$\operatorname*{Res}_{s=1}\zeta_K(s)=\frac{2^{r_1}(2\pi)^{r_2}R_Kh_K}{w_K\sqrt{|\Delta_K|}}.$$

Proof. Todo: [xxx] □

We collect our work in the following theorem:

Theorem 8.3.2. *Let K be a number field of degree d and signature (r_1, r_2). Then $\zeta_K(s)$ is a Selberg class L-function. For $\sigma > 1$, it has a degree d Euler product given by*

$$\zeta_K(s) = \prod_p \prod_{\substack{1 \le i \le r_p \\ 1 \le k \le f_p(\mathfrak{p}_i)}} (1 - \omega_{p,i}^k p^{-s})^{-1},$$

Moreover, it admits meromorphic continuation to $\mathbb{C}$, possesses the functional equation

$$|\Delta_K|^{\frac{s}{2}} \pi^{-\frac{ds}{2}} \Gamma\left(\frac{s}{2}\right)^{r_1+r_2} \Gamma\left(\frac{s+1}{2}\right)^{r_2} \zeta_K(s) = \Lambda(s, \zeta_K) = \Lambda(1-s, \zeta_K),$$

and has a simple pole at $s = 1$ of reside $\frac{2^{r_1}(2\pi)^{r_2} R_K h_K}{w_K \sqrt{|\Delta_K|}}$.

Imprimitivity

Dedekind zeta functions are generally imprimitive L-functions. We illustrate this in the case of quadratic number fields:

Theorem 8.3.3. *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic number field and let χ_{Δ_d} be the quadratic character given by the Kronecker symbol. Then*

$$\zeta_K(s) = \zeta(s) L(s, \chi_{\Delta_d}).$$

Proof. By the identity theorem it suffices to prove this on a set containing a limit point. Therefore we may assume $\sigma > 1$. By Proposition 2.10.4, the Euler product of ζ_K is given by

$$\begin{aligned}
\zeta_K(s) &= \prod_{\substack{p \\ \chi_{\Delta_d}(p)=1}} (1-p^{-s})^{-1}(1-p^{-s})^{-1} \prod_{\substack{p \\ \chi_{\Delta_d}(p)=-1}} (1-p^{-s})^{-1}(1+p^{-s})^{-1} \prod_{\substack{p \\ \chi_{\Delta_d}(p)=0}} (1-p^{-s})^{-1} \\
&= \zeta(s) \prod_{\substack{p \\ \chi_{\Delta_d}(p)=1}} (1-p^{-s})^{-1} \prod_{\substack{p \\ \chi_{\Delta_d}(p)=-1}} (1+p^{-s})^{-1} \prod_{\substack{p \\ \chi_{\Delta_d}(p)=0}} 1 \\
&= \zeta(s) L(s, \chi_{\Delta_p}).
\end{aligned}$$

This claim now follows. ∎

8.4 Hecke L-functions

The Definition & Euler Product

We will investigate the L-functions of holomorphic cusp forms. Let $f \in \mathcal{S}_k(N, \chi)$ and denote its Fourier series by

$$f(z) = \sum_{n \ge 1} a_f(n) n^{\frac{k-1}{2}} e^{2\pi i n z},$$

with $a_f(1) = 1$. Thus if f is a Hecke eigenform, the $a_f(n)$ are the Hecke eigenvalues of f normalized so that they are constant on average. The **Hecke L-series** (respectively **Hecke L-function** if it is an L-function) $L(s, f)$ of f is defined by the following Dirichlet series:

$$L(s, f) = \sum_{n \ge 1} \frac{a_f(n)}{n^s}.$$

We will see that $L(s,f)$ is a Selberg class L-function if f is a primitive Hecke eigenform. From now on, we make this assumption about f. As we have noted, the Hecke relations and the Ramanujan-Petersson conjecture for holomorphic forms together imply $a_f(n) \ll_\varepsilon n^\varepsilon$. Therefore $L(s,f)$ is locally absolutely uniformly convergent for $\sigma > 1$. The L-function will also have an Euler product. Indeed, the Hecke relations imply that the coefficients $a_f(n)$ are multiplicative and satisfy

$$a_f(p^n) = \begin{cases} a_f(p^{n-1})a_f(p) - \chi(p)a_f(p^{n-2}) & \text{if } p \nmid N, \\ (a_f(p))^n & \text{if } p \mid N, \end{cases} \tag{8.10}$$

for all primes p and $n \geq 2$. Because $L(s,f)$ converges absolutely in the region $\sigma > 1$, multiplicativity of the Hecke eigenvalues implies

$$L(s,f) = \prod_p \left(\sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}} \right),$$

in this region. We now simplify the factor inside the product using this Equation (8.10). On the one hand, if $p \nmid N$:

$$\begin{aligned} \sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}} &= 1 + \frac{a_f(p)}{p^s} + \sum_{n\geq 2} \frac{a_f(p^n)}{p^{ns}} \\ &= 1 + \frac{a_f(p)}{p^s} + \sum_{n\geq 2} \frac{a_f(p^{n-1})a_f(p) - \chi(p)a_f(p^{n-2})}{p^{ns}} \\ &= 1 + \frac{a_f(p)}{p^s} + \frac{a_f(p)}{p^s} \sum_{n\geq 1} \frac{a_f(p^n)}{p^{ns}} - \frac{\chi(p)}{p^{2s}} \sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}} \\ &= 1 + \left(\frac{a_f(p)}{p^s} - \frac{\chi(p)}{p^{2s}} \right) \sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}}. \end{aligned}$$

By isolating the sum we find

$$\sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}} = \left(1 - \frac{a_f(p)}{p^s} + \frac{\chi(p)}{p^{2s}} \right)^{-1}.$$

On the other hand, if $p \mid N$ we have

$$\sum_{n\geq 0} \frac{a_f(p^n)}{p^{ns}} = \sum_{n\geq 0} \frac{(a_f(p))^n}{p^{ns}} = \left(1 - a_f(p)p^{-s}\right)^{-1}.$$

Therefore

$$L(s,f) = \prod_{p\nmid N} (1 - a_f(p)p^{-s} + \chi(p)p^{-2s})^{-1} \prod_{p\mid N} (1 - a_f(p)p^{-s})^{-1}.$$

If $p \nmid N$, let $\alpha_1(p)$ and $\alpha_2(p)$ be the roots of $1 - a_f(p)p^{-s} + \chi(p)p^{-2s}$. That is,

$$(1 - \alpha_1(p)p^{-s})(1 - \alpha_2(p)p^{-s}) = (1 - a_f(p)p^{-s} + \chi(p)p^{-2s}).$$

If $p \mid N$, let $\alpha_1(p) = a_f(p)$ and $\alpha_2(p) = 0$. We can then express $L(s,f)$ as a degree 2 Euler product:

$$L(s,f) = \prod_p (1 - \alpha_1(p)p^{-s})^{-1}(1 - \alpha_2(p)p^{-s})^{-1}.$$

The local factor at p is

$$L_p(s, f) = (1 - \alpha_1(p)p^{-s})^{-1}(1 - \alpha_2(p)p^{-s})^{-1},$$

with local roots $\alpha_1(p)$ and $\alpha_2(p)$. Upon applying partial fraction decomposition to the local factor, we find

$$\frac{1}{1-\alpha_1(p)p^{-s}}\frac{1}{1-\alpha_2(p)p^{-s}} = \frac{\frac{\alpha_1(p)}{\alpha_1(p)-\alpha_2(p)}}{1-\alpha_1(p)p^{-s}} + \frac{\frac{-\alpha_2(p)}{\alpha_1(p)-\alpha_2(p)}}{1-\alpha_2(p)p^{-s}}.$$

Expanding both sides as series in p^{-s}, and comparing coefficients gives

$$a_f(p^n) = \frac{\alpha_1(p)^{n+1} - \alpha_2(p)^{n+1}}{\alpha_1(p) - \alpha_2(p)}. \tag{8.11}$$

The Integral Representation

We now want to find an integral representation for $L(s, f)$. Consider the following Mellin transform:

$$\int_0^\infty f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y}.$$

As f has rapid decay at the cusps, this integral exists and defines an analytic function for $\sigma > 1$. In any case, we compute

$$\begin{aligned}
\int_0^\infty f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y} &= \int_0^\infty \sum_{n\geq 1} a_f(n)n^{\frac{k-1}{2}}e^{-2\pi ny}y^{s+\frac{k-1}{2}}\,\frac{dy}{y} \\
&= \sum_{n\geq 1} a_f(n)n^{\frac{k-1}{2}}\int_0^\infty e^{-2\pi ny}y^{s+\frac{k-1}{2}}\,\frac{dy}{y} && \text{FT} \\
&= \sum_{n\geq 1} \frac{a_f(n)}{(2\pi)^{s+\frac{k-1}{2}}n^s}\int_0^\infty e^{-y}y^{s+\frac{k-1}{2}}\,\frac{dy}{y} && y \to \frac{y}{2\pi n} \\
&= \frac{\Gamma\left(s+\frac{k-1}{2}\right)}{(2\pi)^{s+\frac{k-1}{2}}}\sum_{n\geq 1}\frac{a_f(n)}{n^s} \\
&= \frac{\Gamma\left(s+\frac{k-1}{2}\right)}{(2\pi)^{s+\frac{k-1}{2}}}L(s, f).
\end{aligned}$$

Rewriting, we have an integral representation

$$L(s, f) = \frac{(2\pi)^{s+\frac{k-1}{2}}}{\Gamma\left(s+\frac{k-1}{2}\right)}\int_0^\infty f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y}. \tag{8.12}$$

Now split the integral on the right-hand side into two pieces

$$\int_0^\infty f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y} = \int_0^{\frac{1}{\sqrt{N}}} f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y} + \int_{\frac{1}{\sqrt{N}}}^\infty f(iy)y^{s+\frac{k-1}{2}}\,\frac{dy}{y}. \tag{8.13}$$

Now we will rewrite the first piece in the same form and symmetrize the result as much as possible. Begin by performing the change of variables $y \to \frac{1}{Ny}$ to the first piece to obtain

$$\int_{\frac{1}{\sqrt{N}}}^\infty f\left(\frac{i}{Ny}\right)(Ny)^{-s-\frac{k-1}{2}}\,\frac{dy}{y}.$$

Rewriting in terms of the Atkin-Lehner operator and recalling that $\omega_N f = \omega_N(f)\overline{f}$ by Proposition 4.6.6, we have

$$\begin{aligned}
\int_{\frac{1}{\sqrt{N}}}^{\infty} f\left(\frac{i}{Ny}\right)(Ny)^{-s-\frac{k-1}{2}}\frac{dy}{y} &= \int_{\frac{1}{\sqrt{N}}}^{\infty} f\left(-\frac{1}{iNy}\right)(Ny)^{-s-\frac{k-1}{2}}\frac{dy}{y} \\
&= \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\sqrt{N}iy\right)^k (\omega_N f)(iy)(Ny)^{-s-\frac{k-1}{2}}\frac{dy}{y} \\
&= \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\sqrt{N}iy\right)^k \omega_N(f)\overline{f}(iy)(Ny)^{-s-\frac{k-1}{2}}\frac{dy}{y} \\
&= \omega_N(f)i^k N^{\frac{1}{2}-s}\int_{\frac{1}{\sqrt{N}}}^{\infty} \overline{f}(iy)y^{(1-s)-\frac{k-1}{2}}\frac{dy}{y}.
\end{aligned}$$

Substituting this result back into Equation (8.13) with Equation (8.12) yields the integral representation

$$L(s,f) = \frac{(2\pi)^{s+\frac{k-1}{2}}}{\Gamma\left(s+\frac{k-1}{2}\right)}\left[\omega_N(f)i^k N^{\frac{1}{2}-s}\int_{\frac{1}{\sqrt{N}}}^{\infty} \overline{f}(iy)y^{(1-s)+\frac{k-1}{2}}\frac{dy}{y} + \int_{\frac{1}{\sqrt{N}}}^{\infty} f(iy)y^{s+\frac{k-1}{2}}\frac{dy}{y}\right].$$

This integral representation will give analytic continuation. To see this, we know everything outside the brackets is entire. The integrands exhibit rapid decay and therefore the integrals are locally absolutely uniformly convergent on $\mathbb{C}$. Hence we have analytic continuation to all of $\mathbb{C}$. In particular, we have shown that $L(s,f)$ has no poles.

The Functional Equation

An immediate consequence of applying the symmetry $s \to 1-s$ to the integral representation is the following functional equation:

$$N^{\frac{s}{2}}\frac{\Gamma\left(s+\frac{k-1}{2}\right)}{(2\pi)^{s+\frac{k-1}{2}}}L(s,f) = \omega_N(f)i^k N^{\frac{1-s}{2}}\frac{\Gamma\left((1-s)+\frac{k-1}{2}\right)}{(2\pi)^{(1-s)+\frac{k-1}{2}}}L(1-s,\overline{f}).$$

Using the Legendre duplication formula for the gamma function we find that

$$\begin{aligned}
\frac{\Gamma\left(s+\frac{k-1}{2}\right)}{(2\pi)^{s+\frac{k-1}{2}}} &= \frac{1}{(2\pi)^{s+\frac{k-1}{2}}2^{1-\left(s+\frac{k-1}{2}\right)}\sqrt{\pi}}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right) \\
&= \frac{1}{2\pi^{s+\frac{1}{2}}}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right) \\
&= \frac{1}{\sqrt{4\pi}}\pi^{-s}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right).
\end{aligned}$$

The constant factor in front is independent of s and so can be canceled in the functional equation. Therefore we identify the gamma factor as

$$\gamma(s,f) = \pi^{-s}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right),$$

with $\kappa_1 = k-1$ and $\kappa_2 = k+1$ the local roots at infinity. The conductor is $q(f) = N$, so the primes dividing the level ramify, and by the Ramanujan-Petersson conjecture for holomorphic forms, $\alpha_1(p) \neq 0$ and $\alpha_2(p) \neq 0$ for all primes $p \nmid N$. The completed L-function is

$$\Lambda(s,f) = N^{\frac{s}{2}}\pi^{-s}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right)L(s,f),$$

with functional equation

$$\Lambda(s,f) = \omega_N(f)i^k\Lambda(1-s,\overline{f}).$$

This is the functional equation of $L(s,f)$. From it, the root number is $\varepsilon(f) = \omega_N(f)i^k$ and we see that $L(s,f)$ has dual $L(s,\overline{f})$. We will now show that $L(s,f)$ is of order 1. Since $L(s,f)$ has no poles, we do not need to clear any polar divisors. As the integrals in the representation is locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. Since the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s,f)} \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So the reciprocal of the gamma factor is also of order 1. Then

$$L(s,f) \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So $L(s,f)$ is of order 1. We summarize all of our work into the following theorem:

Theorem 8.4.1. *For any primitive Hecke eigenform $f \in \mathcal{S}_k(N,\chi)$, $L(s,f)$ is a Selberg class L-function. For $\sigma > 1$, it has a degree 2 Euler product given by*

$$L(s,f) = \prod_{p\nmid N}(1-a_f(p)p^{-s}+\chi(p)p^{-2s})^{-1}\prod_{p\mid N}(1-a_f(p)p^{-s})^{-1}.$$

Moreover, it admits analytic continuation to $\mathbb{C}$ and possesses the functional equation

$$N^{\frac{s}{2}}\pi^{-s}\Gamma\left(\frac{s+\frac{k-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{k+1}{2}}{2}\right)L(s,f) = \Lambda(s,f) = \omega_N(f)i^k\Lambda(1-s,\overline{f}).$$

Beyond Primitivity

We can still obtain analytic continuation of the L-series $L(s,f)$ if f is not a primitive Hecke eigenform. Indeed, since the primitive Hecke eigenforms form a basis for the space of newforms, we can prove the following:

Theorem 8.4.2. *For any $f \in \mathcal{S}_k(\Gamma_1(N))$, $L(s,f)$ admits analytic continuation to $\mathbb{C}$.*

Proof. If f is a newform, this follows from Theorems 4.6.1 and 8.4.1. We will prove the case for oldforms by induction. The base case is clear since if $N=1$ there are no oldforms. So assume by induction that the claim holds for all proper divisors of N. As f is an oldform, there is a proper divisor $d \mid N$ with $d > 1$ such that

$$f(z) = g(z) + d^{k-1}h(dz) = g(z) + \prod_{p^r\,||\,d}(V_p^r h)(z),$$

for some $g, h \in \mathcal{S}_k\left(\Gamma_1\left(\frac{N}{d}\right)\right)$. Note that $V_p h \in \mathcal{S}_k\left(\Gamma_1\left(\frac{Np}{d}\right)\right)$ by Lemma 4.8.1. Our induction hypothesis then implies that $L(s,g)$ and $L(s,V_p^r h)$, for all $p^r \mid\mid d$, admit analytic continuation to $\mathbb{C}$. Thus so does $L(s,f)$ which completes the proof. $\square$

8.5 Hecke-Maass L-functions

The Definition & Euler Product

We will investigate the L-functions of weight zero Maass cusp forms. Let $f \in \mathcal{C}_\nu(N,\chi)$ and denote its Fourier-Whittaker series by

$$f(z) = \sum_{n\geq 1} a_f(n)\left(\sqrt{y}K_\nu(2\pi ny)e^{2\pi inx} + \frac{a_f(-n)}{a_f(n)}\sqrt{y}K_\nu(2\pi ny)e^{-2\pi inx}\right),$$

with $a_f(1) = 1$. Thus if f is a Hecke eigenform, then as f is even or odd by Proposition 5.7.5, the Fourier-Whittaker series takes the form

$$f(z) = \sum_{n\geq 1} a_f(n)\sqrt{y}K_\nu(2\pi ny)\,\mathrm{SC}(2\pi nx),$$

and the $a_f(n)$ are the Hecke eigenvalues of f normalized so that they are constant on average. The **Hecke-Maass L-series** (respectively **Hecke-Maass L-function** if it is an L-function) $L(s,f)$ of f is defined by the following Dirichlet series:

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

We will see that $L(s,f)$ is a Selberg class L-function if f is a primitive Hecke-Maass eigenform. From now on, we make this assumption about f. The Ramanujan-Petersson conjecture for Maass forms is not know so $L(s,f)$ has not been proven to be a Selberg class L-function. Although, it is conjectured to be, so throughout we will make this additional assumption. As we have noted, the Hecke relations and the Ramanujan-Petersson conjecture for Maass forms together imply $a_f(n) \ll_\varepsilon n^\varepsilon$. Therefore $L(s,f)$ is locally absolutely uniformly convergent for $\sigma > 1$. The L-function will have an Euler product. Indeed, the Hecke relations imply that the coefficients $a_f(n)$ are multiplicative and satisfy

$$a_f(p^n) = \begin{cases} a_f(p^{n-1})a_f(p) - \chi(p)a_f(p^{n-2}) & \text{if } p \nmid N, \\ (a_f(p))^n & \text{if } p \mid N, \end{cases} \tag{8.14}$$

for all primes p and $n \geq 2$. Because $L(s,f)$ converges absolutely in the region $\sigma > 1$, multiplicativity of the Hecke eigenvalues implies

$$L(s,f) = \prod_p \left(\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}}\right),$$

in this region. We now simplify the factor inside the product using this Equation (8.14). On the one hand, if $p \nmid N$:

$$\begin{aligned}
\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}} &= 1 + \frac{a_f(p)}{p^s} + \sum_{n\geq 2}\frac{a_f(p^n)}{p^{ns}} \\
&= 1 + \frac{a_f(p)}{p^s} + \sum_{n\geq 2}\frac{a_f(p^{n-1})a_f(p) - \chi(p)a_f(p^{n-2})}{p^{ns}} \\
&= 1 + \frac{a_f(p)}{p^s} + \frac{a_f(p)}{p^s}\sum_{n\geq 1}\frac{a_f(p^n)}{p^{ns}} - \frac{\chi(p)}{p^{2s}}\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}} \\
&= 1 + \left(\frac{a_f(p)}{p^s} - \frac{\chi(p)}{p^{2s}}\right)\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}}.
\end{aligned}$$

By isolating the sum we find

$$\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}}=\left(1-\frac{a_f(p)}{p^s}+\frac{\chi(p)}{p^{2s}}\right)^{-1}.$$

On the other hand, if $p\mid N$ we have

$$\sum_{n\geq 0}\frac{a_f(p^n)}{p^{ns}}=\sum_{n\geq 0}\frac{(a_f(p))^n}{p^{ns}}=\left(1-a_f(p)p^{-s}\right)^{-1}.$$

Therefore

$$L(s,f)=\prod_{p\nmid N}(1-a_f(p)p^{-s}+\chi(p)p^{-2s})^{-1}\prod_{p\mid N}(1-a_f(p)p^{-s})^{-1}.$$

If $p\nmid N$, let $\alpha_1(p)$ and $\alpha_2(p)$ be the roots of $1-a_f(p)p^{-s}+\chi(p)p^{-2s}$. That is,

$$(1-\alpha_1(p)p^{-s})(1-\alpha_2(p)p^{-s})=(1-a_f(p)p^{-s}+\chi(p)p^{-2s}).$$

If $p\mid N$, let $\alpha_1(p)=a_f(p)$ and $\alpha_2(p)=0$. We can then express $L(s,f)$ as a degree 2 Euler product:

$$L(s,f)=\prod_p(1-\alpha_1(p)p^{-s})^{-1}(1-\alpha_2(p)p^{-s})^{-1}.$$

The local factor at p is

$$L_p(s,f)=(1-\alpha_1(p)p^{-s})^{-1}(1-\alpha_2(p)p^{-s})^{-1},$$

with local roots $\alpha_1(p)$ and $\alpha_2(p)$. Upon applying partial fraction decomposition to the local factor, we find

$$\frac{1}{1-\alpha_1(p)p^{-s}}\frac{1}{1-\alpha_2(p)p^{-s}}=\frac{\frac{\alpha_1(p)}{\alpha_1(p)-\alpha_2(p)}}{1-\alpha_1(p)p^{-s}}+\frac{\frac{-\alpha_2(p)}{\alpha_1(p)-\alpha_2(p)}}{1-\alpha_2(p)p^{-s}}.$$

Expanding both sides as series in p^{-s}, and comparing coefficients gives

$$a_f(p^n)=\frac{\alpha_1(p)^{n+1}-\alpha_2(p)^{n+1}}{\alpha_1(p)-\alpha_2(p)}. \tag{8.15}$$

The Integral Representation

We want to find an integral representation for $L(s,f)$. Recall that f is an eigenfunction for the parity Hecke operator T_{-1} with eigenvalue ± 1. Equivalently, f is even if the eigenvalue is 1 and odd if the eigenvalue is -1. The integral representation will depend upon this parity. To handle both cases simultaneously, let $\mathfrak{a}=0,1$ according to whether f is even or odd. In other words,

$$\mathfrak{a}=\frac{1-a_f(-1)}{2}.$$

Now consider the following Mellin transform:

$$\int_0^\infty\left(\frac{\partial}{\partial x}^{\mathfrak{a}}f\right)(iy)y^{s-\frac{1}{2}+\mathfrak{a}}\frac{dy}{y}.$$

As f has rapid decay at the cusps, this integral exists and defines an analytic function for $\sigma > 1$. The derivative operator is present because if f is odd, $\mathrm{SC}(x) = i\sin(x)$. In any case, the smoothness of f implies that we may differentiate its Fourier-Whittaker series termwise to obtain

$$\left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(z) = \sum_{n \geq 1} a_f(n)(2\pi i n)^{\mathfrak{a}} \sqrt{y} K_\nu(2\pi n y)\cos(2\pi n x).$$

Therefore regardless if f is even or odd, the Fourier-Whittaker series of $\left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(z)$ has $\mathrm{SC}(x) = \cos(x)$ and the integral does not vanish identically. We compute

$$\begin{aligned}
\int_0^\infty \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy) y^{s-\frac{1}{2}+\mathfrak{a}} \frac{dy}{y} &= \int_0^\infty \sum_{n\geq 1} a_f(n)(2\pi i n)^{\mathfrak{a}} K_\nu(2\pi n y) y^{s+\mathfrak{a}} \frac{dy}{y} \\
&= \sum_{n\geq 1} a_f(n)(2\pi i n)^{\mathfrak{a}} \int_0^\infty K_\nu(2\pi n y) y^{s+\mathfrak{a}} \frac{dy}{y} && \text{FT} \\
&= \sum_{n\geq 1} \frac{a_f(n)}{(2\pi)^s n^s} i^{\mathfrak{a}} \int_0^\infty K_\nu(y) y^{s+\mathfrak{a}} \frac{dy}{y} && y \to \frac{y}{2\pi n} \\
&= \frac{\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)}{2^{2-\mathfrak{a}}\pi^s(-i)^{\mathfrak{a}}} \sum_{n\geq 1} \frac{a_f(n)}{n^s} && \text{Appendix E.1} \\
&= \frac{\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)}{2^{2-\mathfrak{a}}\pi^s(-i)^{\mathfrak{a}}} L(s,f).
\end{aligned}$$

This last expression is analytic function for $\sigma > 1$ and so the integral is too. Rewriting, we have an integral representation

$$L(s,f) = \frac{2^{2-\mathfrak{a}}\pi^s(-i)^{\mathfrak{a}}}{\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)} \int_0^\infty \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy) y^{s-\frac{1}{2}+\mathfrak{a}} \frac{dy}{y}. \tag{8.16}$$

Now split the integral on the right-hand side into two pieces

$$\int_0^\infty \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy) y^{s-\frac{1}{2}+\mathfrak{a}} \frac{dy}{y} = \int_0^{\frac{1}{\sqrt{N}}} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy) y^{s-\frac{1}{2}+\mathfrak{a}} \frac{dy}{y} + \int_{\frac{1}{\sqrt{N}}}^\infty \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy) y^{s-\frac{1}{2}+\mathfrak{a}} \frac{dy}{y}. \tag{8.17}$$

Now we will rewrite the first piece in the same form and symmetrize the result as much as possible. Performing the change of variables $y \to \frac{1}{Ny}$ to the first piece to obtain

$$\int_{\frac{1}{\sqrt{N}}}^\infty \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)\left(\frac{i}{Ny}\right)(Ny)^{-s+\frac{1}{2}-\mathfrak{a}} \frac{dy}{y}.$$

We will rewrite this in terms of the Atkin-Lehner operator. But first we require an identity that relates $\frac{\partial}{\partial x}^{\mathfrak{a}}$ with the Atkin-Lehner operator ω_N. By the identity theorem it suffices verify this for $z \in \mathbb{H}$ with $|z|$

fixed. Observe that $-\frac{1}{Nz} = \frac{-x}{N|z|^2} + \frac{iy}{N|z|^2}$. Now differentiate termwise to see that

$$\begin{aligned}
\left(\frac{\partial}{\partial x}^{\mathfrak{a}} \omega_N f\right)(z) &= \left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right) f\left(-\frac{1}{Nz}\right) \\
&= \left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right) \sum_{n\geq 1} a_f(n)\sqrt{\frac{y}{N|z|^2}} K_\nu(2\pi n y)\,\mathrm{SC}\left(-2\pi n \frac{x}{N|z|^2}\right) \\
&= (-N|z|^2)^{-\mathfrak{a}} \sum_{n\geq 1} a_f(n)(2\pi i n)^{\mathfrak{a}} \sqrt{\frac{y}{N|z|^2}} K_\nu\left(2\pi n \frac{y}{N|z|^2}\right)\cos\left(-2\pi n \frac{x}{N|z|^2}\right) \\
&= (-N|z|^2)^{-\mathfrak{a}} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)\left(-\frac{1}{Nz}\right).
\end{aligned}$$

By the identity theorem, we have

$$\left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)\left(-\frac{1}{Nz}\right) = (-N|z|^2)^{\mathfrak{a}} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} \omega_N f\right)(z),$$

for all $z \in \mathbb{H}$. Rewriting in terms of the Atkin-Lehner operator and recalling that $\omega_N f = \omega_N(f)\overline{f}$ by Proposition 5.7.7, we find that

$$\begin{aligned}
\int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)\left(\frac{i}{Ny}\right)(Ny)^{-s+\frac{1}{2}-\mathfrak{a}}\frac{dy}{y} &= \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)\left(-\frac{1}{iNy}\right)(Ny)^{-s+\frac{1}{2}-\mathfrak{a}}\frac{dy}{y} \\
&= \int_{\frac{1}{\sqrt{N}}}^{\infty} (-Ny^2)^{\mathfrak{a}} \left(\left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right)\omega_N f\right)(iy)(Ny)^{-s+\frac{1}{2}-\mathfrak{a}}\frac{dy}{y} \\
&= \int_{\frac{1}{\sqrt{N}}}^{\infty} (-Ny^2)^{\mathfrak{a}}\omega_N(f) \left(\left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right)\overline{f}\right)(iy)(Ny)^{-s+\frac{1}{2}-\mathfrak{a}}\frac{dy}{y} \\
&= w_N(f)(-1)^{\mathfrak{a}} N^{\frac{1}{2}-s} \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right)\overline{f}\right)(iy)y^{(1-s)-\frac{1}{2}+\mathfrak{a}}\frac{dy}{y}.
\end{aligned}$$

Substituting this result back into Equation (8.17) with Equation (8.16) gives the integral representation

$$\begin{aligned}
L(s,f) &= \frac{2^{2-\mathfrak{a}}\pi^s(-i)^{\mathfrak{a}}}{\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)} \\
&\quad \cdot \left[w_N(f)(-1)^{\mathfrak{a}} N^{\frac{1}{2}-s} \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\left(\frac{\partial}{\partial x}^{\mathfrak{a}}\right)\overline{f}\right)(iy)y^{(1-s)-\frac{1}{2}+\mathfrak{a}}\frac{dy}{y} + \int_{\frac{1}{\sqrt{N}}}^{\infty} \left(\frac{\partial}{\partial x}^{\mathfrak{a}} f\right)(iy)y^{s-\frac{1}{2}+\mathfrak{a}}\frac{dy}{y}\right].
\end{aligned}$$

This integral representation will give analytic continuation. To see this, note that everything outside the brackets is entire. The integrands exhibit rapid decay and therefore the integrals are locally absolutely uniformly convergent on $\mathbb{C}$. Hence we have analytic continuation to all of $\mathbb{C}$. In particular, $L(s,f)$ has no poles.

The Functional Equation

An immediate consequence of applying the symmetry $s \to 1-s$ to the integral representation is the following functional equation:

$$N^{\frac{s}{2}} \frac{\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)}{2^{2-\mathfrak{a}}\pi^s(-i)^{\mathfrak{a}}} L(s,f) = \omega_N(f)(-1)^{\mathfrak{a}} N^{\frac{1-s}{2}} \frac{\Gamma\left(\frac{(1-s)+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{(1-s)+\mathfrak{a}-\nu}{2}\right)}{2^{2-\mathfrak{a}}\pi^{1-s}(-i)^{\mathfrak{a}}} L(1-s,\overline{f}).$$

The constant factor in the denominator is independent of s and so can be canceled in the functional equation. Therefore we identify the gamma factor as

$$\gamma(s,f) = \pi^{-s}\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right),$$

with $\kappa_1 = \mathfrak{a}+\nu$ and $\kappa_2 = \mathfrak{a}-\nu$ the local roots at infinity (these are complex conjugates because ν is either purely imaginary of real). The conductor is $q(f) = N$, so the primes dividing the level ramify, and by the Ramanujan-Petersson conjecture for Maass forms, $\alpha_1(p) \neq 0$ and $\alpha_2(p) \neq 0$ for all primes $p \nmid N$. The completed L-function is

$$\Lambda(s,f) = N^{\frac{s}{2}}\pi^{-s}\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)L(s,f),$$

with functional equation

$$\Lambda(s,f) = \omega_N(f)(-1)^{\mathfrak{a}}\Lambda(1-s,\overline{f}).$$

This is the functional equation of $L(s,f)$. From it, the root number is $\varepsilon(f) = \omega_N(f)(-1)^{\mathfrak{a}}$ and we see that $L(s,f)$ has dual $L(s,\overline{f})$. We will now show that $L(s,f)$ is of order 1. Since $L(s,f)$ has no poles, we do not need to clear any polar divisors. As the integrals in the representation is locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. Since the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s,f)} \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So the reciprocal of the gamma factor is also of order 1. Then

$$L(s,f) \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So $L(s,f)$ is of order 1. We summarize all of our work into the following theorem:

Theorem 8.5.1. *For any primitive Hecke-Maass eigenform $f \in \mathcal{C}_\nu(N,\chi)$, $L(s,f)$ is a Selberg class L-function provided the Ramanujan-Petersson conjecture for Maass forms holds. For $\sigma > 1$, it has a degree 2 Euler product given by*

$$L(s,f) = \prod_{p\nmid N}(1-a_f(p)p^{-s}+\chi(p)p^{-2s})^{-1}\prod_{p\mid N}(1-a_f(p)p^{-s})^{-1}.$$

Moreover, it admits analytic continuation to $\mathbb{C}$ and possesses the functional equation

$$N^{\frac{s}{2}}\pi^{-s}\Gamma\left(\frac{s+\mathfrak{a}+\nu}{2}\right)\Gamma\left(\frac{s+\mathfrak{a}-\nu}{2}\right)L(s,f) = \Lambda(s,f) = \omega_N(f)(-1)^{\mathfrak{a}}\Lambda(1-s,\overline{f}).$$

Beyond Primitivity

We can still obtain analytic continuation of the L-series $L(s,f)$ if f is not a primitive Hecke-Maass eigenform. Similarly to the Hecke L-function case, this holds because the primitive Hecke-Maass eigenforms form a basis for the space of newforms:

Theorem 8.5.2. *For any $f \in \mathcal{C}_\nu(\Gamma_1(N))$, $L(s,f)$ admits analytic continuation to $\mathbb{C}$.*

Proof. Argue as in the proof of Theorem 8.4.2. $\square$

8.6 The Rankin-Selberg Method

The Definition & Euler Product

The Rankin-Selberg method is a process by which we can construct new L-functions from old ones. Instead of giving the general definition outright, we first provide a full discussion of the method only in the simplest case. Many technical difficulties arise in the fully general setting. Let $f, g \in \mathcal{S}_k(1)$ be primitive Hecke eigenforms with Fourier series

$$f(z) = \sum_{n \ge 1} a_f(n) n^{\frac{k-1}{2}} e^{2\pi i n z} \quad \text{and} \quad g(z) = \sum_{n \ge 1} a_g(n) n^{\frac{k-1}{2}} e^{2\pi i n z}.$$

The L-function $L(s, f \times g)$ of f and g is given by the L-series

$$L(s, f \times g) = \sum_{n \ge 1} \frac{a_{f \times g}(n)}{n^s} = \sum_{n \ge 1} \frac{a_f(n)\overline{a_g(n)}}{n^s} = \sum_{n \ge 1} \frac{a_f(n)\overline{a_g(n)}}{n^s},$$

The **Rankin-Selberg convolution** $L(s, f \otimes g)$ of f and g is defined by

$$L(s, f \otimes g) = \sum_{n \ge 1} \frac{a_{f \otimes g}(n)}{n^s} = \zeta(2s) L(s, f \times g),$$

where $a_{f \otimes g}(n) = \sum_{n = m\ell^2} a_f(m)\overline{a_g(m)}$. Since $a_f(n) \ll_\varepsilon n^\varepsilon$ and $a_g(n) \ll_\varepsilon n^\varepsilon$, $a_{f \times g}(n) \ll_\varepsilon n^\varepsilon$ as well. Hence $L(s, f \times g)$ is locally absolutely uniformly convergent for $\sigma > 1$. Since $\zeta(2s)$ is also locally absolutely uniformly convergent in this region, the same follows for $L(s, f \times g)$ too. The L-function $L(s, f \times g)$ will also have an Euler product. To see this, let $\alpha_j(p)$ and $\beta_\ell(p)$ be the local roots at p of $L(s, f)$ and $L(s, g)$ respectively. Since $L(s, f \otimes g)$ converges absolutely in the region $\sigma > 1$, multiplicativity of the Hecke eigenvalues implies

$$L(s, f \otimes g) = \zeta(2s) L(s, f \times g) = \prod_{p \nmid NM} (1 - p^{-2s})^{-1} \prod_p \left(\sum_{n \ge 0} \frac{a_f(p^n)\overline{a_g(p^n)}}{p^{ns}} \right),$$

in this region. We now simplify the factor inside the latter product using Equation (8.11):

$$\begin{aligned}
\sum_{n \ge 0} \frac{a_f(p^n)\overline{a_g(p^n)}}{p^{ns}} &= \sum_{n \ge 0} \left(\frac{\alpha_1(p)^{n+1} - \alpha_2(p)^{n+1}}{\alpha_1(p) - \alpha_2(p)} \right) \left(\frac{(\overline{\beta_1(p)})^{n+1} - (\overline{\beta_2(p)})^{n+1}}{\overline{\beta_1(p)} - \overline{\beta_2(p)}} \right) p^{-ns} \\
&= (\alpha_1(p) - \alpha_2(p))^{-1} \left(\overline{\beta_1(p)} - \overline{\beta_2(p)} \right)^{-1} \\
&\cdot \left[\sum_{n \ge 1} \frac{\alpha_1(p)^n (\overline{\beta_1(p)})^n}{p^{(n-1)s}} + \frac{\alpha_2(p)^n (\overline{\beta_2(p)})^n}{p^{(n-1)s}} - \frac{\alpha_1(p)^n (\overline{\beta_2(p)})^n}{p^{(n-1)s}} - \frac{\alpha_2(p)^n (\overline{\beta_1(p)})^n}{p^{(n-1)s}} \right] \\
&= (\alpha_1(p) - \alpha_2(p))^{-1} \left(\overline{\beta_1(p)} - \overline{\beta_2(p)} \right)^{-1} \left[\alpha_1(p)\overline{\beta_1(p)} \left(1 - \alpha_1(p)\overline{\beta_1(p)} p^{-s} \right)^{-1} \right. \\
&+ \alpha_2(p)\overline{\beta_2(p)} \left(1 - \alpha_2(p)\overline{\beta_2(p)} p^{-s} \right)^{-1} - \alpha_1(p)\overline{\beta_2(p)} \left(1 - \alpha_1(p)\overline{\beta_2(p)} p^{-s} \right)^{-1} \\
&\left. - \alpha_2(p)\overline{\beta_1(p)} \left(1 - \alpha_2(p)\overline{\beta_1(p)} p^{-s} \right)^{-1} \right]
\end{aligned}$$

$$
\begin{aligned}
&= (\alpha_1(p) - \alpha_2(p))^{-1} \left(\overline{\beta_1(p)} - \overline{\beta_2(p)}\right)^{-1} \left(1 - \alpha_1(p)\overline{\beta_1(p)}p^{-s}\right)^{-1} \\
&\cdot \left(1 - \alpha_2(p)\overline{\beta_2(p)}p^{-s}\right)^{-1} \left(1 - \alpha_1(p)\overline{\beta_2(p)}p^{-s}\right)^{-1} \left(1 - \alpha_2(p)\overline{\beta_1(p)}p^{-s}\right)^{-1} \\
&\cdot \Big[\alpha_1(p)\overline{\beta_1(p)} \left(1 - \alpha_2(p)\overline{\beta_2(p)}p^{-s}\right) \left(1 - \alpha_1(p)\overline{\beta_2(p)}p^{-s}\right) \left(1 - \alpha_2(p)\overline{\beta_1(p)}p^{-s}\right) \\
&+ \alpha_2(p)\overline{\beta_2(p)} \left(1 - \alpha_1(p)\overline{\beta_1(p)}p^{-s}\right) \left(1 - \alpha_1(p)\overline{\beta_2(p)}p^{-s}\right) \left(1 - \alpha_2(p)\overline{\beta_1(p)}p^{-s}\right) \\
&- \alpha_1(p)\overline{\beta_2(p)} \left(1 - \alpha_1(p)\overline{\beta_1(p)}p^{-s}\right) \left(1 - \alpha_2(p)\overline{\beta_2(p)}p^{-s}\right) \left(1 - \alpha_2(p)\overline{\beta_1(p)}p^{-s}\right) \\
&- \alpha_2(p)\overline{\beta_1(p)} \left(1 - \alpha_1(p)\overline{\beta_1(p)}p^{-s}\right) \left(1 - \alpha_2(p)\overline{\beta_2(p)}p^{-s}\right) \left(1 - \alpha_1(p)\overline{\beta_2(p)}p^{-s}\right) \Big].
\end{aligned}
$$

The term in the brackets simplifies to

$$\left(1 - \alpha_1(p)\alpha_2(p)\overline{\beta_1(p)\beta_2(p)}p^{-2s}\right) (\alpha_1(p) - \alpha_2(p)) \left(\overline{\beta_1(p)} - \overline{\beta_2(p)}\right),$$

because all of the other terms are annihilated by symmetry in $\alpha_1(p)$, $\alpha_2(p)$, $\overline{\beta_1(p)}$, and $\overline{\beta_2(p)}$. The Ramanujan-Petersson conjecture for holomorphic forms implies $\alpha_1(p)\alpha_2(p)\overline{\beta_1(p)\beta_2(p)} = 1$. Therefore the corresponding factor above is $(1 - p^{-2s})$. This factor cancels the local factor at p in the Euler product of $\zeta(2s)$, so that

$$\sum_{n \geq 0} \frac{a_f(p^n)\overline{a_g(p^n)}}{p^{ns}} = \prod_{1 \leq j, \ell \leq 2} \left(1 - \alpha_j(p)\overline{\beta_\ell(p)}p^{-s}\right)^{-1}.$$

Hence

$$L(s, f \otimes g) = \prod_p \left(\sum_{n \geq 0} \frac{a_f(p^n)\overline{a_g(p^n)}}{p^{ns}}\right).$$

In total we have a degree 4 Euler product:

$$L(s, f \otimes g) = \prod_p \prod_{1 \leq j, \ell \leq 2} \left(1 - \alpha_j(p)\overline{\beta_\ell(p)}p^{-s}\right)^{-1}.$$

The local factor at p is

$$L_p(s, f \otimes g) = \prod_{1 \leq j, \ell \leq 2} \left(1 - \alpha_j(p)\overline{\beta_\ell(p)}p^{-s}\right)^{-1},$$

with local roots $\alpha_j(p)\overline{\beta_\ell(p)}$.

The Integral Representation: Part I

We now look for an integral representation for $L(s, f \otimes g)$. Consider the following integral:

$$\int_{\Gamma_\infty \backslash \mathbb{H}} f(z)\overline{g(z)}\mathrm{Im}(z)^{s+k} \, d\mu.$$

This will turn out to be a Mellin transform as we will soon see. Since f and g have rapid decay, this integral exists and defines an analytic function for $\sigma > 1$. We have

$$\begin{aligned}
\int_{\Gamma_\infty\backslash\mathbb{H}} f(z)\overline{g(z)}\mathrm{Im}(z)^{s+k}\,d\mu &= \int_0^\infty \int_0^1 f(x+iy)\overline{g(x+iy)}y^{s+k}\,\frac{dx\,dy}{y^2} \\
&= \int_0^\infty \int_0^1 \sum_{n,m\geq 1} a_f(n)\overline{a_g(m)}(nm)^{\frac{k-1}{2}}e^{2\pi i(n-m)x}e^{-2\pi(n+m)y}y^{s+k}\,\frac{dx\,dy}{y^2} \\
&= \int_0^\infty \sum_{n,m\geq 1} \int_0^1 a_f(n)\overline{a_g(m)}(nm)^{\frac{k-1}{2}}e^{2\pi i(n-m)x}e^{-2\pi(n+m)y}y^{s+k}\,\frac{dx\,dy}{y^2} \qquad \text{FT} \\
&= \int_0^\infty \sum_{n\geq 1} a_f(n)\overline{a_g(n)}n^{k-1}e^{-4\pi ny}y^{s+k}\,\frac{dy}{y^2},
\end{aligned}$$

where the last line follows by Equation (4.1). Observe that this last integral is a Mellin transform. The rest is a computation:

$$\begin{aligned}
\int_0^\infty \sum_{n\geq 1} a_f(n)\overline{a_g(n)}n^{k-1}e^{-4\pi ny}y^{s+k}\,\frac{dy}{y^2} &= \sum_{n\geq 1} a_f(n)\overline{a_g(n)}n^{k-1}\int_0^\infty e^{-4\pi ny}y^{s+k}\,\frac{dy}{y^2} \qquad \text{FT} \\
&= \sum_{n\geq 1} \frac{a_f(n)\overline{a_g(n)}}{(4\pi)^{s+k-1}n^s}\int_0^\infty e^{-y}y^{s+k-1}\,\frac{dy}{y} \qquad y \to \frac{y}{4\pi n} \\
&= \frac{\Gamma(s+k-1)}{(4\pi)^{s+k-1}}\sum_{n\geq 1}\frac{a_f(n)\overline{a_g(n)}}{n^s} \\
&= \frac{\Gamma(s+k-1)}{(4\pi)^{s+k-1}}L(s, f\times g).
\end{aligned}$$

Rewriting, we have an integral representation

$$L(s, f\times g) = \frac{(4\pi)^{s+k-1}}{\Gamma(s+k-1)}\int_{\Gamma_\infty\backslash\mathbb{H}} f(z)\overline{g(z)}\mathrm{Im}(z)^{s+k}\,d\mu.$$

We rewrite the integral as follows:

$$\begin{aligned}
\int_{\Gamma_\infty\backslash\mathbb{H}} f(z)\overline{g(z)}\mathrm{Im}(z)^{s+k}\,d\mu &= \int_{\mathcal{F}} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} f(\gamma z)\overline{g(\gamma z)}\mathrm{Im}(\gamma z)^{s+k}\,d\mu \qquad \text{folding} \\
&= \int_{\mathcal{F}} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} j(\gamma,z)^k\overline{j(\gamma,z)^k}f(z)\overline{g(z)}\mathrm{Im}(\gamma z)^{s+k}\,d\mu \qquad \text{modularity} \\
&= \int_{\mathcal{F}} f(z)\overline{g(z)}\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} |j(\gamma,z)|^{2k}\mathrm{Im}(\gamma z)^{s+k}\,d\mu \\
&= \int_{\mathcal{F}} f(z)\overline{g(z)}\mathrm{Im}(z)^k\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \mathrm{Im}(\gamma z)^{s}\,d\mu \\
&= \int_{\mathcal{F}} f(z)\overline{g(z)}\mathrm{Im}(z)^k E(z,s)\,d\mu.
\end{aligned}$$

Note that $E(z,s)$ is the weight zero Eisenstein series on $\Gamma_1(1)\backslash\mathbb{H}$ at the ∞ cusp. Altogether, this gives the integral representation

$$L(s, f\times g) = \frac{(4\pi)^{s+k-1}}{\Gamma(s+k-1)}\int_{\mathcal{F}} f(z)\overline{g(z)}\mathrm{Im}(z)^k E(z,s)\,d\mu. \tag{8.18}$$

We cannot investigate the integral any further until we understand the Fourier-Whittaker series of $E(z,s)$ and have a functional equation as $s\to 1-s$. Therefore we will take a necessary detour and return to the integral after.

The Integral Representation: Part II

We will compute the Fourier-Whittaker series of $E(z,s)$. To do this we will need the following technical lemma:

Lemma 8.6.1. *For $\sigma>1$ and $b\in\mathbb{Z}$,*

$$\sum_{m\geq 1}\frac{r(b,m)}{m^{2s}} = \begin{cases} \frac{\zeta(2s-1)}{\zeta(2s)} & \text{if } b=0,\\ \frac{\sigma_{1-2s}(|b|)}{\zeta(2s)} & \text{if } b\neq 0,\end{cases}$$

where $\sigma_s(b)$ is the generalized sum of divisors function.

Proof. If $\sigma>1$ then the desired evaluation of the sum is locally absolutely uniformly convergent because the Riemann zeta function is in that region. Hence the sum will be too provided we prove the identity. Suppose $b=0$. Then $r(0,m)=\varphi(m)$. Since $\varphi(m)$ is multiplicative we have

$$\sum_{m\geq 1}\frac{\varphi(m)}{m^{2s}} = \prod_p\left(\sum_{k\geq 0}\frac{\varphi(p^k)}{p^{k(2s)}}\right). \tag{8.19}$$

Recalling that $\varphi(p^k)=p^k-p^{k-1}$ for $k\geq 1$, make the following computation:

$$\begin{aligned}\sum_{k\geq 0}\frac{\varphi(p^k)}{p^{k(2s)}} &= 1+\sum_{k\geq 1}\frac{p^k-p^{k-1}}{p^{k(2s)}}\\ &= \sum_{k\geq 0}\frac{1}{p^{k(2s-1)}} - \frac{1}{p}\sum_{k\geq 1}\frac{1}{p^{k(2s-1)}}\\ &= \sum_{k\geq 0}\frac{1}{p^{k(2s-1)}} - p^{-2s}\sum_{k\geq 0}\frac{1}{p^{k(2s-1)}}\\ &= (1-p^{-2s})\sum_{k\geq 0}\frac{1}{p^{k(2s-1)}}\\ &= \frac{1-p^{-2s}}{1-p^{-(2s-1)}}.\end{aligned} \tag{8.20}$$

Combining Equations (8.19) and (8.20) gives

$$\sum_{m\geq 1}\frac{\varphi(m)}{m^{2s}} = \frac{\zeta(2s-1)}{\zeta(2s)}.$$

Now suppose $b \neq 0$, Proposition 1.4.1 gives the first equality in the following chain:

$$\begin{aligned}
\sum_{m \geq 1} \frac{r(b,m)}{m^{2s}} &= \sum_{m \geq 1} m^{-2s} \sum_{\ell \mid (b,m)} \ell \mu\left(\frac{m}{\ell}\right) \\
&= \sum_{\ell \mid b} \ell \sum_{m \geq 1} \frac{\mu(m)}{(m\ell)^{2s}} \\
&= \left(\sum_{\ell \mid b} \ell^{1-2s}\right)\left(\sum_{m \geq 1} \frac{\mu(m)}{m^{2s}}\right) \\
&= \sigma_{1-2s}(b) \sum_{m \geq 1} \frac{\mu(m)}{m^{2s}} \\
&= \sigma_{1-2s}(|b|) \sum_{m \geq 1} \frac{\mu(m)}{m^{2s}} \\
&= \frac{\sigma_{1-2s}(|b|)}{\zeta(2s)} \qquad \text{Proposition A.2.2.}
\end{aligned}$$

□

We can now compute the Fourier-Whittaker series of $E(z,s)$:

Proposition 8.6.1. *The Fourier-Whittaker series of $E(z,s)$ is given by*

$$E(z,s) = y^s + y^{1-s}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)\zeta(2s-1)}{\Gamma(s)\zeta(2s)} + \sum_{t \geq 1}\left(\frac{2\pi^s|t|^{s-\frac{1}{2}}\sigma_{1-2s}(|t|)}{\Gamma(s)\zeta(2s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi i t x}.$$

Proof. Fix s with $\sigma > 1$. By the Bruhat decomposition for $\Gamma_1(1)$ and Remark 3.2.3, we have

$$E(z,s) = \mathrm{Im}(z)^s + \sum_{\substack{c \geq 1, d \in \mathbb{Z} \\ (c,d)=1}} \frac{\mathrm{Im}(z)^s}{|cz+d|^{2s}}.$$

Summing over all pairs $(c,d) \in \mathbb{Z}^2 - \{\mathbf{0}\}$ with $c \geq 1$, $d \in \mathbb{Z}$, and $(c,d) = 1$ is the same as summing over all triples (c,ℓ,r) with $c \geq 1$, $\ell \in \mathbb{Z}$, r taken modulo c, and $(r,c) = 1$. This is seen by writing $d = c\ell + r$. Therefore

$$\sum_{\substack{c \geq 1, d \in \mathbb{Z} \\ (c,d)=1}} \frac{\mathrm{Im}(z)^s}{|cx+icy+d|^{2s}} = \sum_{(c,\ell,r)} \frac{\mathrm{Im}(z)^s}{|cz+c\ell+r|^{2s}} = \sideset{}{'}\sum_{\substack{c \geq 1 \\ r \pmod{c}}} \sum_{\ell \in \mathbb{Z}} \frac{\mathrm{Im}(z)^s}{|cz+c\ell+r|^{2s}}.$$

where on the right-hand side it is understood that we are summing over all triples (c,ℓ,r) with the prescribed properties. Now let

$$I_{c,r}(z,s) = \sum_{\ell \in \mathbb{Z}} \frac{\mathrm{Im}(z)^s}{|cz+c\ell+r|^{2s}}.$$

We apply the Poisson summation formula to $I_{c,r}(z,s)$. This is allowed since the summands are absolutely integrable as they exhibit polynomial decay of order $\sigma > 1$ and $I_{c,r}(z,s)$ is holomorphic (because $E(z,s)$ is). By the identity theorem it suffices to apply the Poisson summation formula for $z = iy$ with $y > 0$. So let $f(x)$ be given by

$$f(x) = \frac{y^s}{|cx+r+icy|^{2s}}.$$

Then $f(x)$ is absolutely integrable on $\mathbb{R}$ as we have just mentioned. We compute the Fourier transform:

$$
\begin{aligned}
(\mathcal{F}f)(t) = \int_{\mathbb{R}} f(x)e^{-2\pi itx}\,dx &= \int_{\mathbb{R}} \frac{y^s}{|cx+r+icy|^{2s}}e^{-2\pi itx}\,dx \\
&= \int_{\mathbb{R}} \frac{y^s}{((cx+r)^2+(cy)^2)^s}e^{-2\pi itx}\,dx \\
&= e^{2\pi it\frac{r}{c}}\int_{\mathbb{R}} \frac{y^s}{((cx)^2+(cy)^2)^s}e^{-2\pi itx}\,dx && x \to x-\frac{r}{c} \\
&= \frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\int_{\mathbb{R}} \frac{y^s}{(x^2+y^2)^s}e^{-2\pi itx}\,dx \\
&= \frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\int_{\mathbb{R}} \frac{y^{s+1}}{((xy)^2+y^2)^s}e^{-2\pi itxy}\,dx && x \to xy \\
&= \frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\int_{\mathbb{R}} \frac{y^{1-s}}{(x^2+1)^s}e^{-2\pi itxy}\,dx.
\end{aligned}
$$

Appealing to Appendix E.1 to compute this latter integral, we see that

$$
(\mathcal{F}f)(t) = \begin{cases} \frac{y^{1-s}}{c^{2s}}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} & \text{if } t=0, \\ \frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\frac{2\pi^s|t|^{s-\frac{1}{2}}}{\Gamma(s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y) & \text{if } t\neq 0. \end{cases}
$$

By the Poisson summation formula and the identity theorem, we have

$$
I_{c,r}(z,s) = \frac{y^{1-s}}{c^{2s}}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} + \sum_{t\neq 0}\left(\frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\frac{2\pi^s|t|^{s-\frac{1}{2}}}{\Gamma(s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi itx},
$$

for all $z \in \mathbb{H}$. Substituting this back into the Eisenstein series gives a form of the Fourier-Whittaker series:

$$
\begin{aligned}
E_(z,s) &= y^s + \sideset{}{'}\sum_{\substack{c\geq 1 \\ r \pmod{c}}}\left(\frac{y^{1-s}}{c^{2s}}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} + \sum_{t\geq 1}\left(\frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\frac{2\pi^s|t|^{s-\frac{1}{2}}}{\Gamma(s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi itx}\right) \\
&= y^s + y^{1-s}\sideset{}{'}\sum_{\substack{c\geq 1 \\ r \pmod{c}}}\frac{1}{c^{2s}}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} + \sum_{t\geq 1}\left(\sideset{}{'}\sum_{\substack{c\geq 1 \\ r \pmod{c}}}\frac{e^{2\pi it\frac{r}{c}}}{c^{2s}}\frac{2\pi^s|t|^{s-\frac{1}{2}}}{\Gamma(s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi itx} \\
&= y^s + y^{1-s}\sum_{c\geq 1}\frac{r(0,c)}{c^{2s}}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} + \sum_{t\geq 1}\left(\sum_{c\geq 1}\frac{r(t,c)}{c^{2s}}\frac{2\pi^s|t|^{s-\frac{1}{2}}}{\Gamma(s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi itx}.
\end{aligned}
$$

By applying Lemma 8.6.1 to compute the Dirichlet series of Ramanujan sums, we obtain the desired Fourier-Whittaker series:

$$
E(z,s) = y^s + y^{1-s}\frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)\zeta(2s-1)}{\Gamma(s)\zeta(2s)} + \sum_{t\geq 1}\left(\frac{2\pi^s|t|^{s-\frac{1}{2}\sigma_{1-2s}(|t|)}}{\Gamma(s)\zeta(2s)}\sqrt{y}K_{s-\frac{1}{2}}(2\pi|t|y)\right)e^{2\pi itx}.
$$

$\square$

Having computed the Fourier-Whittaker series, we would like to obtain a functional equation for $E(z,s)$ as $s \to 1-s$. To this end, we define $E^*(z,s)$ by

$$
E^*(z,s) = \Lambda(2s,\zeta)E(z,s) = \pi^{-s}\Gamma(s)\zeta(2s)E(z,s).
$$

From Proposition 8.6.1, the Fourier coefficients $a^*(n, y, s)$ of $E^*(z, s)$ in the Fourier series

$$E^*(z, s) = a^*(0, y, s) + \sum_{n \neq 0} a^*(n, y, s) e^{2\pi i n x},$$

are given by

$$a^*(n, y, s) = \begin{cases} y^s \pi^{-s} \Gamma(s) \zeta(2s) + y^{1-s} \pi^{-\left(s-\frac{1}{2}\right)} \Gamma(s - \frac{1}{2}) \zeta(2s - 1) & \text{if } n = 0, \\ 2|n|^{s-\frac{1}{2}} \sigma_{1-2s}(|n|) \sqrt{y} K_{s-\frac{1}{2}}(2\pi |n| y) & \text{if } n \neq 0. \end{cases}$$

We can now derive a functional equation for $E^*(z, s)$. Using the definition and functional equation for $\Lambda(2s - 1, \zeta)$, we can rewrite the second term in the constant coefficient to get

$$a^*(n, y, s) = \begin{cases} y^s \Lambda(2s, \zeta) + y^{1-s} \Lambda(2(1 - s), \zeta) & \text{if } n = 0, \\ 2|n|^{s-\frac{1}{2}} \sigma_{1-2s}(|n|) \sqrt{y} K_{s-\frac{1}{2}}(2\pi |n| y) & \text{if } n \neq 0. \end{cases} \tag{8.21}$$

Now observe that the constant coefficient is invariant under $s \to 1 - s$. Each $n \neq 0$ coefficient is also invariant under $s \to 1 - s$. To see this we will use two facts. First, from Appendix B.6, $K_s(y)$ is invariant under $s \to -s$ and so $K_{s-\frac{1}{2}}(2\pi |n| y)$ is invariant as $s \to 1 - s$. Second, for $n \geq 1$ we have

$$n^{s-\frac{1}{2}} \sigma_{1-2s}(n) = n^{\frac{1}{2}-s} n^{2s-1} \sigma_{1-2s}(n) = n^{\frac{1}{2}-s} n^{2s-1} \sum_{d \mid n} d^{1-2s} = n^{\frac{1}{2}-s} \sum_{d \mid n} \left(\frac{n}{d}\right)^{2s-1} = n^{\frac{1}{2}-s} \sigma_{2s-1}(n),$$

where the second to last equality follows by writing $n^{2s-1} = \left(\frac{n}{d}\right)^{2s-1} d^{2s-1}$ for each $d \mid n$. These two facts together give the invariance of the $n \neq 0$ coefficients under $s \to 1 - s$. Altogether, we have shown the following functional equation for $E^*(z, s)$:

$$E^*(z, s) = E^*(z, 1 - s).$$

We can now obtain meromorphic continuation of $E^*(z, s)$ in s to all of $\mathbb{C}$ for any $z \in \mathbb{H}$. We first write $E^*(z, s)$ as a Fourier-Whittaker series using Equation (8.21):

$$E^*(z, s) = y^s \Lambda(2s, \zeta) + y^{1-s} \Lambda(2(1 - s), \zeta) + \sum_{n \neq 0} 2|n|^{s-\frac{1}{2}} \sigma_{1-2s}(|n|) \sqrt{y} K_{s-\frac{1}{2}}(2\pi |n| y) e^{2\pi i n x}.$$

Since $\Lambda(2s, \zeta)$ is meromorphic on $\mathbb{C}$, the constant term of $E^*(z, s)$ is as well. To finish the meromorphic continuation of $E^*(z, s)$ it now suffices to show

$$\sum_{n \neq 0} 2|n|^{s-\frac{1}{2}} \sigma_{1-2s}(|n|) \sqrt{y} K_{s-\frac{1}{2}}(2\pi |n| y) e^{2\pi i n x},$$

is meromorphic on $\mathbb{C}$. We will actually prove it is locally absolutely uniformly convergent. So let K be a compact subset of $\mathbb{C}$. Then we have to show $E^*(z, s)$ is absolutely convergent on K for any $z \in \mathbb{H}$. To achieve this we need two bounds, one for $\sigma_{1-2s}(|n|)$ and one for $K_{s-\frac{1}{2}}(2\pi |n| y)$. For the first bound, we use the estimate $\sigma_0(|n|) \ll_\varepsilon |n|^\varepsilon$ (recall Proposition A.3.1). Therefore we have the crude bound

$$\sigma_{1-2s}(|n|) = \sum_{d \mid n} d^{1-2s} < \sigma_0(|n|) |n|^{1-2s} \ll_\varepsilon |n|^{1-2s+\varepsilon}.$$

For the second bound, Lemma B.6.2 implies

$$K_{s-\frac{1}{2}}(2\pi|n|y) \ll e^{-2\pi|n|y}.$$

Using these two bounds, we have

$$\sum_{n\neq 0} 2|n|^{s-\frac{1}{2}}\sigma_{1-2s}(|n|)\sqrt{y}K_{s-\frac{1}{2}}(2\pi|n|y)e^{2\pi inx} \ll_\varepsilon \sum_{n\geq 1} n^{\frac{1}{2}-s+\varepsilon}\sqrt{y}e^{-2\pi ny}. \tag{8.22}$$

This latter series is absolutely uniformly convergent on K by the ratio and Weierstrass M-tests. Therefore $E^*(z,s)$ is absolutely convergent on K for any $z \in \mathbb{H}$ and the meromorphic continuation to $\mathbb{C}$ follows. It remains to investigate the poles and residues. We will accomplish this from direct inspection of the Fourier-Whittaker coefficients:

Proposition 8.6.2. *$E^*(z,s)$ has simple poles at $s=0$ and $s=1$, and*

$$\operatorname*{Res}_{s=0} E^*(z,s) = -\frac{1}{2} \quad \textit{and} \quad \operatorname*{Res}_{s=1} E^*(z,s) = \frac{1}{2}.$$

Proof. Since the constant term in the Fourier-Whittaker series of $E^*(z,s)$ is the only non-holomorphic term, poles of $E^*(z,s)$ can only come from that term. So we are reduced to understanding the poles of

$$y^s\Lambda(2s,\zeta) + y^{1-s}\Lambda(2(1-s),\zeta). \tag{8.23}$$

Notice $\Lambda(2s,\zeta)$ has simple poles at $s=0$, $s=\frac{1}{2}$ (one from the Riemann zeta function and one from the gamma factor) and no others. It follows that $E^*(z,s)$ has a simple pole at $s=0$ coming from the y^s term in Equation (8.23), and by the functional equation there is also a pole at $s=1$ coming from the y^{1-s} term. At $s=\frac{1}{2}$, both terms in Equation (8.23) have simple poles and we will show that the singularity there is removable. Recall $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$. Also, by Proposition 8.1.1, $\operatorname{Res}_{s=\frac{1}{2}}\zeta(2s)=\frac{1}{2}$ and $\operatorname{Res}_{s=\frac{1}{2}}\zeta(2(1-s))=-\frac{1}{2}$. So altogether

$$\operatorname*{Res}_{s=\frac{1}{2}} E^*(z,s) = \operatorname*{Res}_{s=\frac{1}{2}}[y^s\Lambda(2s,\zeta)+y^{1-s}\Lambda(2(1-s),\zeta)] = \frac{1}{2}y^{\frac{1}{2}} - \frac{1}{2}y^{\frac{1}{2}} = 0.$$

Hence the singularity at $s=\frac{1}{2}$ is removable. As for the residues at $s=0$ and $s=1$, the functional equation implies that they are negatives of each other. So it suffices to compute the residue at $s=0$. Recall $\zeta(0)=-\frac{1}{2}$ and $\operatorname{Res}_{s=0}\Gamma(s)=1$. Then together we find

$$\operatorname*{Res}_{s=0} E^*(z,s) = \operatorname*{Res}_{s=0} y^s\Lambda(2s,\zeta) = -\frac{1}{2}.$$

□

This completes our study of $E(z,s)$.

The Integral Representation: Part III

We can now continue with the Rankin-Selberg convolution $L(s, f\otimes g)$. Writing Equation (8.18) in terms of $E^*(z,s)$ and $L(s,f\otimes g)$ results in the integral representation

$$L(s,f\otimes g) = \frac{(4\pi)^{s+k-1}\pi^s}{\Gamma(s+k-1)\Gamma(s)}\int_{\mathcal{F}} f(z)\overline{g(z)}\operatorname{Im}(z)^k E^*(z,s)\,d\mu.$$

This integral representation will give analytic continuation. To see this, note that the gamma factors are analytic for $\sigma<0$. By the functional equation for $E^*(z,s)$, the integral is invariant as $s\to 1-s$. These two facts together give analytic continuation to $\mathbb{C}$ outside of the critical strip. The continuation inside of the critical strip will be meromorphic because of the poles of $E^*(z,s)$. To see this, taking the integral representation and substituting the Fourier-Whittaker series for $E^*(z,s)$ gives

$$\begin{aligned} L(s, f\otimes g) = \frac{(4\pi)^{s+k-1}\pi^s}{\Gamma(s+k-1)\Gamma(s)}\bigg[\int_{\mathcal{F}} f(x+iy)\overline{g(x+iy)}y^k(y^s\Lambda(2s,\zeta)+y^{1-s}\Lambda(2(1-s),\zeta))\,\frac{dx\,dy}{y^2} \\ + \int_{\mathcal{F}} f(x+iy)\overline{g(x+iy)}y^k\sum_{n\neq 0}2|n|^{s-\frac{1}{2}}\sigma_{1-2s}(|n|)\sqrt{y}K_{s-\frac{1}{2}}(2\pi|n|y)e^{2\pi inx}\,\frac{dx\,dy}{y^2}\bigg], \end{aligned} \tag{8.24}$$

and we are reduced to showing that both integrals are locally absolutely uniformly convergent in the critical strip and distance ε away from the poles of $E^*(z,s)$. Indeed, the first integral is locally absolutely uniformly convergent in this region since the rapid decay of f and g imply that the integrand is bounded and we are integrating over a region of finite volume. As for the second integral, since s is in the critical strip $0\le\sigma\le 1$ and so Equation (8.22) implies that it is

$$O_\varepsilon\left(\int_{\mathcal{F}} f(x+iy)\overline{g(x+iy)}y^k\sum_{n\ge 1}n^{\frac{1}{2}+\varepsilon}\sqrt{y}e^{-2\pi ny}\,\frac{dx\,dy}{y^2}\right).$$

The Taylor series of the derivative of $\frac{e^y}{1-e^y}$ (upon taking $\varepsilon\le\frac{1}{2}$) gives

$$\sum_{n\ge 1}n^{\frac{1}{2}+\varepsilon}e^{-2\pi ny} = O\left(\sum_{n\ge 1}ne^{-2\pi ny}\right) = O(e^{-2\pi y}),$$

and thus has rapid decay. Together with the rapid decay of f and g, the integrand is bounded and thus is locally absolutely uniformly convergent because we are integrating over a region of finite volume. The meromorphic continuation to the critical strip and hence to all of $\mathbb{C}$ follows. In particular, $L(s,f\otimes g)$ has at most simple poles at $s=0$ and $s=1$. Actually, there is no pole at $s=0$. Indeed, $\gamma(s,f\otimes g)$ has a simple pole at $s=0$ coming from the gamma factors and therefore its reciprocal has a simple zero. This cancels the simple pole at $s=0$ coming from $E^*(z,s)$ and therefore $L(s,f\otimes g)$ is has a removable singularity at $s=0$. So there is at worst a simple pole at $s=1$.

The Functional Equation

An immediate consequence of the symmetry of integral representation is the functional equation:

$$\frac{\Gamma(s+k-1)\Gamma(s)}{(4\pi)^{s+k-1}\pi^s}L(s,f\otimes g) = \frac{\Gamma((1-s)+k-1)\Gamma(1-s)}{(4\pi)^{(1-s)+k-1}\pi^{1-s}}L(1-s,f\otimes g).$$

Applying the Legendre duplication formula for the gamma function twice we see that

$$\begin{aligned} \frac{\Gamma(s+k-1)\Gamma(s)}{(4\pi)^{s+k-1}\pi^s} &= \frac{2^{2s+k-3}}{(4\pi)^{s+k-1}\pi^{s+1}}\Gamma\left(\frac{s+k-1}{2}\right)\Gamma\left(\frac{s+k}{2}\right)\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right) \\ &= \frac{1}{2^{k+1}\pi^k}\pi^{-2s}\Gamma\left(\frac{s+k-1}{2}\right)\Gamma\left(\frac{s+k}{2}\right)\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right). \end{aligned} \tag{8.25}$$

The factor in front is independent of s and can therefore be canceled in the functional equation. We identify the gamma factor as:

$$\gamma(s,f\otimes g) = \pi^{-2s}\Gamma\left(\frac{s+k-1}{2}\right)\Gamma\left(\frac{s+k}{2}\right)\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right),$$

with $\mu_{1,1} = k-1$, $\mu_{2,2} = k$, $\mu_{1,2} = 0$, and $\mu_{2,1} = 1$ the local roots at infinity. The completed L-function is

$$\Lambda(s, f \otimes g) = \pi^{-2s}\Gamma\left(\frac{s+k-1}{2}\right)\Gamma\left(\frac{s+k}{2}\right)\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right)L(s, f \otimes g),$$

so the conductor is $q(f \otimes g) = 1$ and no primes ramify. Clearly $q(f \otimes g) \mid q(f)^2 q(g)^2$ and conductor dropping does not occur. Then

$$\Lambda(s, f \otimes g) = \Lambda(1-s, f \otimes g),$$

is the functional equation of $L(s, f \otimes g)$. In particular, the root number $\varepsilon(f \otimes g) = 1$, and $L(s, f \otimes g)$ is self-dual. We can now show that $L(s, f \otimes g)$ is of order 1. Since the possible pole at $s = 1$ is simple, multiplying by $(s-1)$ clears the possible polar divisor. As the integrals in the integral representation are locally absolutely uniformly convergent, computing the order amounts to estimating the gamma factor. Since the reciprocal of the gamma function is of order 1, we have

$$\frac{1}{\gamma(s, f \otimes g)} \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

So the reciprocal of the gamma factor is also of order 1. Then we find that

$$(s-1)L(s, f \otimes g) \ll_\varepsilon e^{|s|^{1+\varepsilon}}.$$

Thus $(s-1)L(s, f \otimes g)$ is of order 1, and so $L(s, f \otimes g)$ is as well after removing the polar divisor. At last, we compute the residue of $L(s, f \otimes g)$ at $s = 1$:

Proposition 8.6.3. *Let $f, g \in \mathcal{S}_k(1)$ be primitive Hecke eigenforms. Then*

$$\operatorname*{Res}_{s=1} L(s, f \otimes g) = \frac{4^k \pi^{k+1} V}{2\Gamma(k)} \langle f, g \rangle,$$

where $\langle f, g \rangle$ is the Petersson inner product.

Proof. As $V = \frac{\pi}{3}$, Proposition 8.6.2 implies

$$\operatorname*{Res}_{s=1} L(s, f \otimes g) = \frac{4^k \pi^{k+1}}{\Gamma(k)} \operatorname*{Res}_{s=1} \int_{\mathcal{F}} f(z)\overline{g(z)}\operatorname{Im}(z)^k E^*(z, s)\, d\mu = \frac{4^k \pi^{k+1} V}{2\Gamma(k)} \langle f, g \rangle. \qquad \square$$

Notice that if $f = g$, then $\langle f, f \rangle \neq 0$ and therefore the residue at $s = 1$ is not zero and hence not a removable singularity. Actually, this is the only instance in which there is a pole since Theorem 4.6.1 implies that the primitive Hecke eigenforms are orthogonal so that $\langle f, g \rangle = 0$ unless $f = g$. We summarize all of our work into the following theorem:

Theorem 8.6.1. *For any two primitive Hecke eigenforms $f, g \in \mathcal{S}_k(1)$, $L(s, f \otimes g)$ is a Selberg class L-function. For $\sigma > 1$, it has a degree 4 Euler product given by*

$$L(s, f \otimes g) = \prod_p \left(\sum_{n \geq 0} \frac{a_f(p^n)\overline{a_g(p^n)}}{p^{ns}}\right).$$

Moreover, it admits meromorphic continuation to $\mathbb{C}$, possesses the functional equation

$$\pi^{-2s}\Gamma\left(\frac{s+k-1}{2}\right)\Gamma\left(\frac{s+k}{2}\right)\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right)L(s, f \otimes g) = \Lambda(s, f \otimes g) = \Lambda(1-s, f \otimes g),$$

and there is simple pole at $s = 1$ of residue $\frac{4^k \pi^{k+1} V}{2\Gamma(k)} \langle f, g \rangle$ provided $f = g$.

The Rankin-Selberg Method

The Rankin-Selberg method is much more complicated in general, but the argument is essentially the same. Let f and g both be primitive Hecke or Heck-Maass eigenforms with Fourier or Fourier-Whittaker coefficients $a_f(n)$ and $a_g(n)$ respectively. We suppose f has weight k/type ν, level N, and character χ, and g has weight ℓ/type η, level M, and character ψ. The L-function $L(s, f \times g)$ of f and g is given by the L-series

$$L(s, f \times g) = \sum_{n \geq 1} \frac{a_{f \times g}(n)}{n^s} = \sum_{n \geq 1} \frac{a_f(n)\overline{a_g(n)}}{n^s}.$$

The **Rankin-Selberg convolution** $L(s, f \otimes g)$ of f and g is defined by

$$L(s, f \otimes g) = \sum_{n \geq 1} \frac{a_{f \otimes g}(n)}{n^s} = L(2s, \chi\overline{\psi})L(s, f \times g),$$

where $a_{f \otimes g}(n) = \sum_{n=m\ell^2} \chi\overline{\psi}(\ell^2)a_f(m)\overline{a_g(m)}$. The following argument is the **Rankin-Selberg method**:

Method 8.6.1 (Rankin-Selberg method). Let f and b both be primitive Hecke or Heck-Maass eigenforms. Also suppose the following:

(i) f has weight k/type ν, level N, and character χ.

(ii) g has weight ℓ/type η, level M, and character ψ.

(iii) The Ramanujan-Petersson conjecture for Maass forms holds if f or g are Heck-Maass eigenforms.

Then the Rankin-Selberg convolution $L(s, f \otimes g)$ is a Selberg class L-function.

We make a few remarks about the Rankin-Selberg method. Local absolute uniform convergence for $\sigma > 1$ are proved in the exactly the same way as we have described. The argument for the Euler product is also similar. However, if either $N > 1$ or $M > 1$, the computation becomes more difficult to compute since the local p factors for $p \mid NM$ change. Moreover, the situation is increasingly complicated if $(N, M) > 1$ since conductor dropping can occur. The integral representation has a similar argument, but if the weights/types are distinct the resulting Eisenstein series becomes more complicated. In particular, it is on $\Gamma_0(NM)\backslash\mathbb{H}$ and if $NM > 1$, then there is more than just the cusp at ∞. Therefore the functional equation of the Eisenstein series at the ∞ cusp then reflects into a linear combination of Eisenstein series at the other cusps. This requires the Fourier-Whittaker series of all of these Eisenstein series. Moreover, this procedure can be generalized to remove the primitive Hecke and/or primitive Hecke-Maass eigenform conditions by taking linear combinations, but we won't attempt discussing this further.

8.7 Applications of the Rankin-Selberg Method

The Ramanujan-Petersson Conjecture on Average

Let f be a Hecke or Hecke-Maass eigenform. Using Rankin-Selberg convolutions, it is possible to show the weaker result that $a_f(n) \ll_\varepsilon n^\varepsilon$ holds on average without assuming the corresponding Ramanujan-Petersson conjecture:

Proposition 8.7.1. *Let f be a primitive Hecke or Hecke-Maass eigenform. Then for any $X > 0$, we have*

$$\sum_{n \leq X} |a_f(n)| \ll_\varepsilon X^{1+\varepsilon},$$

Proof. By the Cauchy-Schwarz inequality,

$$\left(\sum_{n \leq X} |a_f(n)| \right)^2 \leq X \sum_{n \leq X} |a_f(n)|^2, \tag{8.26}$$

The Rankin-Selberg square $L(s, f \otimes f)$ is locally absolutely uniformly convergent for $\sigma > \frac{3}{2}$. Therefore it still admits meromorphic continuation to $\mathbb{C}$ with a simple pole at $s = 1$. By Landau's theorem, the abscissa of absolute convergence of $L(s, f \otimes f)$, and hence $L(s, f \times f)$ too, is 1 so by Proposition 7.1.3 we have

$$\sum_{n \leq X} |a_f(n)|^2 \ll_\varepsilon X^{1+\varepsilon},$$

for any $\varepsilon > 0$. Substituting this bound into Equation (8.26), we obtain

$$\left(\sum_{n \leq X} |a_f(n)| \right)^2 \ll_\varepsilon X^{2+\varepsilon},$$

and taking the square root yields

$$\sum_{n \leq X} |a_f(n)| \ll_\varepsilon X^{1+\varepsilon}.$$

□

The bound in Proposition 8.7.1 should be compared with the implication $a_f(n) \ll_\varepsilon n^\varepsilon$ that follows from the corresponding Ramanujan-Petersson conjecture. While Proposition 8.7.1 is not useful in the holomorphic form case, it is in the Maass form case. Indeed, recall that if f is a primitive Hecke-Maass eigenform we needed to assume the Ramanujan-Petersson conjecture for Maass forms to ensure $a_f(n) \ll_\varepsilon n^\varepsilon$ so that $L(s, f)$ was locally absolutely uniformly convergent for $\sigma > 1$. However, Propositions 7.1.4 and 8.7.1 now together imply $L(s, f)$ is locally absolutely uniformly convergent for $\sigma > 1$.

Strong Multiplicity One

Let f be a primitive Hecke or Hecke-Maass eigenform. Then f is determined by Hecke eigenvalues at primes for fixed weight/type, and level. Using Rankin-Selberg convolutions, we can prove **strong multiplicity one** for holomorphic or Maass forms which says that f is determined by Hecke eigenvalues at all but finitely many primes:

Theorem 8.7.1 (Strong multiplicity one, holomorphic and Maass versions). *Let f and g both be primitive Hecke or Hecke-Maass eigenforms. Denote the Hecke eigenvalues by $\lambda_f(n)$ and $\lambda_g(n)$ respectively. If $\lambda_f(p) = \lambda_g(p)$ for all but finitely many primes p, then $f = g$.*

Proof. Let S be the set the primes for which $\lambda_f(p) \neq \lambda_g(p)$ including the primes that ramify for $L(s, f)$ and $L(s, g)$. By assumption, S is finite. As the local factors of $L(s, f \otimes g)$ are holomorphic and nonzero at $s = 1$, the order of the pole of $L(s, f \otimes g)$ is the same as the order of the pole of

$$L(s, f \otimes g) \prod_{p \in S} L_p(s, f \otimes g)^{-1} = \prod_{p \notin S} L_p(s, f \otimes g).$$

But as $\lambda_f(p) = \lambda_g(p)$ for all $p \notin S$, we have

$$\prod_{p \notin S} L_p(s, f \otimes g) = \prod_{p \notin S} L_p(s, f \otimes f),$$

and so

$$L(s, f \otimes g) \prod_{p \in S} L_p(s, f \otimes g)^{-1} = L(s, f \otimes f) \prod_{p \in S} L_p(s, f \otimes f)^{-1}.$$

Since $L(s, f \otimes f)$ has a simple pole at $s = 1$, it follows that $L(s, f \otimes g)$ does too. But then $f = g$. $\square$

Chapter 9

Classical Applications

We will discuss some classical applications of L-functions. Our first result is a crowning gem of analytic number theory: Dirichlet's theorem. This result is a consequence of a non-vanishing result for Dirichlet L-series at $s = 1$. Then we discuss Siegel zeros in the case of Dirichlet L-functions and as a result obtain a lower bound for Dirichlet L-functions at $s = 1$. Lastly, we prove the prime number theorem and its variant for primes restricted to a certain residue class, the Siegel–Walfisz theorem, in the classical manner.

9.1 Dirichlet's Theorem

One of the more well-known arithmetic results proved using L-series is **Dirichlet's theorem**:

Theorem 9.1.1 (Dirichlet's theorem). *Let a and m be positive integers such that $(a, m) = 1$. Then the arithmetic progression $\{a + km \mid k \in \mathbb{N}\}$ contain infinitely many primes.*

We will delay the proof for the moment, for it is well-worth understanding the some of the motivation behind why this theorem is interesting and how exactly Dirichlet used the analytic techniques of L-series to attack this purely arithmetic statement. We being by recalling Euclid's famous theorem on the infinitude of the primes. Euclid's proof is completely elementary and arithmetic in nature. He argues that if there were finitely many primes $p_1, p_2, \ldots, p_k$ then a short consideration of $(p_1 p_2 \cdots p_k) + 1$ shows that this number must either be divisible by a prime not in our list or must be prime itself. As primes are the multiplicative building blocks of arithmetic, Euclid assures us that we have an ample amount of primes to work with. Now there is a slightly stronger result due to Euler (see [Eul44]) requiring analytic techniques:

Theorem 9.1.2. *The series*

$$\sum_p \frac{1}{p},$$

diverges.

Proof. For $\sigma > 1$, taking the logarithm of the Euler product of $\zeta(s)$, we get

$$\log \zeta(s) = -\sum_p \log(1 - p^{-s}).$$

The Taylor series of the logarithm gives

$$\log(1 - p^{-s}) = \sum_{k \geq 1} (-1)^{k-1} \frac{(-p^{-s})^k}{k} = \sum_{k \geq 1} (-1)^{2k-1} \frac{1}{kp^{ks}},$$

so that

$$\log \zeta(s) = \sum_p \sum_{k\geq 1} \frac{1}{kp^{ks}}.$$

The double sum restricted to $k \geq 2$ is uniformly bounded for $\sigma > 1$. To see this, first observe

$$\sum_{k\geq 2} \frac{1}{kp^{ks}} \ll \sum_{k\geq 2} \frac{1}{p^k} = \frac{1}{p^2}\sum_{k\geq 0}\frac{1}{p^k} = \frac{1}{p^2}(1-p^{-1})^{-1} \leq \frac{2}{p^2},$$

where the last inequality follows because $p \geq 2$. Then

$$\sum_p \sum_{k\geq 2} \frac{1}{kp^{ks}} \ll 2\sum_p \frac{1}{p^2} < 2\sum_{n\geq 1}\frac{1}{n^2} = 2\zeta(2).$$

Therefore

$$\log \zeta(s) - \sum_p \frac{1}{p^s} = \sum_p \sum_{k\geq 2} \frac{1}{kp^{ks}},$$

and remains bounded as $s \to 1$. The claim now follows since $\zeta(s)$ has a simple pole at $s = 1$. □

Theorem 9.1.2 tells us that there are infinitely many primes but also that the primes are not too sparse in the integers for otherwise the series would converge. The idea Dirichlet used to prove his result on primes in arithmetic progressions was in a very similar spirit. He sought out to prove the divergence of the series

$$\sum_{p\equiv a \pmod m} \frac{1}{p},$$

for positive integers a and m with $(a, m) = 1$ as the divergence immediately implies there are infinitely many primes p of the form $p \equiv a \pmod m$. In the case $a = 1$ and $m = 2$ we recover Theorem 9.1.2 exactly since every prime is odd. Dirichlet's proof proceeds in a similar way to that of Theorem 9.1.2 and this is where Dirichlet used what are now known as Dirichlet characters and Dirichlet L-series. The proof can be broken into three steps. The first is to proceed as Euler did, but with the Dirichlet L-series $L(s, \chi)$ where χ has modulus m. That is, write $L(s, \chi)$ as a sum over primes and a bounded term as $s \to 1$. The next step is to use the orthogonality relations of the characters to sieve out the correct sum. The last step is to show the non-vanishing result $L(1, \chi) \neq 0$ for all non-principal characters χ. This is the essential part of the proof as it is what assures us that the sum diverges. Luckily, we have done most of the hard work to prove this already:

Theorem 9.1.3. *Let χ be a non-principal Dirichlet character. Then $L(1, \chi)$ is finite and nonzero.*

Proof. This follows immediately by applying Lemma 7.10.1 to $\zeta(s)L(s, \chi)$ and noting that $L(s, \chi)$ is holomorphic. □

We now prove Dirichlet's theorem:

Proof of Dirichlet's theorem. Let χ be a Dirichlet character modulo m. Then for $\sigma > 1$, taking the logarithm of the Euler product of $L(s, \chi)$ gives

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

The Taylor series of the logarithm implies

$$\log(1-\chi(p)p^{-s}) = \sum_{k\ge1}(-1)^{k-1}\frac{(-\chi(p)p^{-s})^k}{k} = \sum_{k\ge1}(-1)^{2k-1}\frac{\chi(p^k)}{kp^{ks}},$$

so that

$$\log L(s,\chi) = \sum_p\sum_{k\ge1}\frac{\chi(p^k)}{kp^{ks}}.$$

The double sum restricted to $k \ge 2$ is uniformly bounded for $\sigma > 1$. Indeed, first observe

$$\left|\sum_{k\ge2}\frac{\chi(p^k)}{kp^{ks}}\right| \ll \sum_{k\ge2}\frac{1}{p^k} = \frac{1}{p^2}\sum_{k\ge0}\frac{1}{p^k} = \frac{1}{p^2}(1-p^{-1})^{-1} \le \frac{2}{p^2},$$

where the last inequality follows because $p > 2$. Then

$$\left|\sum_p\sum_{k\ge2}\frac{\chi(p^k)}{kp^{ks}}\right| \le 2\sum_p\frac{1}{p^2} < 2\sum_{n\ge1}\frac{1}{n^2} = 2\zeta(2),$$

as desired. Now write

$$\sum_{\chi \pmod m}\overline{\chi(a)}\log L(s,\chi) = \sum_{\chi \pmod m}\sum_p\frac{\overline{\chi(a)}\chi(p)}{p^s} + \sum_{\chi \pmod m}\overline{\chi(a)}\sum_p\sum_{k\ge2}\frac{\chi(p^k)}{kp^{ks}}.$$

By the orthogonality relations (Proposition 1.3.1 (ii)), we find that

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

and so

$$\sum_{\chi \pmod m}\overline{\chi(a)}\log L(s,\chi) - \sum_{\chi \pmod m}\overline{\chi(a)}\sum_p\sum_{k\ge2}\frac{\chi(p^k)}{kp^{ks}} = \varphi(m)\sum_{p\equiv a \pmod m}\frac{1}{p^s}.$$

The triple sum is uniformly bounded for $\sigma > 1$ because the inner double sum is and there are finitely many Dirichlet characters modulo m. Therefore it suffices to show that the first sum on the left-hand side diverges as $s \to 1$. For $\chi = \chi_{m,0}$,

$$L(s,\chi_{m,0}) = \zeta(s)\prod_{p\mid m}(1-p^{-s}),$$

so the corresponding term in the sum is

$$\overline{\chi_{m,0}}(a)\log L(s,\chi_{m,0}) = \log\zeta(s) + \sum_{p\mid m}\log(1-p^{-s}),$$

which diverges as $s \to 1$ because $\zeta(s)$ has a simple pole at $s = 1$. We will be done if $\log L(s,\chi)$ remains bounded as $s \to 1$ for all $\chi \ne \chi_{m,0}$. So assume χ is non-principal. Then we must show $L(1,\chi)$ is finite and nonzero. This follows from Theorem 9.1.3. $\square$

For primitive χ of conductor $q > 1$, we know from Theorem 9.1.3 that $L(1,\chi)$ is finite and nonzero. It is interesting to know whether or not this value is computable in general. Indeed it is. The computation is fairly straightforward and only requires some basic properties of Ramanujan and Gauss sums that we have already developed. The idea is to rewrite the character values $\chi(n)$ so that we can collapse the infinite series into a Taylor series. Our result is the following:

Theorem 9.1.4. *Let χ be a primitive Dirichlet character with conductor $q > 1$. Then*

$$L(1,\chi) = -\frac{\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)\log\left(\sin\left(\frac{\pi a}{q}\right)\right) \quad \text{or} \quad L(1,\chi) = \frac{\tau(\chi)\pi i}{q^2}\sum_{a \pmod q}\overline{\chi}(a)a,$$

according to whether χ is even or odd.

Proof. First compute

$$\begin{aligned}
\chi(n) &= \frac{1}{\tau(\overline{\chi})}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}} && \text{Corollary 1.4.1}\\
&= \frac{\chi(-1)}{\overline{\tau(\chi)}}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}} && \text{Proposition 1.4.2 (i) and } \chi(-1)^2 = 1\\
&= \frac{\chi(-1)\tau(\chi)}{\tau(\chi)\overline{\tau(\chi)}}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}}\\
&= \frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}} && \text{Theorem 1.4.1}\\
&= \frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}}.
\end{aligned}$$

Substituting this result into the definition of $L(1,\chi)$, we find that

$$\begin{aligned}
L(1,\chi) &= \sum_{n\geq 1}\frac{1}{n}\left(\frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)e^{\frac{2\pi ian}{q}}\right)\\
&= \frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)\sum_{n\geq 1}\frac{e^{\frac{2\pi ian}{q}}}{n} \qquad (9.1)\\
&= \frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)\log\left(\left(1-e^{\frac{2\pi ia}{q}}\right)^{-1}\right),
\end{aligned}$$

where in the last line we have used the Taylor series of the logarithm. We will now simplify the last expression in Equation (9.1). Since $\sin(\theta) = \frac{e^{i\theta}-e^{-i\theta}}{2i}$, we have

$$1 - e^{\frac{2\pi ia}{q}} = -2ie^{\frac{\pi ia}{q}}\left(\frac{e^{\frac{\pi ia}{q}} - e^{-\frac{\pi ia}{q}}}{2i}\right) = -2ie^{\frac{\pi ia}{q}}\sin\left(\frac{\pi a}{q}\right).$$

Therefore the last expression in Equation (9.1) becomes

$$\frac{\chi(-1)\tau(\chi)}{q}\sum_{a \pmod q}\overline{\chi}(a)\log\left(\left(-2ie^{\frac{\pi ia}{q}}\sin\left(\frac{\pi a}{q}\right)\right)^{-1}\right).$$

As a is taken modulo q, we have $0 < \frac{\pi a}{q} < \pi$ so that $\sin\left(\frac{\pi a}{q}\right)$ is never negative. Therefore we can split up the logarithm term and obtain

$$-\frac{\chi(-1)\tau(\chi)}{q}\left(\log(-2i)\sum_{a \pmod q}\overline{\chi}(a)+\frac{\pi i}{q}\sum_{a \pmod q}\overline{\chi}(a)a+\sum_{a \pmod q}\overline{\chi}(a)\log\left(\sin\left(\frac{\pi a}{q}\right)\right)\right).$$

By the orthogonality relations (Corollary 1.3.1 (i)), the first sum above vanishes. Therefore

$$L(1,\chi)=-\frac{\chi(-1)\tau(\chi)}{q}\left(\frac{\pi i}{q}\sum_{a \pmod q}\chi(a)a+\sum_{a \pmod q}\chi(a)\log\left(\sin\left(\frac{\pi a}{q}\right)\right)\right). \tag{9.2}$$

Equation (9.2) simplifies in that one of the two sums vanish depending on if χ is even or odd. For the first sum in Equation (9.2), observe that

$$\frac{\pi i}{q}\sum_{a \pmod q}\chi(a)a=-\frac{\chi(-1)\pi i}{q}\sum_{a \pmod q}\chi(-a)(-a).$$

As $a \to -a$ is a bijection on $\mathbb{Z}/q\mathbb{Z}$, this sum vanishes if χ is even. For the second sum in Equation (9.2), we have an analogous relation of the form

$$\sum_{a \pmod q}\chi(a)\log\left(\sin\left(\frac{\pi a}{q}\right)\right)=\chi(-1)\sum_{a \pmod q}\chi(-a)\log\left(\sin\left(\frac{\pi a}{q}\right)\right).$$

As $a \to -a$ is a bijection on $\mathbb{Z}/q\mathbb{Z}$, this sum vanishes if χ is odd. This finishes the proof. □

Theorem 9.1.4 encodes some interesting identities. For example, if χ is the non-principal Dirichlet character modulo 4, then χ is uniquely defined by $\chi(1) = 1$ and $\chi(3) = \chi(-1) = -1$. In particular, χ is odd and its conductor is 4. Now

$$\tau(\chi)=\sum_{a \pmod 4}\chi(a)e^{\frac{2\pi i a}{4}}=e^{\frac{2\pi i}{4}}-e^{\frac{6\pi i}{4}}=i-(-i)=2i,$$

so by Theorem 9.1.4 we get

$$L(1,\chi)=-\frac{\chi(-1)\tau(\chi)\pi i}{16}(1-3)=\frac{\pi}{4}.$$

Expanding out $L(1,\chi)$ gives

$$1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\cdots=\frac{\pi}{4},$$

which is the famous **Madhava–Leibniz formula** for π.

9.2 Siegel's Theorem

The discussion of Siegel zeros first arose during the study of zero-free regions for Dirichlet L-functions. Refining the argument used in Theorem 7.10.1, we can show that Siegel zeros only exist when the character χ is quadratic. But first we improve the zero-free region for the Riemann zeta function:

Theorem 9.2.1. *There exists a constant $c > 0$ such that $\zeta(s)$ has no zeros in the region*

$$\sigma \geq 1-\frac{c}{\log(|t|+3)}.$$

Proof. By Theorem 7.10.1 applied to $\zeta(s)$, it suffices to show that $\zeta(s)$ has no real nontrivial zeros. To see this, let $\eta(s)$ be defined by

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

Note that $\eta(s)$ converges for $\sigma > 0$ by Proposition 7.1.1. Now for $0 < s < 1$ and even n, $\frac{1}{n^s} - \frac{1}{(n+1)^s} > 0$ so that $\eta(s) > 0$. But for $\sigma > 0$, we have

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

Therefore $\zeta(s)$ cannot admit a zero for $0 < s < 1$ because then $\eta(s)$ would be zero too. This completes the proof. $\square$

Theorem 9.2.1 shows that $\zeta(s)$ has no Siegel zeros. Moreover, since $1 - 2^{1-s} < 0$ for $0 < s < 1$, the proof shows that $\zeta(s) < 0$ in this interval as well. As for the height of the first zero, it occurs on the critical line (as predicted by the Riemann hypothesis) at height $t \approx 14.134$ (see [Dav80] for a further discussion). The first 15 zeros were computed by Gram in 1903 (see [Gra03]). Since then, billions of zeros have been computed and have all been verified to lie on the critical line. The analogous situation for Dirichlet L-functions is only slightly different but causes increasing complexity in further study. We first show that if a Siegel zero exists for the Dirichlet L-function of a primitive character, then the character is necessarily quadratic:

Theorem 9.2.2. *Let χ be a primitive Dirichlet character of conductor $q > 1$. Then there exists a constant $c > 0$ such that $L(s, \chi)$ has no zeros in the region*

$$\sigma \geq 1 - \frac{c}{\log(q(|t| + 3))},$$

except for possibly one simple real zero β_χ with $\beta_{chi} < 1$ in the case χ is quadratic.

Proof. By Theorem 7.10.1 applied to $\zeta(s)L(s, \chi)$, and shrinking c if necessary, it remains to show that there not a simple real zero β_χ if χ is not quadratic. For this, let $L(s, g)$ be the L-series defined by

$$L(s, g) = L^3(s, \chi_{q,0})L^4(s, \chi)L(s, \chi^2).$$

We have $d_g = 8$ and $\mathfrak{q}(g)$ satisfies

$$\mathfrak{q}(g) \leq \mathfrak{q}(\chi_{q,0})^3 \mathfrak{q}(\chi)^4 \mathfrak{q}(\chi^2) \leq q^8 3^5 < (3q)^8.$$

Moreover, $\mathrm{Re}(\Lambda_g(n)) \geq 0$ for $(n, q) = 1$. To see this, suppose p is an unramified prime. The local roots of $L(s, g)$ at p are 1 with multiplicity three, $\chi(p)$ with multiplicity four, and $\chi^2(p)$ with multiplicity one. So for any $k \geq 1$, the sum of k-th powers of these local roots is

$$3 + 4\chi^k(p) + \chi^{2k}(p).$$

Writing $\chi(p) = e^{i\theta}$, the real part of this expression is

$$3 + 4\cos(\theta) + \cos(2\theta) = 2(1 + \cos(\theta))^2 \geq 0,$$

where we have also made use of the identity $3 + 4\cos(\theta) + \cos(2\theta) = 2(1 + \cos(\theta))^2$. Thus $\mathrm{Re}(\Lambda_g(n)) \geq 0$ for $(n, q) = 1$, and the conditions of Lemma 7.10.1 are satisfied for $L(s, g)$ (recall Equation (8.9) for the L-series $L(s, \chi_{q,0})$ and $L(s, \chi^2)$). On the one hand, if β be a real nontrivial zero of $L(s, \chi)$ then $L(s, g)$ has a real nontrivial zero at $s = \beta$ of order at least 4. On the other hand, using Equation (8.9) and that $\chi^2 \neq \chi_{q,0}$, $L(s, g)$ has a pole at $s = 1$ of order 3. Then, upon shrinking c if necessary, Lemma 7.10.1 gives a contradiction since $r_g = 3$. This completes the proof. $\square$

Siegel zeros present an unfortunate obstruction to zero-free region results for Dirichlet L-functions when the primitive character χ is quadratic. However, if we no longer require the constant c in the zero-free region to be effective, we can obtain a much better result for how close the Siegel zero can be to 1. Ultimately, this improved bound results from a lower bound for $L(1,\chi)$ (recall that this is nonzero from our discussion about Dirichlet's theorem). **Siegel's theorem** refers to either this lower bound or to the improved zero-free region. In the lower bound version, Siegel's theorem is the following:

Theorem 9.2.3 (Siegel's theorem, lower bound version). *Let χ be a primitive quadratic Dirichlet character modulo $q > 1$. Then there exists a positive constant $c_1(\varepsilon)$ such that*

$$L(1,\chi) \geq \frac{c_1(\varepsilon)}{q^{\varepsilon}}.$$

In the zero-free region version, Siegel's theorem takes the following form:

Theorem 9.2.4 (Siegel's theorem, zero-free region version). *Let χ be a primitive quadratic Dirichlet character modulo $q > 1$. Then there exists a positive constant $c_2(\varepsilon)$ such that $L(s,\chi)$ has no real zeros in the segment*

$$\sigma \geq 1 - \frac{c_2(\varepsilon)}{q^{\varepsilon}}.$$

The largest defect of Siegel's theorem, in either version, is that the implicit constants $c_1(\varepsilon)$ and $c_2(\varepsilon)$ are ineffective (and not necessarily equal). Actually, the lower bound result is slightly stronger as it implies the zero-free region result. We will first prove the zero-free region given the lower bound, and then we will prove the lower bound. Before we begin, we need two small lemmas about the size of $L'(\sigma,\chi)$ and $L(\sigma,\chi)$ for σ close to 1:

Lemma 9.2.1. *Let χ be a non-principal Dirichlet character modulo $m > 1$. Then $L'(\sigma,\chi) = O(\log^2(m))$ for any σ such that $0 \leq 1-\sigma \leq \frac{1}{\log(m)}$.*

Proof. Setting $A(X) = \sum_{n \leq X} \chi(n)$ we have $A(X) \ll 1$ by Corollary 1.3.1 (i) and that χ is periodic. Therefore $\sigma_c \leq 0$ by Proposition 7.1.1. Hence for σ in the prescribed region, $L(\sigma,\chi)$ is holomorphic and its derivative is given by

$$L'(\sigma,\chi) = \sum_{n \geq 1} \frac{\chi(n)\log(n)}{n^{\sigma}} = \sum_{n<m} \frac{\chi(n)\log(n)}{n^{\sigma}} + \sum_{n \geq m} \frac{\chi(n)\log(n)}{n^{\sigma}}.$$

We will show that the last two sums are both $O(\log^2(m))$. For the first sum, if $n < m$, we have

$$\left|\frac{\chi(n)\log(n)}{n^{\sigma}}\right| \leq \frac{1}{n^{\sigma}}\log(n) = \frac{n^{1-\sigma}}{n}\log(n) < \frac{m^{1-\sigma}}{n}\log(n) < \frac{e}{n}\log(m),$$

where the last inequality follows because $1-\sigma \leq \frac{1}{\log(m)}$. Then

$$\left|\sum_{n \leq m} \frac{\chi(n)\log(n)}{n^{s}}\right| < e\log(m)\sum_{n<m}\frac{1}{n} < e\log(m)\int_1^m \frac{1}{n}\,dn \ll \log^2(m).$$

For the second sum, $A(Y) \ll 1$ so that $A(Y)\log(Y)Y^{-\sigma} \to 0$ as $Y \to \infty$. Then Abel's summation formula (see Appendix B.3) gives

$$\sum_{n \geq m} \frac{\chi(n)\log(n)}{n^{\sigma}} = -A(m)\log(m)m^{-\sigma} - \int_m^{\infty} A(u)(1-\sigma\log(u))u^{-(\sigma+1)}\,du. \tag{9.3}$$

Since $0 \le 1-\sigma \le \frac{1}{\log(m)}$, we have $1-\sigma\log(u) \le \frac{\log(u)}{\log(m)}$. Also, we have the more precise estimate $|A(X)| \le m$ because χ is m-periodic and $|\chi(n)| \le 1$. With these estimates and Equation (9.3) we make the following computation:

$$\begin{aligned}
\left|\sum_{n\ge m}\frac{\chi(n)\log(n)}{n^\sigma}\right| &\le |A(m)|\log(m)m^{-\sigma} + \int_m^\infty |A(u)|(1-\sigma\log(u))u^{-(\sigma+1)}\,du \\
&\le |A(m)|\log(m)m^{-\sigma} + \log(m)\int_m^\infty |A(u)|\log(u)u^{-(\sigma+1)}\,du \\
&\le m^{1-\sigma}\log(m) + m\int_m^\infty \log(u)u^{-(\sigma+1)}\,du \\
&= m^{1-\sigma}\log(m) + m\left(-\log(u)\frac{u^{-\sigma}}{\sigma}\Big|_m^\infty + \int_m^\infty \frac{u^{-(\sigma+1)}}{s}\,du\right) \\
&= m^{1-\sigma}\log(m) + m\left(-\log(u)\frac{u^{-\sigma}}{\sigma} - \frac{u^{-\sigma}}{\sigma^2}\right)\Big|_m^\infty \\
&= m^{1-\sigma}\log(m) + m\left(\log(m)\frac{m^{-\sigma}}{\sigma} + \frac{m^{-\sigma}}{\sigma^2}\right) \\
&\ll m^{1-\sigma}\log(m) \\
&\ll e\log(m),
\end{aligned}$$

where in the fourth line we have used integration by parts and the last line holds because $1-\sigma \le \frac{1}{\log(m)}$. But $e\log(m) = O(\log^2(m))$ so the second sum is also $O(\log^2(m))$. Therefore we have shown $L'(\sigma,\chi) = O(\log^2(m))$ finishing the proof. $\square$

The second lemma is even easier and is proved in exactly the same way:

Lemma 9.2.2. *Let χ be a non-principal Dirichlet character modulo $m > 1$. Then $L(\sigma,\chi) = O(\log(m))$ for any σ such that $0 \le 1-\sigma \le \frac{1}{\log(m)}$.*

Proof. Note that

$$L(\sigma,\chi) = \sum_{n\ge 1}\frac{\chi(n)}{n^\sigma} = \sum_{n<m}\frac{\chi(n)}{n^\sigma} + \sum_{n\ge m}\frac{\chi(n)}{n^\sigma}.$$

It suffices to show that the last two sums are both $O(\log^2(m))$. For the first sum, since $n < m$, we have

$$\left|\frac{\chi(n)}{n^\sigma}\right| \le \frac{1}{n^\sigma} = \frac{n^{1-\sigma}}{n} < \frac{m^{1-\sigma}}{n} < \frac{e}{n},$$

where the last inequality follows because $1-\sigma \le \frac{1}{\log(m)}$. Therefore

$$\left|\sum_{n\le m}\frac{\chi(n)}{n^s}\right| < e\sum_{n<m}\frac{1}{n} < e\log(m)\int_1^m \frac{1}{n}\,dn \ll \log(m).$$

As for the second sum, setting $A(Y) = \sum_{n\le Y}\chi(n)$ we have $A(Y) \ll 1$ by Corollary 1.3.1 (i) and that χ is periodic. Thus $A(Y)Y^{-\sigma} \to 0$ as $Y \to \infty$. Then Abel's summation formula (see Appendix B.3) gives

$$\sum_{n\ge m}\frac{\chi(n)}{n^\sigma} = -A(m)m^{-\sigma} - \int_m^\infty A(u)u^{-(\sigma+1)}\,du. \tag{9.4}$$

Using the more precise estimate $|A(X)| \leq m$, because χ is m-periodic and $|\chi(n)| \leq 1$, with Equation (9.4), we make the following computation:

$$\begin{aligned}
\left| \sum_{n \geq m} \frac{\chi(n)\log(n)}{n^\sigma} \right| &\leq |A(m)|m^{-\sigma} + \int_m^\infty |A(u)|u^{-(\sigma+1)}\,du \\
&\leq |A(m)|m^{-\sigma} + \int_m^\infty |A(u)|\log(u)u^{-(\sigma+1)}\,du \\
&\leq m^{1-\sigma} + m\int_m^\infty \log(u)u^{-(\sigma+1)}\,du \\
&= m^{1-\sigma} + m\left(-\log(u)\frac{u^{-\sigma}}{\sigma}\bigg|_m^\infty + \int_m^\infty \frac{u^{-(\sigma+1)}}{s}\,du \right) \\
&= m^{1-\sigma} + m\left(-\log(u)\frac{u^{-\sigma}}{\sigma} - \frac{u^{-\sigma}}{\sigma^2} \right)\bigg|_m^\infty \\
&= m^{1-\sigma} + m\left(\log(m)\frac{m^{-\sigma}}{\sigma} + \frac{m^{-\sigma}}{\sigma^2} \right) \\
&\ll m^{1-\sigma} \\
&\ll e,
\end{aligned}$$

where in the fourth line we have used integration by parts and the last line holds because $1 - \sigma \leq \frac{1}{\log(m)}$. But $e = O(\log^2(m))$ so the second sum is also $O(\log^2(m))$. Therefore we have shown $L(\sigma, \chi) = O(\log(m))$ which completes the proof. $\square$

We will now prove the zero-free region version of Siegel's theorem, assuming the lower bound version, and using Lemma 9.2.1:

Proof of Siegel's theorem, zero-free region version. We we will prove the theorem by contradiction. Clearly the result holds for a single q, and notice that the result also holds provided we bound q from above by taking the maximum of all the $c_2(\varepsilon)$. Therefore we may suppose q is arbitrarily large. In this case, if there was a real zero β with $\beta \geq 1 - \frac{c_2(\varepsilon)}{q^\varepsilon}$, equivalently $1 - \beta \leq \frac{c_2(\varepsilon)}{q^\varepsilon}$, then for large enough q we have $0 \leq 1 - \beta \leq \frac{1}{\log(q)}$ so that $L'(\sigma, \chi) = O(\log^2(q))$ for $\beta \leq \sigma \leq 1$ by Lemma 9.2.1. These two estimates and the mean value theorem together give

$$L(1, \chi) = L(1, \chi) - L(\beta, \chi) = L'(\sigma, \chi)(1 - \beta) \ll \frac{\log^2(q)}{q^\varepsilon}.$$

Upon taking $\frac{\varepsilon}{2}$ in the lower bound version of Siegel's theorem, we obtain

$$\frac{1}{q^{\frac{\varepsilon}{2}}} \ll L(1, \chi) \ll \frac{\log^2(q)}{q^\varepsilon},$$

which is a contradiction for large q. $\square$

It remains to prove the lower bound version of Siegel's theorem. The idea is to combine two Dirichlet L-series attached to distinct characters with distinct moduli and use this new L-series to derivative a lower bound for a single Dirichlet L-function at $s = 1$. We first need a lemma:

Lemma 9.2.3. *Let χ_1 and χ_2 be two quadratic Dirichlet characters and let $L(s,g)$ be the L-series defined by*

$$L(s,g) = \zeta(s)L(s,\chi_1)L(s,\chi_2)L(s,\chi_1\chi_2).$$

Then $\Lambda_g(n) \geq 0$. In particular, $a_g(n) \geq 0$ and $a_g(0) = 1$.

Proof. For any prime p, the local roots at p are 1 with multiplicity one, $\chi_1(p)$ with multiplicity one, $\chi_2(p)$ with multiplicity one, and $\chi_1\chi_2(p)$ with multiplicity one. So for any $k \geq 1$, the sum of k-th powers of these local roots is

$$(1 + \chi_1^k(p))(1 + \chi_2^k(p)) \geq 0.$$

Thus $\Lambda_g(n) \geq 0$. It follows immediately from the Euler product of $L(s,g)$ that $a_g(n) \geq 0$ too. Also, it is clear from the Euler product of $L(s,g)$ that $a_g(0) = 1$. □

The key ingredient in the proof of the lower bound version of Siegel's theorem is an estimate for the L-series $L(s,g)$ in Lemma 9.2.3 relative to the modulus q_1q_2 in a small interval on the real axis close to 1. We now prove the theorem:

Proof of Siegel's theorem, lower bound version. Let χ_1 and χ_2 be two distinct primitive quadratic and non-principal characters modulo q_1 and q_2 respectively. Let $L(s,g)$ be the L-series defined by

$$L(s,g) = \zeta(s)L(s,\chi_1)L(s,\chi_2)L(s,\chi_1\chi_2).$$

Observe that $L(s,g)$ is holomorphic on $\mathbb{C}$ except for a simple pole at $s = 1$. Let λ be the residue at this pole so that

$$\lambda = L(1,\chi_1)L(1,\chi_2)L(1,\chi_1\chi_2).$$

Now $L(s,g)$ is represented as an absolutely convergent series for $\sigma > 1$ so that it has a power series about $s = 2$ with radius 1:

$$L(s,g) = \sum_{m\geq 0} \frac{L^{(m)}(2,g)}{m!}(s-2)^m,$$

for $|s-2| < 1$. We can compute $L^{(m)}(2,g)$ using the Dirichlet series by differentiating termwise:

$$L^{(m)}(2,g) = \frac{d^m}{ds^m}\left(\sum_{n\geq 1} \frac{a_g(n)}{n^s}\right)\Bigg|_{s=2} = (-1)^m \sum_{n\geq 1} \frac{a_g(n)\log^m(n)}{n^s}\Bigg|_{s=2} = (-1)^m \sum_{n\geq 1} \frac{a_g(n)\log^m(n)}{n^2}. \tag{9.5}$$

Since the $a_g(n)$ are nonnegative by Lemma 9.2.3, it follows that $L^{(m)}(2,g)$ is nonnegative and therefore we may write

$$L(s,g) = \sum_{m\geq 0} b_g(m)(2-s)^m,$$

for $|s-2| < 1$ and with $b_g(m)$ nonnegative. Also, Equation (9.5) and the fact that the $a_g(n)$ are nonnegative with $a_g(0) = 1$ together imply that $b_g(0) > 1$. Then

$$L(s,g) - \frac{\lambda}{s-1} = L(s,g) - \lambda\sum_{m\geq 0}(2-s)^m = \sum_{m\geq 0}(b_g(m) - \lambda)(2-s)^m, \tag{9.6}$$

and the last series must be absolutely convergent for say $|s-2| < 2$ because $L(s,g) - \frac{\lambda}{s-1}$ is holomorphic as we have removed the pole at $s = 1$. We now wish to estimate $L(s,g)$ and $\frac{\lambda}{s-1}$ on the circle $|s-2| = \frac{3}{2}$. Let

χ be a non-principal Dirichlet character modulo m and let $A(X) = \sum_{n \le X} \chi(n)$. Then Abel's summation formula and that $A(X) \ll 1$ (by Corollary 1.3.1 (i) and that χ is periodic) together imply

$$L(s,\chi) = s \int_1^\infty A(u) u^{-(s+1)}\, du,$$

for $\sigma > 0$. Now suppose $\sigma \ge \frac{1}{2}$. As $|A(X)| \le m$, we obtain

$$|L(s,\chi)| \le m|s| \int_1^\infty u^{-(\sigma+1)}\, du = -m|s| \frac{u^{-\sigma}}{\sigma}\bigg|_1^\infty = \frac{m|s|}{\sigma} \le 2m|s|.$$

In particular, on the disk $|s-2| \le \frac{3}{2}$ we have the estimates

$$L(s,\chi_1) \ll q_1, \quad L(s,\chi_2) \ll q_2, \quad \text{and} \quad L(s,\chi_1\chi_2) \ll q_1 q_2.$$

Since $\zeta(s)$ is bounded on the circle $|s-2| = \frac{3}{2}$ (it's a compact set) and $\lambda = L(1,\chi_1)L(1,\chi_2)L(1,\chi_1\chi_2)$, we obtain the bounds

$$L(s,g) \ll q_1^2 q_2^2 \quad \text{and} \quad \frac{\lambda}{s-1} \ll q_1^2 q_2^2,$$

on this circle as well. Cauchy's inequality for the size of coefficients of a power series applied to Equation (9.6) on the circle $|s-2| = \frac{3}{2}$ gives

$$b_g(m) - \lambda \ll q_1^2 q_2^2 \left(\frac{2}{3}\right)^m. \tag{9.7}$$

Let M be a positive integer. For real s with $\frac{7}{8} < s < 1$ we have $2 - s < \frac{9}{8}$ and using Equations (9.6) and (9.7) together we can upper bound the tail of $L(s,g) - \frac{\lambda}{s-1}$:

$$\begin{aligned}
\left| \sum_{m \ge M} (b_g(m) - \lambda)(2-s)^m \right| &\le \sum_{m \ge M} |b_g(m) - \lambda| (2-s)^m \\
&\ll q_1^2 q_2^2 \sum_{m \ge M} \left(\frac{2}{3}(2-s)\right)^m \\
&\ll q_1^2 q_2^2 \sum_{m \ge M} \left(\frac{3}{4}\right)^m \\
&\ll q_1^2 q_2^2 \left(\frac{3}{4}\right)^M \\
&\ll q_1^2 q_2^2 e^{-\frac{M}{4}},
\end{aligned}$$

where the last estimate follows because $(\frac{3}{4})^M < e^{-\frac{M}{4}}$ (which is equivalent to $\log\left(\frac{3}{4}\right) < -\frac{1}{4}$). Let c be the implicit constant. Using that the $b_g(m)$ are nonnegative, $b(0) > 1$, and the previous estimate for the tail, we can estimate $L(s,g) - \frac{\lambda}{s-1}$ from below for $\frac{7}{8} < s < 1$. Indeed, discarding the $b_g(m)$ terms for $1 \le m \le M$, bounding the constant term below by 1, and use the tail estimate, gives

$$L(s,g) - \frac{\lambda}{s-1} \ge 1 - \lambda \sum_{0 \le m \le M-1} (2-s)^m - c q_1^2 q_2^2 e^{-\frac{M}{4}} = 1 - \lambda \frac{(2-s)^M - 1}{1-s} - c q_1^2 q_2^2 e^{-\frac{M}{4}}, \tag{9.8}$$

which is valid for any positive integer M. Now chose M such that

$$\frac{1}{2}e^{-\frac{1}{4}} \leq cq_1^2q_2^2e^{-\frac{M}{4}} < \frac{1}{2}. \tag{9.9}$$

Upon isolating $L(s,g)$ in Equation (9.8) and using the second estimate in Equation (9.9), we get

$$L(s,g) \geq \frac{1}{2} - \lambda\frac{(2-s)^M}{1-s}. \tag{9.10}$$

Taking the logarithm of the first estimate in Equation (9.9) and isolating M, we obtain

$$M \leq 8\log(q_1q_2) + c, \tag{9.11}$$

for some different constant c. It follows that

$$(2-s)^M = e^{M\log(2-s)} < e^{M(1-s)} \leq c(q_1q_2)^{8(1-s)}, \tag{9.12}$$

for some different constant c, where in the first estimate we have used the Taylor series of the logarithm truncated at the first term and in the second estimate we have used Equation (9.11). Since $1-s$ is positive for $\frac{7}{8} < s < 1$, we can combine Equations (9.10) and (9.12) which gives

$$L(s,g) \geq \frac{1}{2} - \lambda\frac{c}{1-s}(q_1q_2)^{8(1-s)}. \tag{9.13}$$

This is our desired lower bound for $L(s,g)$. We will now choose the character χ_1. If there exists a Siegel zero β_1 with $1-\frac{\varepsilon}{16} < \beta_1 < 1$, let χ_1 be the character corresponding to the Dirichlet L-function that admits this Siegel zero. Then $L(\beta_1,g) = 0$ independent of the choice of χ_2. If there is no such Siegel zero, choose χ_1 to be any quadratic primitive character and β_1 to be any number such that $1-\frac{\varepsilon}{16} < \beta_1 < 1$. Then $L(\beta_1,g) < 0$ independent of the choice of χ_2. Indeed, $\zeta(s)$ is negative in this segment (actually for $0 \leq s < 1$) and each of the Dirichlet L-series defining $L(s,g)$ is positive at $s=1$ (the Euler product implies Dirichlet L-series are positive for $s > 1$ and they are in fact nonzero for $s=1$ by Theorem 9.1.3) and do not admit a zero for $\beta_1 < s \leq 1$ by our choice of β_1. In either case, $L(\beta_1,g) \leq 0$ so isolating λ and disregarding the constants in Equation (9.13) with $s=\beta_1$ gives the weaker estimate

$$\lambda \gg \frac{1-\beta_1}{(q_1q_2)^{8(1-\beta_1)}}. \tag{9.14}$$

We will now choose $\chi_2 = \chi$ and hence $q_2 = q$ as in the statement of the theorem. Notice that, independent of any work we have done, the theorem holds for a single q. Moreover, the theorem holds provided we bound q from above by taking the minimum of the $c_1(\varepsilon)$. Therefore we may assume q is arbitrarily large and in particular that $q > q_1$. Using Lemma 9.2.2 with $\sigma = 1$ applied to $L(\sigma,\chi_1)$ and $L(\sigma,\chi_1\chi)$, we obtain

$$\lambda \ll \log(q_1)\log(q_1q)L(1,\chi). \tag{9.15}$$

Combining Equations (9.14) and (9.15) yields

$$\frac{1-\beta_1}{(q_1q)^{8(1-\beta_1)}} \ll \log(q_1)\log(q_1q)L(1,\chi).$$

As β_1 and q_1 are fixed and $\log(q_1q) = O(\log(q))$, isolating $L(1,\chi)$ gives the weaker estimate

$$\frac{1}{q^{8(1-\beta_1)}\log(q)} \ll L(1,\chi). \tag{9.16}$$

But $1-\frac{\varepsilon}{16}<\beta_1<1$ so that $0<8(1-\beta_1)<\frac{\varepsilon}{2}$ which combined with Equation (9.16) yields

$$\frac{1}{q^{\varepsilon}} \ll_{\varepsilon} \frac{1}{q^{\frac{\varepsilon}{2}}\log(q)} \ll L(1,\chi),$$

where the first estimate follows because $\log(q) \ll_{\varepsilon} q^{\frac{\varepsilon}{2}}$ for sufficiently large q. This is equivalent to the statement in the theorem. □

The part of the proof of the lower bound version of Siegel's theorem which makes $c_1(\varepsilon)$ (and hence $c_2(\varepsilon)$) ineffective is the value of β_1. The choice of β_1 depends upon the existence of a Siegel zero near 1 and relative to the given ε. Since we don't know if Siegel zeros exist, this makes estimating β_1 relative to ε ineffective. Many results in analytic number theory make use of Siegel's theorem and hence are also ineffective. Many important problems investigate methods to get around using Siegel's theorem in favor of a weaker result that is effective. So far, no Siegel zero has been shown to exist or not exist for Dirichlet L-functions. But some progress has been made to showing that they are rare:

Proposition 9.2.1. *Let χ_1 and χ_2 be two distinct quadratic Dirichlet characters of conductor q_1 and q_2. If $L(s,\chi_1)$ and $L(s,\chi_2)$ have Siegel zeros β_1 and β_2 respectively and $\chi_1\chi_2$ is not principal, then there exists a positive constant c such that*

$$\min(\beta_1,\beta_2) < 1-\frac{c}{\log(q_1q_2)}.$$

Proof. We may assume χ_1 and χ_2 are primitive since if $\widetilde{\chi}_i$ is the primitive character inducing χ_i, for $i=1,2$, the only difference in zeros between $L(s,\chi_i)$ and $L(s,\widetilde{\chi}_i)$ occur on the line $\sigma=0$. Now let $\widetilde{\chi}$ be the primitive character of conductor q inducing $\chi_1\chi_2$. From Equation (8.9) with $\chi_1\chi_2$ in place of χ, we find that

$$\left|\frac{L'}{L}(s,\chi_1\chi_2)-\frac{L'}{L}(s,\widetilde{\chi})\right| = \left|\sum_{p\mid q_1q_2}\frac{\widetilde{\chi}(p)\log(p)p^{-s}}{1-\widetilde{\chi}(p)p^{-s}}\right| \le \sum_{p\mid q_1q_2}\frac{\log(p)p^{-\sigma}}{1-p^{-\sigma}} \le \sum_{p\mid q_1q_2}\log(p) \le \log(q_1q_2). \tag{9.17}$$

Let $s=\sigma$ with $1<\sigma\le 2$. Using the reverse triangle inequality, we deduce from Equation (9.17) that

$$-\frac{L'}{L}(\sigma,\chi_1\chi_2) < c\log(q_1q_2), \tag{9.18}$$

for some positive constant c. By Lemma 7.9.1 (iv) applied to $\zeta(s)$ while discarding all of the terms in both sums, we have

$$-\frac{\zeta'}{\zeta}(\sigma) < A+\frac{1}{\sigma-1}, \tag{9.19}$$

for some positive constant A. By Lemma 7.9.1 (iv) applied to $L(s,\chi_i)$ and only retaining the term corresponding to β_i, we have

$$-\frac{L'}{L}(\sigma,\chi_i) < A\log(q_i)+\frac{1}{\sigma-\beta_i}, \tag{9.20}$$

for $i=1,2$ and some possibly larger constant A. Now by Lemma 9.2.3, $-\frac{L'}{L}(\sigma,g)\ge 0$. Combining Equations (9.17) to (9.20) with this fact implies

$$0 < A+\frac{1}{\sigma-1}+A\log(q_1)-\frac{1}{\sigma-\beta_1}+A\log(q_2)-\frac{1}{\sigma-\beta_2}+c\log(q_1q_2).$$

Taking c larger, if necessary, we arrive at the simplified estimate

$$0 < \frac{1}{\sigma - 1} - \frac{1}{\sigma - \beta_1} - \frac{1}{\sigma - \beta_2} + c\log(q_1q_2),$$

which we rewrite as

$$\frac{1}{\sigma - \beta_1} + \frac{1}{\sigma - \beta_2} < \frac{1}{\sigma - 1} + c\log(q_1q_2),$$

Now let $\sigma = 1 + \frac{\delta}{\log(q_1q_2)}$ for some $\delta > 0$. Upon substituting, we have

$$\frac{1}{\sigma - \beta_1} + \frac{1}{\sigma - \beta_2} < \left(c + \frac{1}{\delta}\right)\log(q_1q_2).$$

If $\min(\beta_1, \beta_2) \geq 1 - \frac{c}{\log(q_1q_2)}$, then we arrive at

$$2(\delta + c) < c + \frac{1}{\delta},$$

which is a contradiction if we take δ small enough so that $2\delta^2 + c\delta < 1$. □

From Proposition 9.2.1 we immediately see that for every modulus $m > 1$ there is at most one primitive quadratic Dirichlet character that can admit a Siegel zero:

Proposition 9.2.2. *For every integer $m > 1$, there is at most one Dirichlet character χ modulo m such that $L(s, \chi)$ has a Siegel zero. If this Siegel zero exists, χ is necessarily quadratic.*

Proof. Let $\widetilde{\chi}$ be the primitive character inducing χ. As the zeros of $L(s, \chi)$ and $L(s, \widetilde{\chi})$ differ only on the line $\sigma = 0$, Theorem 9.2.2 implies that χ must be quadratic. Suppose χ_1 and χ_2 are two distinct character modulo m, of conductors q_1 and q_2, admitting Siegel zeros β_1 and β_2. Then $\chi_1\chi_2 \neq \chi_{m,0}$. Moreover, $\beta_1 \geq 1 - \frac{c_1}{\log(q_1)}$ and $\beta_2 \geq 1 - \frac{c_2}{\log(q_2)}$ for some positive constants c_1 and c_2. Taking c smaller, if necessary, we have $\min(\beta_1, \beta_2) \geq 1 - \frac{c}{\log(q_1q_2)}$ which contradicts Proposition 9.2.1. □

Todo: [The Analytic Class Number Formula]

9.3 The Prime Number Theorem

The function $\psi(x)$ is defined by

$$\psi(x) = \sum_{n \leq x} \Lambda(n),$$

for $x > 0$. We will obtain an explicit formula for $\psi(x)$ analogous to the explicit formula for the Riemann zeta function. The explicit formula for $\psi(x)$ will be obtained by applying truncated Perron's formula to the logarithmic derivative of $\zeta(s)$. Since $\psi(x)$ is discontinuous when x is a prime power, we need to work with a slightly modified function to apply the Mellin inversion formula. Define $\psi_0(x)$ by

$$\psi_0(x) = \begin{cases} \psi(x) & \text{if } x \text{ is not a prime power,} \\ \psi(x) - \frac{1}{2}\Lambda(x) & \text{if } x \text{ is a prime power.} \end{cases}$$

Equivalently, $\psi_0(x)$ is $\psi(x)$ except that its value is halfway between the limit values when x is a prime power. Stated another way, if x is a prime power the last term in the sum for $\psi_0(x)$ is multiplied by $\frac{1}{2}$. The **explicit formula** for $\psi(x)$ is the following:

Theorem 9.3.1 (Explicit formula for $\psi(x)$). *For $x \geq 2$,*

$$\psi_0(x) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} - \frac{\zeta'}{\zeta}(0) - \frac{1}{2}\log(1 - x^{-2}),$$

where the sum is counted with multiplicity and ordered with respect to the size of the ordinate.

A few comments are in order before we prove the explicit formula for $\psi(x)$. First, since ρ is conjectured to be of the form $\rho = \frac{1}{2} + i\gamma$ via the Riemann hypothesis for the Riemann zeta function, x is conjectured to be the main term in the explicit formula. The constant $\frac{\zeta'}{\zeta}(0)$ can be shown to be $\log(2\pi)$ (see [Dav80] for a proof). Also, using the Taylor series of the logarithm, the last term can be expressed as

$$\frac{1}{2}\log(1 - x^{-2}) = \frac{1}{2}\sum_{m \geq 1}(-1)^{m-1}\frac{(-x^{-2})^m}{m} = \sum_{m \geq 1}(-1)^{2m-1}\frac{x^{-2m}}{2m} = \sum_{m \geq 1}\frac{x^{-2m}}{-2m} = \sum_{\omega}\frac{x^{\omega}}{\omega},$$

where ω runs over the trivial zeros of $\zeta(s)$. We will now prove the explicit formula for $\psi(x)$:

Proof of the explicit formula for $\psi(x)$. Applying truncated Perron's formula to $-\frac{\zeta'}{\zeta}(s)$ gives

$$\psi_0(x) - J(x,T) \ll x^c \sum_{\substack{n \geq 1 \\ n \neq x}} \frac{\Lambda(n)}{n^c} \min\left(1, \frac{1}{T\left|\log\left(\frac{x}{n}\right)\right|}\right) + \delta_x \Lambda(x)\frac{c}{T}, \tag{9.21}$$

where

$$J(x,T) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} -\frac{\zeta'}{\zeta}(s)x^s\,\frac{ds}{s},$$

$c > 1$, and it is understood that $\delta_x = 0$ unless x is a prime power. Take $T > 2$ not coinciding with the ordinate of a nontrivial zero and let $c = 1 + \frac{1}{\log(x^2)}$ so that $x^c = \sqrt{e}x$ and $1 < c < 2$. The first step is to estimate the right-hand side of Equation (9.21). We deal with the terms corresponding to n such that n is bounded away from x before anything else. So suppose $n \leq \frac{3}{4}x$ or $n \geq \frac{5}{4}x$. For these n, $\log\left(\frac{x}{n}\right)$ is bounded away from zero so that their contribution is

$$\ll \frac{x^c}{T}\sum_{n \geq 1}\frac{\Lambda(n)}{n^c} \ll \frac{x^c}{T}\left(-\frac{\zeta'}{\zeta}(c)\right) \ll \frac{x\log(x)}{T}, \tag{9.22}$$

where the last estimate follows from Lemma 7.9.1 (iv) applied to $\zeta(s)$ while discarding all of the terms in both sums and our choice of c (in particular $\log(c) \ll \log(x)$). Now we deal with the terms n close to x. Consider those n for which $\frac{3}{4}x < n < x$. Let x_1 be the largest prime power less than x. We may also suppose $\frac{3}{4}x < x_1 < x$ since otherwise $\Lambda(n) = 0$ and these terms do not contribute anything. Moreover, $\frac{x^c}{n^c} \ll 1$. For the term $n = x_1$, we have

$$\log\left(\frac{x}{n}\right) = -\log\left(1 - \frac{x - x_1}{x}\right) \geq \frac{x - x_1}{x},$$

where we have obtained the inequality by using Taylor series of the logarithm truncated after the first term. The contribution of this term is then

$$\ll \Lambda(x_1)\min\left(1, \frac{x}{T(x - x_1)}\right) \ll \log(x)\min\left(1, \frac{x}{T(x - x_1)}\right). \tag{9.23}$$

For the other such n, we can write $n = x_1 - v$, where v is an integer satisfying $0 < v < \frac{1}{4}x$, so that

$$\log\left(\frac{x}{n}\right) \geq \log\left(\frac{x_1}{n}\right) = -\log\left(1 - \frac{v}{x_1}\right) \geq \frac{v}{x_1},$$

where we have obtained the latter inequality by using Taylor series of the logarithm truncated after the first term. The contribution for these n is then

$$\ll \sum_{0<v<\frac{1}{4}x} \Lambda(x_1 - v)\frac{x_1}{Tv} \ll \frac{x}{T}\sum_{0<v<\frac{1}{4}x} \frac{\Lambda(x_1 - v)}{v} \ll \frac{x\log(x)}{T}\sum_{0<v<\frac{1}{4}x} \frac{1}{v} \ll \frac{x\log^2(x)}{T}. \tag{9.24}$$

The contribution for those n for which $x < n < \frac{5}{4}x$ is handled in exactly the same way with x_1 being the least prime power larger than x. Let $\langle x \rangle$ be the distance between x and the nearest prime power other than x if x itself is a prime power. Combining Equations (9.23) and (9.24) with our previous comment, the contribution for those n with $\frac{3}{4}x < n < \frac{5}{4}x$ is

$$\ll \frac{x\log^2(x)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x \rangle}\right). \tag{9.25}$$

Putting Equations (9.22) and (9.25) together and noticing that the error term in Equation (9.22) is absorbed by the second error term in Equation (9.25), we obtain

$$\psi_0(x) - J(x,T) \ll \frac{x\log^2(x)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x \rangle}\right). \tag{9.26}$$

This is the first part of the proof. Now we estimate $J(x,T)$ by appealing to the residue theorem. Let $U \geq 1$ be an odd integer. Let Ω be the region enclosed by the contours $\eta_1, \ldots, \eta_4$ in Figure 9.1 and set $\eta = \sum_{1\leq i\leq 4} \eta_i$ so that $\eta = \partial\Omega$.

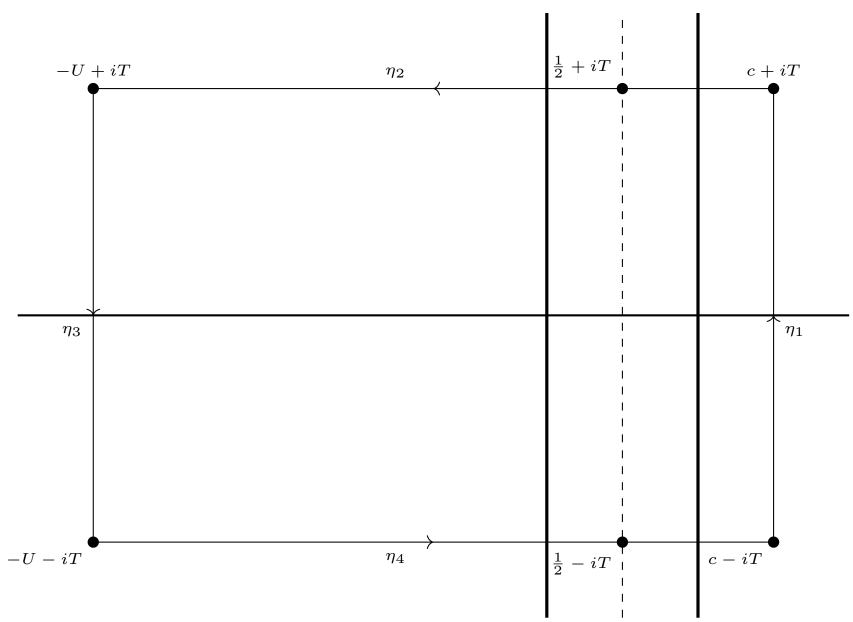

Figure 9.1: Contour for the explicit formula for $\psi(x)$

We may express $J(x,T)$ as

$$J(x,T) = \frac{1}{2\pi i}\int_{\eta_1} -\frac{\zeta'}{\zeta}(s)x^s\,\frac{ds}{s}.$$

The residue theorem together with the formula for the negative logarithmic derivative in Proposition 7.8.1 applied to $\zeta(s)$ and Corollary 1.7.1 imply

$$J(x,T) = x - \sum_{|\gamma|<T} \frac{x^\rho}{\rho} - \frac{\zeta'}{\zeta}(0) - \sum_{0<2m<U} \frac{x^{-2m}}{-2m} + \frac{1}{2\pi i}\int_{\eta_2+\eta_3+\eta_4} -\frac{\zeta'}{\zeta}(s)x^s\,\frac{ds}{s}, \tag{9.27}$$

where $\rho = \beta + i\gamma$ is a nontrivial zero of ζ. We will estimate $J(x,T)$ by estimating the remaining integral. By Lemma 7.9.1 (ii) applied to $\zeta(s)$, the number of nontrivial zeros satisfying $|\gamma - T| < 1$ is $O(\log(T))$. Among the ordinates of these nontrivial zeros, there must be a gap of size $\gg \frac{1}{\log(T)}$. Upon varying T by a bounded amount (we are varying in the interval $[T-1, T+1]$) so that it belongs to this gap, we can additionally ensure

$$\gamma - T \gg \frac{1}{\log(T)},$$

for all the nontrivial zeros of $\zeta(s)$. To estimate part of the horizontal integrals over η_2 and η_4, Lemma 7.9.1 (iv) applied to $\zeta(s)$ gives

$$\frac{\zeta'}{\zeta}(s) = \sum_{|\gamma-T|<1} \frac{1}{s-\rho} + O(\log(T)),$$

on the parts of these segments with $-1 \le \sigma \le 2$. By our choice of T, $|s-\rho| \ge |\gamma - T| \gg \frac{1}{\log(T)}$ so that each term in the sum is $O(\log(T))$. There are at most $O(\log(T))$ such terms by Lemma 7.9.1 (ii) applied to $\zeta(s)$, so we find that

$$\frac{\zeta'}{\zeta}(s) = O(\log^2(T)),$$

on the parts of these segments with $-1 \le \sigma \le 2$. It follows that the parts of the horizontal integrals over η_2 and η_4 with $-1 \le \sigma \le c$ (recall $c < 2$) contribute

$$\ll \frac{\log^2(T)}{T}\int_{-1}^{c} x^\sigma\,d\sigma \ll \frac{\log^2(T)}{T}\int_{-\infty}^{c} x^\sigma\,d\sigma \ll \frac{x\log^2(T)}{T\log(x)}, \tag{9.28}$$

where in the last estimate we have used the choice of c. To estimate the remainder of the horizontal integrals, we need a bound for $\frac{\zeta'}{\zeta}(s)$ when $\sigma < -1$ and away from the trivial zeros. To this end, write the functional equation for $\zeta(s)$ in the form

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

and take the logarithmic derivative to get

$$\frac{\zeta'}{\zeta}(s) = \log(\pi) + \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{1-s}{2}\right) - \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s}{2}\right) + \frac{\zeta'}{\zeta}(1-s).$$

Let s be such that $\sigma < -1$ and suppose s is distance $\frac{1}{2}$ away from the trivial zeros. We will estimate every term on the right-hand side of the previous identity. The first term is constant and the last term is bounded since it is an absolutely convergent Dirichlet series. As for the digamma terms, since s is away from the trivial zeros, Proposition 1.7.3 implies $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{1-s}{2}\right) = O(\log|1-s|)$ and $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s}{2}\right) = O(\log|s|)$. However, as $\sigma < -1$ and s is away from the trivial zeros, s and $1-s$ are bounded away from zero so that $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{1-s}{2}\right) = O(\log|s|)$. Putting these estimates together gives

$$\frac{\zeta'}{\zeta}(s) \ll \log(|s|), \tag{9.29}$$

for $\sigma < -1$. Using Equation (9.29), the parts of the horizontal integrals over η_2 and η_4 with $-U \le \sigma \le -1$ contribute

$$\ll \frac{\log(T)}{T} \int_{-U}^{-1} x^{\sigma}\, d\sigma \ll \frac{\log(T)}{Tx\log(x)}. \tag{9.30}$$

Combining Equations (9.28) and (9.30) gives

$$\frac{1}{2\pi i} \int_{\eta_2+\eta_4} -\frac{\zeta'}{\zeta}(s) x^s \frac{ds}{s} \ll \frac{x\log^2(T)}{T\log(x)} + \frac{\log(T)}{Tx\log(x)} \ll \frac{x\log^2(T)}{T\log(x)}. \tag{9.31}$$

To estimate the vertical integral, we use Equation (9.29) again to conclude that

$$\frac{1}{2\pi i} \int_{\eta_3} -\frac{\zeta'}{\zeta}(s) x^s \frac{ds}{s} \ll \frac{\log(U)}{U} \int_{-T}^{T} x^{-U}\, dt \ll \frac{T\log(U)}{Ux^U}. \tag{9.32}$$

Combining Equations (9.27), (9.31) and (9.32) and taking the limit as $U \to \infty$, the error term in Equation (9.32) vanishes and the sum over m in Equation (9.27) evaluates to $\frac{1}{2}\log(1-x^{-2})$ (as we have already mentioned) giving

$$J(x,T) = x - \sum_{|\gamma|<T} \frac{x^{\rho}}{\rho} - \frac{\zeta'}{\zeta}(0) - \frac{1}{2}\log(1-x^{-2}) + \frac{x\log^2(T)}{T\log(x)}. \tag{9.33}$$

Substituting Equation (9.33) into Equation (9.26), we at last obtain

$$\psi_0(x) = x - \sum_{|\gamma|<T} \frac{x^{\rho}}{\rho} - \frac{\zeta'}{\zeta}(0) - \frac{1}{2}\log(1-x^{-2}) + \frac{x\log^2(xT)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right), \tag{9.34}$$

where the second to last term on the right-hand side is obtained by combining the error term in Equation (9.31) with the first error term in Equation (9.26). The theorem follows by taking the limit as $T \to \infty$. □

Note that the convergence of the right-hand side in the explicit formula for $\psi(x)$ is uniform in any interval not containing a prime power since $\psi(x)$ is continuous there. Moreover, we have an approximate formula for $\psi(x)$ as a corollary:

Corollary 9.3.1. *For $x \ge 2$ and $T > 2$,*

$$\psi_0(x) = x - \sum_{|\gamma|<T} \frac{x^{\rho}}{\rho} + R(x,T),$$

where ρ runs over the nontrivial zeros of $\zeta(s)$ counted with multiplicity and ordered with respect to the size of the ordinate, and

$$R(x,T) \ll \frac{x\log^2(xT)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right),$$

where $\langle x\rangle$ is the distance between x and the nearest prime power other than x if x itself is a prime power. Moreover, if x is an integer, we have the simplified estimate

$$R(x,T) \ll \frac{x\log^2(xT)}{T}.$$

Proof. This follows from Equation (9.34) since $\frac{\zeta'}{\zeta}(0)$ is constant and $\frac{1}{2}\log(1-x^2)$ is bounded for $x \ge 2$. If x is an integer, then $\langle x \rangle \ge 1$ so that $\log(x)\min\left(1, \frac{x}{T\langle x \rangle}\right) \le \frac{x\log(x)}{T}$ and this term can be absorbed into $O\left(\frac{x\log^2(xT)}{T}\right)$. □

With our refined explicit formula in hand, we are ready to discuss and prove the prime number theorem. The **prime counting function** $\pi(x)$ is defined by

$$\pi(x) = \sum_{p \le x} 1,$$

for $x > 0$. Equivalently, $\pi(x)$ counts the number of primes that no larger than x. Euclid's infinitude of the primes is equivalent to $\pi(x) \to \infty$ as $x \to \infty$. A more interesting question is to ask how the primes are distributed among the integers. The classical **prime number theorem** answers this question and the precise statement is the following:

Theorem 9.3.2 (Prime number theorem, classical version). *For $x \ge 2$,*

$$\pi(x) \sim \frac{x}{\log(x)}.$$

We will delay the proof for the moment and give some intuition and historical context to the result. Intuitively, the prime number theorem is a result about how dense the primes are in the integers. To see this, notice that the result is equivalent to the asymptotic

$$\frac{\pi(x)}{x} \sim \frac{1}{\log(x)}.$$

Letting $x \ge 2$, the left-hand side is the probability that a randomly chosen positive integer no larger than x is prime. Thus the asymptotic result says that for large enough x, the probability that a randomly chosen integer no larger than x is prime is approximately $\frac{1}{\log(x)}$. We can also interpret this as saying that the average gap between primes no larger than x is approximately $\frac{1}{\log(x)}$. As a consequence, a positive integer with at most $2n$ digits is about half as likely to be prime than a positive integer with at most n digits. Indeed, there are $10^n - 1$ numbers with at most n digits, $10^{2n} - 1$ with at most $2n$ digits, and $\log(10^{2n} - 1)$ is approximately $2\log(10^n)$. Note that the prime number theorem says nothing about the exact error $\pi(x) - \frac{x}{\log(x)}$ as $x \to \infty$. The theorem only says that the relative error tends to zero:

$$\lim_{x \to \infty} \frac{\pi(x) - \frac{x}{\log(x)}}{\frac{x}{\log(x)}} = 0.$$

Now for some historical context. While Gauss was not the first to put forth a conjectural form of the prime number theorem, he was known for compiling extensive tables of primes and he suspected that the density of the primes up to x was roughly $\frac{1}{\log(x)}$. How might one suspect this is the correct density? Well, let $d\delta_p$ be the weighted point measure that assigns $\frac{1}{p}$ at the prime p and zero everywhere else. Then

$$\sum_{p \le x} \frac{1}{p} = \int_1^x d\delta_p(u).$$

We can interpret the integral as integrating the density $d\delta_p$ over $[1,x]$. Let's try and find a more explicit expression for the density $d\delta_p$. Euler (see [Eul44]), argued

$$\sum_{p\leq x}\frac{1}{p}\sim\log\log(x).$$

But notice that

$$\log\log(x)=\int_1^{\log(x)}\frac{du}{u}=\int_e^x\frac{1}{u}\frac{du}{\log u},$$

where in the second equality we have made the change of variables $u\to\log(u)$. So altogether,

$$\sum_{p\leq x}\frac{1}{p}\sim\int_e^x\frac{1}{u}\frac{du}{\log u}.$$

This is an asymptotic that gives a more explicit representation of the density $d\delta_p$. Notice that both sides of this asymptotic are weighted the same, the left-hand side by $\frac{1}{p}$, and the right-hand side by $\frac{1}{u}$. If we remove these weight (this is not strictly allowed), then we might hope

$$\pi(x)=\sum_{p\leq x}1\sim\int_e^x\frac{du}{\log(u)}.$$

Accordingly, we define the **logarithmic integral** $\mathrm{Li}(x)$ by

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

for $x\geq 2$. Notice that $\mathrm{Li}(x)\sim\frac{x}{\log x}$ because

$$\lim_{x\to\infty}\left|\frac{\mathrm{Li}(x)}{\frac{x}{\log x}}\right|=\lim_{x\to\infty}\left|\frac{\int_2^x\frac{dt}{\log(t)}}{\frac{x}{\log x}}\right|=\lim_{x\to\infty}\left|\frac{\frac{1}{\log(x)}}{\frac{\log(x)-1}{\log^2(x)}}\right|=\lim_{x\to\infty}\left|\frac{\log(x)}{\log(x)-1}\right|=1.$$

where in the second equality we have used L'Hôpital's rule. So an equivalent statement is the logarithmic integral version of the **prime number theorem**:

Theorem 9.3.3 (Prime number theorem, logarithmic integral version). *For $x\geq 2$,*

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

Interpreting the logarithmic integral as an integral of density, then for large x the density of primes up to x is approximately $\frac{1}{\log(x)}$ which is what both versions of the prime number theorem claim. Legendre was the first to put forth a conjectural form of the prime number theorem. In 1798 (see [Leg98]) he claimed that $\pi(x)$ was of the form

$$\frac{x}{A\log(x)+B},$$

for some constants A and B. In 1808 (see [Leg08]) he refined his conjecture by claiming

$$\frac{x}{\log(x)+A(x)},$$

where $\lim_{x\to\infty} A(x) \approx 1.08366$. Riemann's 1859 manuscript (see [Rie59]) contains an outline for how to prove the prime number theorem, but it was not until 1896 that the prime number theorem was proved independently by Hadamard and de la Vallée Poussin (see [Had96] and [Pou97]). Their proofs, as well as every proof thereon out until 1949, used complex analytic methods in an essential way (there are now elementary proofs due to Erdös and Selberg). We are now ready to prove the prime number theorem. Strictly speaking, we will prove the absolute error version of the **prime number theorem**, due to de la Vallée Poussin, which bounds the absolute error between $\pi(x)$ and $\mathrm{Li}(x)$:

Theorem 9.3.4 (Prime number theorem, absolute error version). *For $x \geq 2$, there exists a positive constant c such that*

$$\pi(x) = \mathrm{Li}(x) + O\left(xe^{-c\sqrt{\log(x)}}\right).$$

Proof. It suffices to assume x is an integer, because $\pi(x)$ can only change value at integers and the other functions in the statement are increasing. We will first prove

$$\psi(x) = x + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.35}$$

for some positive constant c. To achieve this, we estimate the sum over the nontrivial zeros of $\zeta(s)$ in Corollary 9.3.1. So let $T > 2$ not coinciding with the ordinate of a nontrivial zero, and suppose $\rho = \beta + i\gamma$ is a nontrivial zero with $|\gamma| < T$. By Theorem 9.2.1, we know $\beta < 1 - \frac{c}{\log(T)}$ for some positive constant c. It follows that

$$|x^\rho| = x^\beta < x^{1-\frac{c}{\log(T)}} = xe^{-c\frac{\log(x)}{\log(T)}}. \tag{9.36}$$

As $|\rho| > |\gamma|$, letting $\gamma_1 > 0$ (which is bounded away from zero since the zeros of $\zeta(s)$ are discrete and we know that there is no real nontrivial zero) be the ordinate of the first nontrivial zero, applying integration by parts gives

$$\sum_{|\gamma|<T} \frac{1}{\rho} \ll \sum_{\gamma_1 \leq |\gamma| < T} \frac{1}{\gamma} \ll \int_{\gamma_1}^{T} \frac{dN(t)}{t} = \frac{N(T)}{T} + \int_{\gamma_1}^{T} \frac{N(t)}{t^2}\,dt \ll \log^2(T), \tag{9.37}$$

where in the last estimate we have used that $N(t) \ll t\log(t)$ which follows from Corollary 7.9.1. Putting Equations (9.36) and (9.37) together gives

$$\sum_{|\gamma|<T} \frac{x^\rho}{\rho} \ll x\log^2(T)e^{-c\frac{\log(x)}{\log(T)}}. \tag{9.38}$$

As $\psi(x) \sim \psi_0(x)$ and x is an integer, Equation (9.38) with Corollary 9.3.1 together imply

$$\psi(x) - x \ll x\log^2(T)e^{-c\frac{\log(x)}{\log(T)}} + \frac{x\log^2(xT)}{T}. \tag{9.39}$$

We will now let T be determined by

$$\log^2(T) = \log(x),$$

or equivalently,

$$T = e^{\sqrt{\log(x)}}.$$

With this choice of T (note that if $x \geq 2$ then $T > 2$), we can estimate Equation (9.39) as follows:

$$\begin{aligned}\psi(x) - x &\ll x\log(x)e^{-c\sqrt{\log(x)}} + x\left(\log^2(x) + \log(x)\right)e^{-\sqrt{\log(x)}} \\ &\ll x\log(x)e^{-c\sqrt{\log(x)}} + x\log^2(x)e^{-\sqrt{\log(x)}} \\ &\ll x\log^2(x)e^{-\min(1,c)\sqrt{\log(x)}}.\end{aligned}$$

As $\log(x) \ll_\varepsilon e^{-\varepsilon\sqrt{\log(x)}}$, we conclude that

$$\psi(x) - x \ll xe^{-c\sqrt{\log(x)}},$$

for some smaller c with $c < 1$. This is equivalent to Equation (9.35). Now let

$$\pi_1(x) = \sum_{n \le x} \frac{\Lambda(n)}{\log(n)}.$$

We can write $\pi_1(x)$ in terms of $\psi(x)$ as follows:

$$\begin{aligned}
\pi_1(x) &= \sum_{n \le x} \frac{\Lambda(n)}{\log(n)} \\
&= \sum_{n \le x} \Lambda(n) \int_n^x \frac{dt}{t\log^2(t)} + \frac{1}{\log(x)} \sum_{n \le x} \Lambda(n) \\
&= \int_2^x \sum_{n \le t} \Lambda(n) \frac{dt}{t\log^2(t)} + \frac{1}{\log(x)} \sum_{n \le x} \Lambda(n) \\
&= \int_2^x \frac{\psi(t)}{t\log^2(t)}\, dt + \frac{\psi(x)}{\log(x)}.
\end{aligned}$$

Applying Equation (9.35) to the last expression yields

$$\pi_1(x) = \int_2^x \frac{t}{t\log^2(t)}\, dt + \frac{x}{\log(x)} + O\left(\int_2^x \frac{e^{-c\sqrt{\log(t)}}}{\log^2(t)}\, dt + \frac{xe^{-c\sqrt{\log(x)}}}{\log(x)}\right). \tag{9.40}$$

Upon applying integrating by parts to the main term in Equation (9.40), we obtain

$$\int_2^x \frac{t}{t\log^2(t)}\, dt + \frac{x}{\log(x)} = \int_2^x \frac{dt}{\log(t)} + \frac{2}{\log(2)} = \mathrm{Li}(x) + \frac{2}{\log(2)}. \tag{9.41}$$

As for the error term in Equation (9.40), $\log^2(t)$ and $\log(x)$ are both bounded away from zero so that

$$\int_2^x \frac{e^{-c\sqrt{\log(t)}}}{\log^2(t)}\, dt + \frac{xe^{-c\sqrt{\log(x)}}}{\log(x)} \ll \int_2^x e^{-c\sqrt{\log(t)}}\, dt + xe^{-c\sqrt{\log(x)}}.$$

For $t \le x^{\frac{1}{4}}$, we use the bound $e^{-c\sqrt{\log(t)}} < 1$ so that

$$\int_2^{x^{\frac{1}{4}}} e^{-c\sqrt{\log(t)}}\, dt < \int_2^{x^{\frac{1}{4}}} dt \ll x^{\frac{1}{4}}.$$

For $t > x^{\frac{1}{4}}$, $\sqrt{\log(t)} > \frac{1}{2}\sqrt{\log(x)}$ and thus

$$\int_2^{x^{\frac{1}{4}}} e^{-c\sqrt{\log(t)}}\, dt \le e^{-c\frac{1}{2}\sqrt{\log(x)}} \int_2^{x^{\frac{1}{4}}} dt \ll x^{\frac{1}{4}} e^{-c\frac{1}{2}\sqrt{\log(x)}}.$$

All of these estimates together imply

$$\int_2^x \frac{e^{-c\sqrt{\log(t)}}}{\log^2(t)}\, dt + \frac{xe^{-c\sqrt{\log(x)}}}{\log(x)} \ll xe^{-c\sqrt{\log(x)}}, \tag{9.42}$$

for some smaller c. Combining Equations (9.40) to (9.42) yields

$$\pi_1(x) = \mathrm{Li}(x) + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.43}$$

where the constant in Equation (9.41) has been absorbed into the error term. We now pass from $\pi_1(x)$ to $\pi(x)$. If p is a prime such that $p^m < x$ for some $m \geq 1$, then $p < x^{\frac{1}{2}} < x^{\frac{1}{3}} < \cdots < x^{\frac{1}{m}}$. Therefore

$$\pi_1(x) = \sum_{n \leq x} \frac{\Lambda(n)}{\log(n)} = \sum_{p^m \leq x} \frac{\log(p)}{m\log(p)} = \pi(x) + \frac{1}{2}\pi(x^{\frac{1}{2}}) + \cdots. \tag{9.44}$$

Moreover, as $\pi(x^{\frac{1}{n}}) < x^{\frac{1}{n}}$ for any $n \geq 1$, we see that $\pi(x) - \pi_1(x) = O(x^{\frac{1}{2}})$. This estimate together with Equations (9.43) and (9.44) gives

$$\pi(x) = \mathrm{Li}(x) + O\left(xe^{-c\sqrt{\log(x)}}\right),$$

because $x^{\frac{1}{2}} \ll xe^{-c\sqrt{\log(x)}}$. This completes the proof. □

The proof of the logarithmic integral and classical versions of the prime number theorem are immediate:

Proof of prime number theorem, logarithmic integral and classical versions. By the absolute error version of the prime number theorem,

$$\pi(x) = \mathrm{Li}(x)\left(1 + O\left(\frac{xe^{-c\sqrt{\log(x)}}}{\mathrm{Li}(x)}\right)\right).$$

But we have shown $\mathrm{Li}(x) \sim \frac{x}{\log(x)}$ so that

$$\frac{xe^{-c\sqrt{\log(x)}}}{\mathrm{Li}(x)} \sim \log(x)e^{-c\sqrt{\log(x)}} = o(1),$$

where the equality holds since $\log(x) \ll_\varepsilon e^{-\varepsilon\sqrt{\log(x)}}$. The logarithm integral version follows. The classical version also holds using the asymptotic $\mathrm{Li}(x) \sim \frac{x}{\log(x)}$. □

In the proof of the logarithmic integral and classical versions of the prime number theorem, we saw that $xe^{-c\sqrt{\log(x)}} < \frac{x}{\log(x)}$ for sufficiently large x. Therefore the exact error $\pi(x) - \mathrm{Li}(x)$ grows slower than $\pi(x) - \frac{x}{\log x}$ for sufficiently large x. This means that $\mathrm{Li}(x)$ is a better numerical approximation to $\pi(x)$ than $\frac{x}{\log(x)}$. There is also the following result due to Hardy and Littlewood (see [HL16]) which gives us more information:

Proposition 9.3.1. *$\pi(x) - \mathrm{Li}(x)$ changes sign infinitely often as $x \to \infty$.*

So in addition, Proposition 9.3.1 implies that $\mathrm{Li}(x)$ never underestimates or overestimates $\pi(x)$ continuously. On the other hand, the exact error $\pi(x) - \frac{x}{\log(x)}$ is positive provided $x \geq 17$ (see [RS62]). It is also worthwhile to note that in 1901 Koch showed that the Riemann hypothesis improves the error term in the absolute error version of the prime number theorem (see [Koc01]):

Proposition 9.3.2. *For $x \geq 2$, we have*

$$\pi(x) = \mathrm{Li}(x) + O(\sqrt{x}\log(x)),$$

provided the Riemann hypothesis for the Riemann zeta function holds.

Proof. If ρ is a nontrivial zero of $\zeta(s)$, the Riemann hypothesis implies $|x^\rho| = \sqrt{x}$. Therefore as in the proof of the absolute error version of the prime number theorem,

$$\sum_{|\gamma|<T} \frac{x^\rho}{\rho} \ll \sqrt{x}\log^2(T),$$

for $T > 2$ not coinciding with the ordinate of a nontrivial zero. Repeating the same argument with T determined by

$$T^2 = x,$$

gives

$$\psi(x) = x + O(\sqrt{x}\log^2(x)),$$

and then transferring to $\pi_1(x)$ and finally $\pi(x)$ gives

$$\pi(x) = x + O(\sqrt{x}\log(x)).$$

□

9.4 The Siegel-Walfisz Theorem

Let χ be a Dirichlet character modulo $m > 1$. The function $\psi(x, \chi)$ is defined by

$$\psi(x, \chi) = \sum_{n \le x} \chi(n)\Lambda(n),$$

for $x > 0$. This function plays the analogous role of $\psi(x)$ but for Dirichlet L-series. Accordingly, we will derive an explicit formula for $\psi(x, \chi)$ in a similar manner to that of $\psi(x)$. Because $\psi(x, \chi)$ is discontinuous when x is a prime power, we also introduce a slightly modified function. Define $\psi_0(x, \chi)$ by

$$\psi_0(x, \chi) = \begin{cases} \psi(x, \chi) & \text{if } x \text{ is not a prime power,} \\ \psi(x, \chi) - \frac{1}{2}\chi(x)\Lambda(x) & \text{if } x \text{ is a prime power.} \end{cases}$$

Thus $\psi_0(x, \chi)$ is $\psi(x, \chi)$ except that its value is halfway between the limit values when x is a prime power. We will also need to define a particular constant that will come up. For a character χ, define $b(\chi)$ to be $\frac{L'}{L}(0, \chi)$ if χ is odd and to be the constant term in the Laurent series of $\frac{L'}{L}(s, \chi)$ if χ is even (as in the even case $\frac{L'}{L}(s, \chi)$ has a pole at $s = 0$). The **explicit formula** for $\psi(x, \chi)$ is the following:

Theorem 9.4.1 (Explicit formula for $\psi(x, \chi)$). *Let χ be a primitive Dirichlet character of conductor $q > 1$. Then for $x \ge 2$,*

$$\psi_0(x, \chi) = -\sum_{\rho} \frac{x^\rho}{\rho} - b(\chi) + \tanh^{-1}(x^{-1}),$$

if χ is odd, and

$$\psi_0(x, \chi) = -\sum_{\rho} \frac{x^\rho}{\rho} - \log(x) - b(\chi) - \frac{1}{2}\log(1 - x^{-2}),$$

if χ is even, and where in both expressions, ρ runs over the nontrivial zeros of $L(s, \chi)$ counted with multiplicity and ordered with respect to the size of the ordinate.

As for $\psi(x)$, a few comments are in order. Unlike the explicit formula for $\psi(x)$, there is no main term x in the explicit formula for $\psi(x,\chi)$. This is because $L(s,\chi)$ does not have a pole at $s=1$. The constant $b(\chi)$ can be expressed in terms of $B(\chi)$ (see [Dav80] for a proof). Also, in the case χ is odd the Taylor series of the inverse hyperbolic tangent lets us write

$$\tanh^{-1}(x^{-1}) = \sum_{m\geq 1}\frac{x^{-(2m-1)}}{2m-1} = -\sum_{m\geq 1}\frac{x^{-(2m-1)}}{-(2m-1)} = -\sum_{\omega}\frac{x^{\omega}}{\omega},$$

where ω runs over the trivial zeros of $L(s,\chi)$. In the case χ is even, $\frac{1}{2}\log(1-x^{-2})$ accounts for the contribution of the trivial zeros just as for $\zeta(s)$. We will now prove the explicit formula for $\psi(x,\chi)$:

Proof of the explicit formula for $\psi(x,\chi)$. By truncated Perron's formula applied to $-\frac{L'}{L}(s,\chi)$, we get

$$\psi_0(x,\chi) - J(x,T,\chi) \ll x^c \sum_{\substack{n\geq 1\\ n\neq x}}\frac{\chi(n)\Lambda(n)}{n^c}\min\left(1, \frac{1}{T\left|\log\left(\frac{x}{n}\right)\right|}\right) + \delta_x\chi(x)\Lambda(x)\frac{c}{T}, \tag{9.45}$$

where

$$J(x,T,\chi) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} -\frac{L'}{L}(s,\chi)x^s\,\frac{ds}{s},$$

$c>1$, and it is understood that $\delta_x = 0$ unless x is a prime power. Take $T>2$ not coinciding with the ordinate of a nontrivial zero and let $c = 1+\frac{1}{\log(x^2)}$ so that $x^c = \sqrt{e}x$. We will estimate the right-hand side of Equation (9.45). First, we estimate the terms corresponding to n such that n is bounded away from x. So suppose $n \leq \frac{3}{4}x$ or $n\geq \frac{5}{4}x$. For these n, $\log\left(\frac{x}{n}\right)$ is bounded away from zero so that their contribution is

$$\ll \frac{x^c}{T}\sum_{n\geq 1}\frac{\chi(n)\Lambda(n)}{n^c} \ll \frac{x^c}{T}\left(-\frac{L'}{L}(c,\chi)\right) \ll \frac{x\log(x)}{T}, \tag{9.46}$$

where the last estimate follows from Lemma 7.9.1 (iv) applied to $L(s,\chi)$ while discarding all of the terms in both sums and our choice of c (in particular $\log(c)\ll\log(x)$). Now we estimate the terms n close to x. So consider those n for which $\frac{3}{4}x<n<x$ and let x_1 be the largest prime power less than x. We may also assume $\frac{3}{4}x < x_1 < x$ since otherwise $\Lambda(n)=0$ and these terms do not contribute anything. Moreover, $\frac{x^c}{n^c}\ll 1$. For the term $n=x_1$, we have the estimate

$$\log\left(\frac{x}{n}\right) = -\log\left(1-\frac{x-x_1}{x}\right) \geq \frac{x-x_1}{x},$$

where we have obtained the inequality by using Taylor series of the logarithm truncated after the first term. The contribution of this term is

$$\ll \chi(x_1)\Lambda(x_1)\min\left(1,\frac{x}{T(x-x_1)}\right) \ll \log(x)\min\left(1,\frac{x}{T(x-x_1)}\right). \tag{9.47}$$

For the other n, we write $n = x_1 - v$, where v is an integer satisfying $0<v<\frac{1}{4}x$, so that

$$\log\left(\frac{x}{n}\right) \geq \log\left(\frac{x_1}{n}\right) = -\log\left(1-\frac{v}{x_1}\right) \geq \frac{v}{x_1},$$

where we have obtained the latter inequality by using Taylor series of the logarithm truncated after the first term. The contribution for these n is

$$\ll \sum_{0<v<\frac{1}{4}x}\chi(x_1-v)\Lambda(x_1-v)\frac{x_1}{Tv} \ll \frac{x}{T}\sum_{0<v<\frac{1}{4}x}\frac{\Lambda(x_1-v)}{v} \ll \frac{x\log(x)}{T}\sum_{0<v<\frac{1}{4}x}\frac{1}{v} \ll \frac{x\log^2(x)}{T}. \tag{9.48}$$

The contribution for those n for which $x < n < \frac{5}{4}x$ is handled in exactly the same way with x_1 being the least prime power larger than x. Let $\langle x \rangle$ be the distance between x and the nearest prime power other than x if x itself is a prime power. Combining Equations (9.47) and (9.48) with our previous comment, the contribution for those n with $\frac{3}{4}x < n < \frac{5}{4}x$ is

$$\ll \frac{x\log^2(x)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x \rangle}\right). \tag{9.49}$$

Putting Equations (9.46) and (9.49), the error term in Equation (9.46) is absorbed by the second error term in Equation (9.49) and we obtain

$$\psi_0(x,\chi) - J(x,T,\chi) \ll \frac{x\log^2(x)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x \rangle}\right). \tag{9.50}$$

Now we estimate $J(x,T,\chi)$ by using the residue theorem. Let $U \geq 1$ be an integer with U is even if χ is odd and odd if χ is even. Let Ω be the region enclosed by the contours $\eta_1, \ldots, \eta_4$ in Figure 9.2 and set $\eta = \sum_{1 \leq i \leq 4} \eta_i$ so that $\eta = \partial\Omega$. We may write $J(x,T,\chi)$ as

$$J(x,T,\chi) = \frac{1}{2\pi i}\int_{\eta_1} -\frac{L'}{L}(s,\chi)x^s\,\frac{ds}{s}.$$

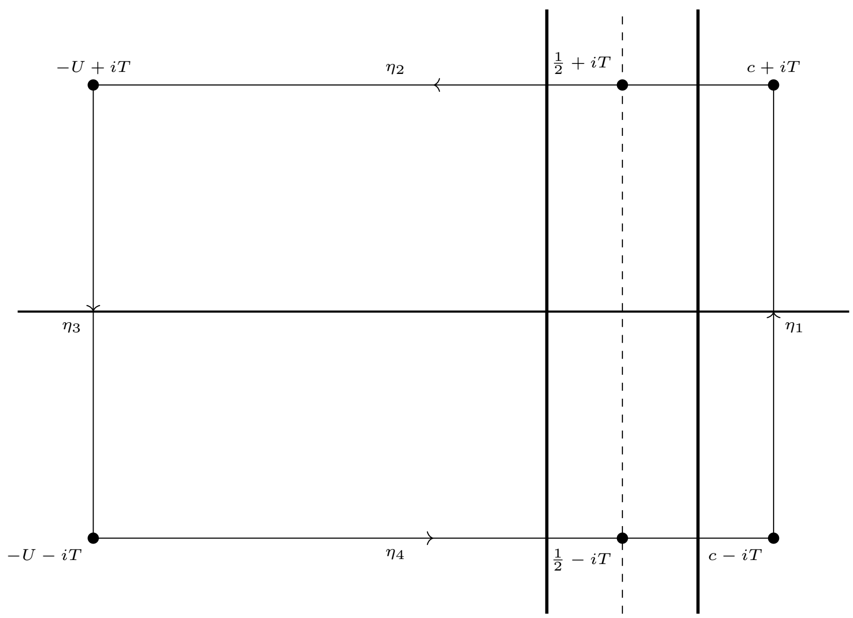

Figure 9.2: Contour for the explicit formula for $\psi(x,\chi)$

We now separate the cases that χ is even or odd. If χ is odd, then the residue theorem, the formula for the negative logarithmic derivative in Proposition 7.8.1 applied to $L(s,\chi)$, and Corollary 1.7.1 together give

$$J(x,T,\chi) = -\sum_{|\gamma|<T}\frac{x^\rho}{\rho} - b(\chi) - \sum_{0<2m+1<U}\frac{x^{-(2m-1)}}{-(2m-1)} + \frac{1}{2\pi i}\int_{\eta_2+\eta_3+\eta_4} -\frac{L'}{L}(s,\chi)x^s\,\frac{ds}{s}, \tag{9.51}$$

where $\rho = \beta + i\gamma$ is a nontrivial zero of $L(s,\chi)$. If χ is even, then there is a minor complication because $L(s,\chi)$ has a simple zero at $s = 0$ and so the integrand has a double pole at $s = 0$. To find the residue, the Laurent series are

$$\frac{L'}{L}(s,\chi) = \frac{1}{s} + b(\chi) + \cdots \quad \text{and} \quad \frac{x^s}{s} = \frac{1}{s} + \log(x) + \cdots,$$

and thus the residue of the integrand is $-(\log(x)+b(\chi))$. Now as before, the residue theorem, the formula for the negative logarithmic derivative in Proposition 7.8.1 applied to $L(s,\chi)$, and Corollary 1.7.1 together give

$$J(x,T,\chi) = -\sum_{|\gamma|<T} \frac{x^{\rho}}{\rho} - \log(x) - b(\chi) - \sum_{0<2m<U} \frac{x^{-2m}}{-2m} + \frac{1}{2\pi i}\int_{\eta_2+\eta_3+\eta_4} -\frac{L'}{L}(s,\chi)x^{s}\,\frac{ds}{s}, \tag{9.52}$$

where $\rho = \beta+i\gamma$ is a nontrivial zero of $L(s,\chi)$. We now estimate the remaining integrals in Equations (9.51) and (9.52). For this estimate, the parity of χ does not matter so we make no such restriction. By Lemma 7.9.1 (ii) applied to $L(s,\chi)$, the number of nontrivial zeros satisfying $|\gamma - T| < 1$ is $O(\log(qT))$. Among the ordinates of these nontrivial zeros, there must be a gap of size $\gg \frac{1}{\log(qT)}$. Upon varying T by a bounded amount (we are varying in the interval $[T-1, T+1]$) so that it belongs to this gap, we can additionally ensure

$$\gamma - T \gg \frac{1}{\log(qT)},$$

for all the nontrivial zeros of $L(s,\chi)$. To estimate part of the horizontal integrals over η_2 and η_4, Lemma 7.9.1 (iv) applied to $L(s,\chi)$ gives

$$\frac{L'}{L}(s,\chi) = \sum_{|\gamma-T|<1} \frac{1}{s-\rho} + O(\log(qT)),$$

on the parts of these segments with $-1 \le \sigma \le 2$. Our choice of T implies $|s-\rho| \ge |\gamma - T| \gg \frac{1}{\log(qT)}$ so that each term in the sum is $O(\log(qT))$. As there are at most $O(\log(qT))$ such terms by Lemma 7.9.1 (ii) applied to $L(s,\chi)$, we have

$$\frac{L'}{L}(s,\chi) = O(\log^2(qT)),$$

on the parts of these segments with $-1 \le \sigma \le 2$. It follows that the parts of the horizontal integrals over η_2 and η_4 with $-1 \le \sigma \le c$ (recall $c < 2$) contribute

$$\ll \frac{\log^2(qT)}{T}\int_{-1}^{c} x^{\sigma}\,d\sigma \ll \frac{\log^2(qT)}{T}\int_{-\infty}^{c} x^{\sigma}\,d\sigma \ll \frac{x\log^2(qT)}{T\log(x)}. \tag{9.53}$$

where in the last estimate we have used the choice of c. To estimate the remainder of the horizontal integrals, we require a bound for $\frac{L'}{L}(s,\chi)$ when $\sigma < -1$ and away from the trivial zeros. To find such a bound, write the functional equation for $L(s,\chi)$ in the form

$$L(s,\chi) = \frac{\varepsilon_{\chi}}{i^{\mathfrak{a}}} q^{\frac{1}{2}-s}\pi^{s-1}\frac{\Gamma\left(\frac{(1-s)+\mathfrak{a}}{2}\right)}{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)} L(1-s,\chi),$$

and take the logarithmic derivative to get

$$\frac{L'}{L}(s,\chi) = -\log(q) + \log(\pi) + \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{(1-s)+\mathfrak{a}}{2}\right) - \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) + \frac{L'}{L}(1-s,\chi).$$

Now let s be such that $\sigma < -1$ and suppose s is distance $\frac{1}{2}$ away from the trivial zeros. We will estimate every term on the right-hand side of the identity above. The second term is constant and the last term is bounded since it is an absolutely convergent Dirichlet series. For the digamma terms,

s is away from the trivial zeros so Proposition 1.7.3 implies $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{(1-s)+\mathfrak{a}}{2}\right) = O(\log|(1-s)+\mathfrak{a}|)$ and $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) = O(\log|s+\mathfrak{a}|)$. However, as $\sigma < -1$ and s is away from the trivial zeros, $s+\mathfrak{a}$ and $(1-s)+\mathfrak{a}$ are bounded away from zero so that $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{(1-s)+\mathfrak{a}}{2}\right) = O(\log|s|)$ and $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) = O(\log|s|)$. Putting these estimates together with the first term yields

$$\frac{L'}{L}(s,\chi) \ll \log(q|s|), \tag{9.54}$$

for $\sigma < -1$. Using Equation (9.54), the parts of the horizontal integrals over η_2 and η_4 with $-U \leq \sigma \leq -1$ contribute

$$\ll \frac{\log(qT)}{T}\int_{-U}^{-1} x^{\sigma}\,d\sigma \ll \frac{\log(qT)}{Tx\log(x)}. \tag{9.55}$$

Combining Equations (9.53) and (9.55) gives

$$\frac{1}{2\pi i}\int_{\eta_2+\eta_4} -\frac{L'}{L}(s)x^s\frac{ds}{s} \ll \frac{x\log^2(qT)}{T\log(x)} + \frac{\log(qT)}{Tx\log(x)} \ll \frac{x\log^2(qT)}{T\log(x)}. \tag{9.56}$$

To estimate the vertical integral, we use Equation (9.54) again to conclude that

$$\frac{1}{2\pi i}\int_{\eta_3} -\frac{L'}{L}(s,\chi)x^s\frac{ds}{s} \ll \frac{\log(qU)}{U}\int_{-T}^{T} x^{-U}\,dt \ll \frac{T\log(qU)}{Ux^U}. \tag{9.57}$$

Combining Equations (9.51), (9.56) and (9.57) and taking the limit as $U \to \infty$, the error term in Equation (9.57) vanishes and the sum over m in Equations (9.51) and (9.52) evaluates to $-\tanh^{-1}(x^{-1})$ or $\frac{1}{2}\log(1-x^{-2})$ respectively (as we have already mentioned) giving

$$J(x,T,\chi) = -\sum_{|\gamma|<T}\frac{x^{\rho}}{\rho} - b(\chi) + \tanh^{-1}(x^{-1}) + \frac{x\log^2(qT)}{T\log(x)}, \tag{9.58}$$

if χ is odd, and

$$J(x,T,\chi) = -\sum_{|\gamma|<T}\frac{x^{\rho}}{\rho} - \log(x) - b(\chi) - \frac{1}{2}\log(1-x^{-2}) + \frac{x\log^2(qT)}{T\log(x)}, \tag{9.59}$$

if χ is even. Substituting Equations (9.58) and (9.59) into Equation (9.50) in the respective cases, we obtain

$$\psi_0(x,\chi) = -\sum_{|\gamma|<T}\frac{x^{\rho}}{\rho} - b(\chi) + \tanh^{-1}(x^{-1}) + \frac{x\log^2(xqT)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right), \tag{9.60}$$

if χ is odd, and

$$\psi_0(x,\chi) = -\sum_{|\gamma|<T}\frac{x^{\rho}}{\rho} - \log(x) - b(\chi) - \frac{1}{2}\log(1-x^{-2}) + \frac{x\log^2(xqT)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right), \tag{9.61}$$

if χ is even, and where the second to last term on the right-hand side in both equations are obtained by combining the error term in Equation (9.56) with the first error term in Equation (9.50). The theorem follows by taking the limit as $T \to \infty$. $\square$

As was the case for $\psi(x)$, the convergence of the right-hand side in the explicit formula for $\psi(x,\chi)$ is uniform in any interval not containing a prime power since $\psi(x,\chi)$ is continuous there. Moreover, there is an approximate formula for $\psi(x,\chi)$ as a corollary which holds for all Dirichlet characters:

Corollary 9.4.1. *Let χ be a Dirichlet character modulo $m > 1$. Then for $2 \le T \le x$,*

$$\psi_0(x,\chi) = -\frac{x^{\beta_\chi}}{\beta_\chi} - \sum_{|\gamma|<T}{}' \frac{x^\rho}{\rho} + R(x,T,\chi),$$

where ρ runs over the nontrivial zeros of $L(s,\chi)$ counted with multiplicity and ordered with respect to the size of the ordinate, the $'$ in the sum indicates that we are excluding the terms corresponding to real zeros, the term corresponding to a Siegel zero β_χ is present only if $L(s,\chi)$ admits a Siegel zero, and

$$R(x,T,\chi) \ll \frac{x\log^2(xmT)}{T} + x^{1-\beta_\chi}\log(x) + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right),$$

where $\langle x\rangle$ is the distance between x and the nearest prime power other than x if x itself is a prime power and again the term corresponding to a Siegel zero β_χ is present only if $L(s,\chi)$ admits a Siegel zero. Moreover, if x is an integer, we have the simplified estimate

$$R(x,T,\chi) \ll \frac{x\log^2(xmT)}{T} + x^{1-\beta_\chi}\log(x).$$

Proof. We first reduced to the case that χ is primitive. Let $\widetilde{\chi}$ be the primitive character inducing χ and denote its conductor by q. We estimate

$$\begin{aligned} |\psi_0(x,\chi) - \psi_0(x,\widetilde{\chi})| &\le \sum_{\substack{n\le x\\(n,m)>1}} \Lambda(n) \\ &= \sum_{p|m}\sum_{\substack{v\ge1\\p^v\le x}} \log(p) \\ &\ll \log(x)\sum_{p|m}\log(p) \\ &\ll \log(x)\log(m) \\ &\ll \log^2(xm), \end{aligned} \tag{9.62}$$

where the third line holds because $p^v \le x$ implies $v \le \frac{\log(x)}{\log(p)}$ so that there are $O(\log(x))$ many terms in the inner sum and in the last line we have used the simple estimates $\log(x) \ll \log(xm)$ and $\log(m) \ll \log(xm)$. Therefore the difference between $\psi_0(x,\chi)$ and $\psi_0(x,\widetilde{\chi})$ is $O(\log^2(xm))$. Now for $2 \le T \le x$, we have $\log^2(xm) \ll \frac{x\log^2(xmT)}{T}$, which implies that the difference is absorbed into $O\left(\frac{x\log^2(xmT)}{T}\right)$ which is the first term in the error for $R(x,T,\chi)$. As $R(x,T,\widetilde{\chi}) \ll R(x,T,\chi)$ because $q \le m$, and there are finitely many nontrivial zeros of $L(s,\chi)$ that are not nontrivial zeros of $L(s,\widetilde{\chi})$ (all occurring on the line $\sigma = 0$), it suffices to assume that χ is primitive. The claim will follow from estimating terms in Equations (9.60) and (9.61). We will estimate the constant $b(\chi)$ first. The formula for the negative logarithmic derivative in Proposition 7.8.1 applied to $L(s,\chi)$ at $s = 2$ implies

$$0 = -\frac{L'}{L}(2,\chi) - \frac{1}{2}\log(q) + \frac{1}{2}\log(\pi) - \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{2+\mathfrak{a}}{2}\right) + B(\chi) + \sum_\rho\left(\frac{1}{2-\rho} + \frac{1}{\rho}\right). \tag{9.63}$$

Adding Equation (9.63) to the formula for the negative logarithmic derivative in Proposition 7.8.1 applied to $L(s,\chi)$ gives

$$-\frac{L'}{L}(s,\chi) = -\frac{L'}{L}(2,\chi) - \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{2+\mathfrak{a}}{2}\right) + \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) - \sum_{\rho}\left(\frac{1}{s-\rho} - \frac{1}{2-\rho}\right).$$

As the first two terms are constant, we obtain a weaker estimate

$$-\frac{L'}{L}(s,\chi) = \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) - \sum_{\rho}\left(\frac{1}{s-\rho} - \frac{1}{2-\rho}\right) + O(1).$$

If χ is odd, we set $s = 0$ since $\frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right)$ does not have a pole there. If χ is even, we compare constant terms in the Laurent series using the Laurent series

$$\frac{L'}{L}(s,\chi) = \frac{1}{s} + b(\chi) + \cdots \quad \text{and} \quad \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s+\mathfrak{a}}{2}\right) = \frac{1}{s} + b + \cdots,$$

for some constant b. In either case, our previous estimate gives

$$b(\chi) = -\sum_{\rho}\left(\frac{1}{\rho} + \frac{1}{2-\rho}\right) + O(1). \tag{9.64}$$

Let $\rho = \beta + i\gamma$. For all the terms with $|\gamma| > 1$, we estimate

$$\sum_{|\gamma|>1}\left(\frac{1}{\rho} + \frac{1}{2-\rho}\right) \le \sum_{|\gamma|>1}\left|\frac{1}{\rho} + \frac{1}{2-\rho}\right| = \sum_{|\gamma|>1}\frac{2}{|\rho(2-\rho)|} \ll \sum_{|\gamma|>1}\frac{1}{|\rho|^2} \ll \log(q), \tag{9.65}$$

where the second to last estimate holds since $2-\rho \gg \rho$ because β is bounded and the last estimate holds by the convergent sum in Proposition 7.8.1and Lemma 7.9.1 (ii) both applied to $L(s,\chi)$ (recall that the tail of a convergent series is bounded). For the terms corresponding to $2-\rho$ with $|\gamma| \le 1$, we have

$$\sum_{|\gamma|\le 1}\frac{1}{2-\rho} \le \sum_{|\gamma|\le 1}\frac{1}{|2-\rho|} \ll \log(q), \tag{9.66}$$

where the last estimate holds by using Lemma 7.9.1 (ii) applied to $L(s,\chi)$ and because the nontrivial zeros are bounded away from 2. Combining Equations (9.64) to (9.66) yields

$$b(\chi) = -\sum_{|\gamma|\le 1}\frac{1}{\rho} + O(\log(q)). \tag{9.67}$$

Inserting Equation (9.67) into Equations (9.60) and (9.61) and noting that $\tanh^{-1}(x^{-1})$ and $\frac{1}{2}\log(1-x^{-2})$ are both bounded for $x \ge 2$ gives

$$\psi_0(x,\chi) = -\sum_{|\gamma|<T}\frac{x^{\rho}}{\rho} + \sum_{|\gamma|\le 1}\frac{1}{\rho} + R'(x,T,\chi), \tag{9.68}$$

where

$$R'(x,T,\chi) \ll \frac{x\log^2(xqT)}{T} + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right),$$

and we have absorbed the error in Equation (9.67) into $O\left(\frac{x\log^2(xqT)}{T}\right)$ because $2 \le T \le x$. Extracting the terms corresponding to the possible real zeros β_χ and $1-\beta_\chi$ in Equation (9.68), we obtain

$$\psi_0(x,\chi) = -\sum_{|\gamma|<T}{}' \frac{x^\rho}{\rho} + \sum_{|\gamma|\le 1}{}' \frac{1}{\rho} - \frac{x^{\beta_\chi}-1}{\beta_\chi} - \frac{x^{1-\beta_\chi}-1}{1-\beta_\chi} + R'(x,T,\chi). \tag{9.69}$$

We now estimate some of the terms in Equation (9.69). For the second sum, we have $\rho \gg \frac{1}{\log(q)}$ since γ is bounded and $\beta < 1 - \frac{c}{\log(q|\gamma|)}$ for some positive constant c by Theorem 9.2.2. Thus

$$\sum_{|\gamma|\le 1}{}' \frac{1}{\rho} \ll \sum_{|\gamma|\le 1}{}' \log(q) \ll \log^2(q), \tag{9.70}$$

where the last estimate holds by Lemma 7.9.1 (ii) applied to $L(s,\chi)$. Similarly,

$$\frac{x^{1-\beta_\chi}-1}{1-\beta_\chi} \ll x^{1-\beta_\chi}\log(x), \tag{9.71}$$

because $\rho \gg \frac{1}{\log(q)}$ implies $1-\beta_\chi \gg \frac{1}{\log(q)} \gg \frac{1}{\log(x)}$. Substituting Equations (9.70) and (9.71) into Equation (9.69) and noting that β_χ is bounded yields

$$\psi_0(x,\chi) = -\frac{x^{\beta_\chi}}{\beta_\chi} - \sum_{|\gamma|<T}{}' \frac{x^\rho}{\rho} + R(x,T,\chi), \tag{9.72}$$

where

$$R(x,T,\chi) \ll \frac{x\log^2(xqT)}{T} + x^{1-\beta_\chi}\log(x) + \log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right)$$

and we have absorbed the error in Equation (9.70) into $O\left(\frac{x\log^2(xqT)}{T}\right)$ because because $2 \le T \le x$. If x is an integer, then $\langle x\rangle \ge 1$ so that $\log(x)\min\left(1, \frac{x}{T\langle x\rangle}\right) \le \frac{x\log(x)}{T}$ and this term can be absorbed into $O\left(\frac{x\log^2(xqT)}{T}\right)$. $\square$

We can now discuss the Siegel–Walfisz theorem. Let a and m be positive integers with $m > 1$ and $(a,m) = 1$. The **prime counting function** $\pi(x : a, m)$ is defined by

$$\pi(x;a,m) = \sum_{\substack{p \le x \\ p \equiv a \pmod m}} 1,$$

for $x > 0$. Equivalently, $\pi(x;a,m)$ counts the number of primes that no larger than x and are equivalent to a modulo m. This is the analog of $\pi(x)$ that is naturally associated to Dirichlet characters modulo m. Accordingly, there are asymptotics for $\pi(x;a,m)$ analogous to those for $\pi(x)$. To prove them, we will require an auxiliary function. The function $\psi(x;a,m)$ is defined by

$$\psi(x;a,m) = \sum_{\substack{n \le x \\ n \equiv a \pmod m}} \Lambda(n),$$

for $x \ge 1$. This is just $\psi(x)$ restricted to only those terms equivalent to a modulo m. As for the asymptotic, the classical **Siegel–Walfisz theorem** is the first of them and the precise statement is the following:

Theorem 9.4.2 (Siegel–Walfisz theorem, classical version). *Let a and m be positive integers with $m > 1$, $(a, m) = 1$, and let $N \geq 1$. For $x \geq 2$,*

$$\pi(x; a, m) \sim \frac{x}{\varphi(m)\log(x)},$$

provided $m \leq \log^N(x)$.

The logarithmic integral version of the **Siegel–Walfisz theorem** is equivalent and sometimes more useful:

Theorem 9.4.3 (Siegel–Walfisz theorem, logarithmic integral version). *Let a and m be positive integers with $m > 1$, $(a, m) = 1$, and let $N \geq 1$. For $x \geq 2$,*

$$\pi(x; a, m) \sim \frac{\mathrm{Li}(x)}{\varphi(m)},$$

provided $m \leq \log^N(x)$.

We will prove the absolute error version of the **Siegel–Walfisz theorem** which is slightly stronger as it bounds the absolute error between $\pi(x; a, m)$ and $\frac{\mathrm{Li}(x)}{\varphi(m)}$:

Theorem 9.4.4 (Siegel–Walfisz theorem, absolute error version). *Let a and m be positive integers with $m > 1$, $(a, m) = 1$, and let $N \geq 1$. For $x \geq 2$, there exists a positive constant c such that*

$$\pi(x; a, m) = \frac{\mathrm{Li}(x)}{\varphi(m)} + O\left(xe^{-c\sqrt{\log(x)}}\right),$$

provided $m \leq \log^N(x)$.

Proof. It suffices to assume x is an integer, because $\pi(x; a, m)$ can only change value at integers and the other functions in the statement are increasing. We being with the identity

$$\psi(x; a, m) = \frac{1}{\varphi(m)} \sum_{\chi \pmod{m}} \overline{\chi}(a)\psi(x, \chi), \tag{9.73}$$

which holds by the orthogonality relations (Proposition 1.3.1 (ii)). Let $\widetilde{\chi}$ be the primitive character inducing χ. Then Equation (9.62) implies

$$\psi(x, \chi) = \psi(x, \widetilde{\chi}) + O(\log^2(xm)). \tag{9.74}$$

When $\chi = \chi_{m,0}$ we have $\psi(x, \widetilde{\chi}) = \psi(x)$, and as $\psi(x, \chi) \sim \psi_0(x, \chi)$, substituting Equation (9.35) into Equation (9.74) gives

$$\psi(x, \chi_{m,0}) = \psi(x) + O\left(xe^{-c\sqrt{\log(x)}} + \log^2(xm)\right), \tag{9.75}$$

for some positive constant c. Upon combining Equations (9.73) and (9.75), we obtain

$$\psi(x; a, m) = \frac{x}{\varphi(m)} + \frac{1}{\varphi(m)} \sum_{\substack{\chi \pmod{m} \\ \chi \neq \chi_{m,0}}} \overline{\chi}(a)\psi(x, \chi) + O\left(\frac{1}{\varphi(m)}\left(xe^{-c\sqrt{\log(x)}} + \log^2(xm)\right)\right). \tag{9.76}$$

We now prove

$$\psi(x, \chi) = -\frac{x^{\beta_\chi}}{\beta_\chi} + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.77}$$

for some potentially different constant c, where $\chi \neq \chi_{m,0}$, and the term corresponding to β_χ appears if and only if $L(s,\chi)$ admits a Siegel zero. To accomplish this, we estimate the sum over the nontrivial zeros of $L(s,\chi)$ in Corollary 9.4.1. So fix a non-principal χ modulo m and let $\widetilde{\chi}$ be the primitive character inducing χ. Let $2 \le T \le x$ not coinciding with the ordinate of a nontrivial zero and let $\rho = \beta + i\gamma$ be a complex nontrivial zero of $L(s,\chi)$ with $|\gamma| < T$. By Theorem 9.2.2, all of the zeros ρ satisfy $\beta < 1 - \frac{c}{\log(mT)}$ for some possibly smaller c (recall that the nontrivial zeros of $L(s,\chi)$ that are not nontrivial zeros of $L(s,\widetilde{\chi})$ lie on the line $\sigma = 0$). It follows that

$$|x^\rho| = x^\beta < x^{1-\frac{c}{\log(mT)}} = xe^{-c\frac{\log(x)}{\log(mT)}}. \tag{9.78}$$

As $|\rho| > |\gamma|$, for those terms with $|\gamma| > 1$ (unlike the Riemann zeta function we do not have a positive lower bound for the first ordinate γ_1 of a nontrivial zero that is not real since Siegel zeros may exist), applying integration by parts gives

$$\sum_{1<|\gamma|<T} \frac{1}{\rho} \ll \sum_{1<|\gamma|<T} \frac{1}{\gamma} \ll \int_1^T \frac{dN(t,\widetilde{\chi})}{t} = \frac{N(T,\widetilde{\chi})}{T} + \int_1^T \frac{N(t,\widetilde{\chi})}{t^2}\,dt \ll \log^2(mT) \ll \log^2(xm), \tag{9.79}$$

where in the second to last estimate we have used that $N(t,\widetilde{\chi}) \ll t\log(qt) \ll t\log(mt)$ which follows from Corollary 7.9.1 applied to $L(s,\widetilde{\chi})$ and that $q \le m$ and in the last estimate we have used the bound $T \le x$. For the remaining terms with $|\gamma| \le 1$ (that do not correspond to real nontrivial zeros), Equation (9.70) along with $q \le m$ gives

$$\sideset{}{'}\sum_{|\gamma|\le 1} \frac{1}{\rho} \ll \log^2(m). \tag{9.80}$$

Combining Equations (9.77) to (9.80) yields

$$\sideset{}{'}\sum_{|\gamma|<T} \frac{x^\rho}{\rho} \ll x\log^2(xm)e^{-c\frac{\log(x)}{\log(mT)}}, \tag{9.81}$$

where the $'$ in the sum indicates that we are excluding the terms corresponding to real zeros and the error in Equation (9.80) has been absorbed by that in Equation (9.79). As $\psi(x,\chi) \sim \psi_0(x,\chi)$ and x is an integer, inserting Equation (9.81) into Corollary 9.4.1 results in

$$\psi(x,\chi) + \frac{x^{\beta_\chi}}{\beta_\chi} \ll x\log^2(xm)e^{-c\frac{\log(x)}{\log(mT)}} + \frac{x\log^2(xmT)}{T} + x^{1-\beta_\chi}\log(x), \tag{9.82}$$

where the terms corresponding to real zeros are present if and only if $L(s,\chi)$ admits a Siegel zero. We now let T be determined by $T = x$ for $2 \le x < 3$ and

$$\log^2(T) = \log(x),$$

or equivalently,

$$T = e^{\sqrt{\log(x)}},$$

for $x \ge 3$. With this choice of T (note that if $x \ge 2$ then $2 \le T \le x$) and that $m \ll \log^N(x)$, we can estimate Equation (9.82) as follows:

$$
\begin{aligned}
\psi(x) + \frac{x^{\beta_\chi}}{x} &\ll x(\log^2(x) + \log^2(m))e^{-c\frac{\log(x)}{\log(m)+\sqrt{\log(x)}}} + x(\log^2(x) + \log^2(m) + \log(x))e^{-\sqrt{\log(x)}} + x^{1-\beta_\chi}\log(x) \\
&\ll x(\log^2(x) + \log^2(m))e^{-c\frac{\log(x)}{\log(m)+\sqrt{\log(x)}}} + x(\log^2(x) + \log^2(m) + \log(x))e^{-\sqrt{\log(x)}} + x^{1-\beta_\chi}\log(x) \\
&\ll x(\log^2(x) + \log^2(m))e^{-c\sqrt{\log(x)}} + x(\log^2(x) + \log^2(m))e^{-\sqrt{\log(x)}} + x^{1-\beta_\chi}\log(x) \\
&\ll x\log^2(x)e^{-c\sqrt{\log(x)}} + x\log^2(x)e^{-\sqrt{\log(x)}} + x^{1-\beta_\chi}\log(x) \\
&\ll x\log^2(x)e^{-\min(1,c)\frac{\sqrt{\log(x)}}{\log(m)}},
\end{aligned}
$$

where in the last estimate we have used that $x^{1-\beta_\chi} \le x^{\frac{1}{2}}$ because $\beta_\chi \ge \frac{1}{2}$. As $\log(x) \ll_\varepsilon e^{-\varepsilon\sqrt{\log(x)}}$, we conclude that

$$
\psi(x) + \frac{x^{\beta_\chi}}{x} \ll xe^{-c\sqrt{\log(x)}},
$$

for some smaller c with $c < 1$. This is equivalent to Equation (9.77). Substituting Equation (9.77) into Equation (9.76) and noting that there is at most one Siegel zero for characters modulo m by Proposition 9.2.2, we arrive at

$$
\psi(x;a,m) = \frac{x}{\varphi(m)} - \frac{\overline{\chi}_1(a)x^{\beta_\chi}}{\varphi(m)\beta_\chi} + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.83}
$$

where χ_1 is the single quadratic character modulo m such that $L(s,\chi_1)$ admits a Siegel zero if it exists and we have absorbed the second term in the error in Equation (9.76) into the first since $\log(x) \ll_\varepsilon e^{-\varepsilon\sqrt{\log(x)}}$ and $m \ll \log^N(x)$. Taking $\varepsilon = \frac{1}{2N}$ in the zero-free region version of Siegel's theorem, $m^{\frac{1}{2N}} \ll \sqrt{\log(x)}$ so that $\beta_\chi < 1 - \frac{c}{\sqrt{\log(x)}}$ for some potentially smaller constant c. It follows that m^{2N} Therefore

$$
x^{\beta_\chi} < x^{1-\frac{c}{\sqrt{\log(x)}}} = xe^{-c\log(x)}. \tag{9.84}
$$

Combining Equations (9.83) and (9.84) gives the simplified estimate

$$
\psi(x;a,m) = \frac{x}{\varphi(m)} + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.85}
$$

for some potentially smaller constant c. Now let

$$
\pi_1(x;a,m) = \sum_{\substack{n \le x \\ n \equiv a \pmod{m}}} \frac{\Lambda(n)}{\log(n)}.
$$

We can write $\pi_1(x;a,m)$ in terms of $\psi(x;a,m)$ as follows:

$$\begin{aligned}\pi_1(x;a,m) &= \sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \frac{\Lambda(n)}{\log(n)}\\ &= \sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \Lambda(n)\int_n^x \frac{dt}{t\log^2(t)} + \frac{1}{\log(x)}\sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \Lambda(n)\\ &= \int_2^x \sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \Lambda(n)\frac{dt}{t\log^2(t)} + \frac{1}{\log(x)}\sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \Lambda(n)\\ &= \int_2^x \frac{\psi(t;a,m)}{t\log^2(t)}\,dt + \frac{\psi(x;a,m)}{\log(x)}.\end{aligned}$$

Applying Equation (9.85) to the last expression yields

$$\pi_1(x;a,m) = \int_2^x \frac{t}{\varphi(m)t\log^2(t)}\,dt + \frac{x}{\varphi(m)\log(x)} + O\left(\int_2^x \frac{e^{-c\sqrt{\log(t)}}}{\log^2(t)}\,dt + \frac{xe^{-c\sqrt{\log(x)}}}{\log(x)}\right). \tag{9.86}$$

Applying integrating by parts to the main term in Equation (9.86), we obtain

$$\int_2^x \frac{t}{\varphi(m)t\log^2(t)}\,dt + \frac{x}{\varphi(m)\log(x)} = \int_2^x \frac{dt}{\varphi(m)\log(t)} + \frac{2}{\varphi(m)\log(2)} = \frac{\mathrm{Li}(x)}{\varphi(m)} + \frac{2}{\varphi(m)\log(2)}. \tag{9.87}$$

As for the error term in Equation (9.86), we use Equation (9.42). Combining Equations (9.42), (9.86) and (9.87) yields

$$\pi_1(x;a,m) = \frac{\mathrm{Li}(x)}{\varphi(m)} + O\left(xe^{-c\sqrt{\log(x)}}\right), \tag{9.88}$$

for some smaller c and where the constant in Equation (9.87) has been absorbed into the error term. As last we pass from $\pi_1(x;a,m)$ to $\pi(x;a,m)$. If p is a prime such that $p^m < x$ for some $m \geq 1$, then $p < x^{\frac{1}{2}} < x^{\frac{1}{3}} < \cdots < x^{\frac{1}{m}}$. Therefore

$$\pi_1(x;a,m) = \sum_{\substack{n\leq x\\ n\equiv a \pmod m}} \frac{\Lambda(n)}{\log(n)} = \sum_{\substack{p^m\leq x\\ p^m\equiv a \pmod m}} \frac{\log(p)}{m\log(p)} = \pi(x;a,m) + \frac{1}{2}\pi(x^{\frac{1}{2}};a,m) + \cdots. \tag{9.89}$$

Moreover, as $\pi(x^{\frac{1}{n}};a,m) < x^{\frac{1}{n}}$ for any $n \geq 1$, we see that $\pi(x;a,m) - \pi_1(x;a,m) = O(x^{\frac{1}{2}})$. This estimate together with Equations (9.88) and (9.89) gives

$$\pi(x) = \frac{\mathrm{Li}(x)}{\varphi(m)} + O\left(xe^{-c\sqrt{\log(x)}}\right),$$

because $x^{\frac{1}{2}} \ll xe^{-c\sqrt{\log(x)}}$. This completes the proof. $\square$

It is interesting to note that the constant c in the Siegel-Walfisz theorem is ineffective because of the use of Siegel's theorem (it also depends upon N). This is unlike the prime number theorem, where the constant c can be made to be effective. The proof of the logarithmic integral and classical versions of the Siegel-Walfisz theorem are immediate:

Proof of Siegel-Walfisz theorem, logarithmic integral and classical versions. By the absolute error version of the Siegel-Walfisz theorem,

$$\pi(x; a, m) = \frac{\mathrm{Li}(x)}{\varphi(m)}\left(1 + O\left(\frac{\varphi(m)xe^{-c\sqrt{\log(x)}}}{\mathrm{Li}(x)}\right)\right).$$

As $\mathrm{Li}(x) \sim \frac{x}{\log(x)}$, we have

$$\frac{\varphi(m)xe^{-c\sqrt{\log(x)}}}{\mathrm{Li}(x)} \sim \varphi(m)\log(x)e^{-c\sqrt{\log(x)}} = o(1),$$

where the equality holds since $m \ll \log^N(x)$ and $\log(x) \ll_\varepsilon e^{-\varepsilon\sqrt{\log(x)}}$. The logarithm integral version follows. The classical version also holds using the asymptotic $\mathrm{Li}(x) \sim \frac{x}{\log(x)}$. $\square$

Also, we have an optimal error term, in a much wider range of m, assuming the Riemann hypothesis for Dirichlet L-functions:

Proposition 9.4.1. *Let a and m be positive integers with $m > 1$ and $(a, m) = 1$. For $x \geq 2$, we have*

$$\pi(x; a, m) = \frac{\mathrm{Li}(x)}{\varphi(m)} + O(\sqrt{x}\log(x)),$$

provided $m \leq x$ and the Riemann hypothesis for Dirichlet L-functions holds.

Proof. Let χ be a Dirichlet character modulo m. If ρ is a nontrivial zero of $L(s, \chi)$, the Riemann hypothesis for Dirichlet L-functions implies $|x^\rho| = \sqrt{x}$ and that Siegel zeros do not exist so we may merely assume $m \leq x$. Therefore as in the proof of the absolute error version of the Siegel-Walfisz theorem,

$$\sum_{|\gamma|<T} \frac{x^\rho}{\rho} \ll \sqrt{x}\log^2(x),$$

for $2 \leq T \leq x$ not coinciding with the ordinate of a nontrivial zero. Repeating the same argument with T determined by $T = x$ for $2 \leq x < 3$ and

$$T^2 = x,$$

for $x \geq 3$ gives

$$\psi(x; a, m) = \frac{x}{\varphi(m)} + O(\sqrt{x}\log^2(x)),$$

and then transferring to $\pi_1(x)$ and finally $\pi(x)$ gives

$$\pi(x; a, m) = \frac{x}{\varphi(m)} + O(\sqrt{x}\log(x)).$$

$\square$

Part V

An Introduction to Moments

Chapter 10

The Katz-Sarnak Philosophy

The Katz-Sarnak philosophy is an idea that certain statistics about families of L-functions should match statistics for random matrices coming from some particular compact matrix group. One starts with some class of zeros to look at, say zeros of an individual L-function high up the critical strip or zeros of for some collection of L-functions low down on the critical strip. Actually, one works with the corresponding unfolded nontrivial zeros since they are evenly spaced on average. Then some class of test functions are introduced to carry out the statistical calculations in order to reveal the similarity with some class of matrices. In the following, we give a loose introduction to the Katz-Sarnak philosophy.

10.1 The Work of Montgomery & Dyson

The beginning of the connection between random matrix theory and analytic number theory was at Princeton in the 1970s via discussions between Montgomery and Dyson. They found similarities between statistical information about the nontrivial distribution of the zeros of the Riemann zeta function and calculations in random matrix theory about unitary matrices. To do this, they considered the unfolded nontrivial zeros $\rho_{\text{unf}} = \beta + i\omega$ of $\zeta(s)$ with positive ordinate, that is $\omega > 0$, and indexed them according to the size of ordinate. So let $\Omega = (\omega_n)_{n\ge 1}$ denote the increasing sequence of positive ordinates of the unfolded nontrivial zeros of $\zeta(s)$. Montgomery and Dyson considered the **two-point correlation function** $F(\alpha, \beta; \zeta, W)$ for $\zeta(s)$, defined by

$$F(\alpha, \beta; \zeta, W) = \frac{1}{W}|\{(\omega_n, \omega_m) \in \Omega^2 : \omega_n, \omega_m \le W \text{ and } \omega_n - \omega_m \in [\alpha, \beta]\}|,$$

for any real α and β with $\alpha < \beta$ and $W > 0$. What this function measures is the probability of how close pairs of zeros tend to be with respect to some fixed distance and up to some fixed height. In other words, the correlation between distances of zeros. They wanted to understand if the limiting distribution

$$F(\alpha, \beta; \zeta) = \lim_{W\to\infty} F(\alpha, \beta; \zeta, W),$$

exists and what can be said about it. The following conjecture made by Montgomery, known as Montgomery's pair correlation conjecture, answers this:

Conjecture 10.1.1 (Montgomery's pair correlation conjecture). *For any α and β with $\alpha < \beta$, $F(\alpha, \beta; \zeta)$ exists provided the Riemann hypothesis for the Riemann zeta function holds. Moreover,*

$$F(\alpha, \beta; \zeta) = \int_\alpha^\beta \left(1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2 + \delta(x)\right) dx,$$

where $\delta(x)$ is the Dirac delta function.

Montgomery's pair correlation conjecture still remains out of reach, but there is very good numerical evidence supporting it from some unpublished work of Odlyzko (see [Odl92]). Dyson recognized that Montgomery's pair correlation conjecture models a similar situation in random matrix theory that he had investigated earlier. Consider an $N \times N$ unitary matrix $A \in U(N)$ with eigenphases θ_n for $1 \leq n \leq N$ denoted in increasing order. Clearly the average density of the eigenphases of A in $[0, 2\pi)$ is $\frac{N}{2\pi}$. For any eigenphase θ, let ϕ be the **unfolded eigenphase** corresponding to θ be defined by

$$\phi = \frac{N}{2\pi}\theta.$$

It follows that the average density of the unfolded eigenphases of A in $[0, N)$ is 1. Let $\Phi = (\phi_n)_{1 \leq n \leq N}$ denote the increasing sequence of unfolded eigenphases of A. We consider the **two-point correlation function** $F(\alpha, \beta; A, U(N))$ for A, defined by

$$F(\alpha, \beta; A, U(N)) = \frac{1}{N}|\{(\phi_n, \phi_m) \in \Phi^2 : \phi_n - \phi_m \in [\alpha, \beta]\}|,$$

for any real α and β with $\alpha < \beta$. Since $U(N)$ has a Haar measure dA, we can compute the global distribution of $F(\alpha, \beta; A, U(N))$ over $U(N)$, namely $F(\alpha, \beta; U(N))$, defined by

$$F(\alpha, \beta; U(N)) = \int_{U(N)} F(\alpha, \beta; A, U(N))\, dA.$$

Analogously, we want to understand if the limiting distribution

$$F(\alpha, \beta; U) = \lim_{N \to \infty} F(\alpha, \beta; U(N)),$$

exists and what can be said about it. Dyson showed the following (see [Dys62] for a proof):

Proposition 10.1.1. *For any real α and β with $\alpha < \beta$, $F(\alpha, \beta; U)$ exists. Moreover,*

$$F(\alpha, \beta; U) = \int_{\alpha}^{\beta} \left(1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2 + \delta(x)\right) dx,$$

where $\delta(x)$ is the Dirac delta function.

The right-hand side of Proposition 10.1.1 is exactly the same formula given in Montgomery's pair correlation conjecture. In other words, if Montgomery's pair correlation conjecture is true then the two-point correlation of the unfolded nontrivial zeros of the Riemann zeta function in the limit as we move up the critical line exactly match the two-point correlation of the unfolded eigenphases of unitary matrices in the limit as the size of the matrices increase. In short, statistical information about the Riemann zeta function agrees with statistical information about the eigenvalues of unitary matrices. This is the origin of the Katz-Sarnak philosophy.

10.2 The Work of Katz & Sarnak

Katz and Sarnak generalized the work of Montgomery and Dyson by establishing a connection between families of L-functions and other compact matrix groups. For the ease of categorization, Katz and Sarnak associated a **symmetry type** to each compact matrix group that they studied. The underlying compact matrix group associated to each symmetric type is called a **matrix ensemble**. The symmetry types and associated matrix ensembles are described in the following table:

Symmetry Type	Matrix Ensemble
Unitary (U)	U(N)
Orthogonal (O^+)	SO($2N$)
Orthogonal (O^-)	SO($2N+1$)
Symplectic (Sp)	USp($2N$)

Let $G(N)$ be a matrix ensemble, where G denotes the symmetry type, and let dA denote the Haar measure. For any $A \in G(N)$, let $\Phi = (\phi_n)_{1\le n\le N}$ denote the increasing sequence of unfolded eigenphases of A. Katz and Sarnak considered two local spacing distributions between the unfolded eigenphases of A. The first was the **two-point correlation function** $F(\alpha, \beta; A, G(N))$ for A, defined by

$$F(\alpha, \beta; A, G(N)) = \frac{1}{N}|\{(\phi_n, \phi_m) \in \Phi^2 : \phi_n - \phi_m \in [\alpha, \beta]\}|,$$

for any real α and β with $\alpha < \beta$. They computed the global distribution of $F(\alpha, \beta; A, G(N))$ over $G(N)$, namely $F(\alpha, \beta; G(N))$, defined by

$$F(\alpha, \beta; G(N)) = \int_{G(N)} F(\alpha, \beta; A, G(N))\, dA,$$

and sought to understand if the limiting distribution

$$F(\alpha, \beta; G) = \lim_{N\to\infty} F(\alpha, \beta; G(N)),$$

exists and what can be said about it. Note that in the case $G(N) = U(N)$, this is exactly the two-point correlation function considered by Dyson. Katz and Sarnak succeeded in generalizing Dyson's work (see [KS23] for a proof):

Proposition 10.2.1. *For symmetry type G and any real α and β with $\alpha < \beta$, $F(\alpha, \beta; G)$ exists. Moreover,*

$$F(\alpha, \beta; G) = \int_\alpha^\beta \left(1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2 + \delta(x)\right) dx,$$

where $\delta(x)$ is the Dirac delta function.

The second local spacing distribution was the **k-th consecutive spacing function** for A, defined by

$$\mu_k(\alpha, \beta; A, G(N)) = \frac{1}{N}|\{1 \le j \le N : (\phi_{j+k}, \phi_j) \in \Phi^2 \text{ and } \phi_{j+k} - \phi_j \in [\alpha, \beta]\}|,$$

for any real α and β with $\alpha < \beta$ and $k \ge 1$. Again, they proceeded to compute the global distribution over $G(N)$, namely $\mu_k(\alpha, \beta; G(N))$, defined by

$$\mu_k(\alpha, \beta; G(N)) = \int_{G(N)} \mu_k(\alpha, \beta; A, G(N))\, dA,$$

and asked if the limiting distribution

$$\mu_k(\alpha, \beta; G) = \lim_{N\to\infty} \mu_k(\alpha, \beta; G(N)),$$

exists and what can be said about it. They were able the show the following (see [KS23] for a proof):

Proposition 10.2.2. *For any symmetry type G and any real α and β with $\alpha < \beta$ and $k \geq 1$, $\mu_k(\alpha, \beta; G)$ exists. Moreover, it is independent of the particular symmetry type G.*

In particular, Propositions 10.2.1 and 10.2.2 show that the limiting distributions $F(\alpha, \beta; G)$ and $\mu_k(\alpha, \beta; G)$ are both independent of the symmetry type G. However, this symmetry independence is not true for all limiting distributions. Katz and Sarnak also considered a local and global distributions associated to single eigenphases. The local distribution they considered, associated to a single eigenphase, was the **one-level density function** $\Delta(\alpha, \beta; A, G(N))$ for A, defined by

$$\Delta(\alpha, \beta; A, G(N)) = |\{\phi \in \Phi : \phi \in [\alpha, \beta]\}|.$$

They computed the global distribution, namely $\Delta(\alpha, \beta; G(N))$, defined by

$$\Delta(\alpha, \beta; G(N)) = \int_{G(N)} \Delta(\alpha, \beta; A, G(N))\, dA,$$

and asked if the limiting distribution

$$\Delta(\alpha, \beta; G) = \lim_{N \to \infty} \Delta(\alpha, \beta; G(N)),$$

exists and what can be said about it. Precisely, they proved the following (see [KS23] for a proof):

Proposition 10.2.3. *For any symmetry type G and any real α and β with $\alpha < \beta$ and $k \geq 1$, $\Delta(\alpha, \beta; G)$ exists. Moreover, it depends upon the particular symmetry type G.*

The global distribution they considered, associated to a single eigenphase, was the **k-th eigenphase function** $\nu_k(\alpha, \beta, G(N))$ for $G(N)$, defined by

$$\nu_k(\alpha, \beta; G(N)) = dA\left(\{A \in G(N) : \phi_k \in [\alpha, \beta]\}\right),$$

for any real α and β with $\alpha < \beta$ and $k \geq 1$. Again, they asked if the limiting distribution

$$\nu_k(\alpha, \beta; G) = \lim_{N \to \infty} \nu_k(\alpha, \beta; G(N)),$$

exists and what can be said about it. They were able to show the following (see [KS23] for a proof):

Proposition 10.2.4. *For any symmetry type G and any real α and β with $\alpha < \beta$ and $k \geq 1$, $\nu_k(\alpha, \beta; G)$ exists. Moreover, it depends upon the particular symmetry type G.*

In short, Katz and Sarnak studied four limiting distributions $F(\alpha, \beta; G)$, $\mu_k(\alpha, \beta; G)$, $\Delta(\alpha, \beta; G)$, and $\nu_k(\alpha, \beta; G)$. The former two are distributions about collections of eigenphases of unitary matrices and are independent of the symmetry type G while the latter two are distributions about single eigenphases of unitary matrices and depend upon the symmetry type G. Analogous distributions can be defined for families of L-functions. We say that a collection of L-functions $\mathcal{F} = (L(s_\alpha, f_\alpha))_{\alpha \in I}$, for some infinite indexing set $I \subset \mathbb{R}_{\geq 0}$, is a **family** if it is an ordered set with respect to the analytic conductor and if $\mathfrak{q}(s_\alpha, f_\alpha) \to \infty$ as $\alpha \to \infty$. We say that the family $\mathcal{F}$ is **continuous** if $f_\alpha = f_\beta$ for all $\alpha, \beta \in I$ and $s_\alpha = \sigma + it_\alpha$ where t_α is a continuous function of α. Necessarily, $t_\alpha \to \infty$ as $\alpha \to \infty$ and I is a half-open ray. We say that a family $\mathcal{F}$ is **discrete** if $f_\alpha \neq f_\beta$ for all distinct $\alpha, \beta \in I$ and I is discrete. Necessarily, $q(f_\alpha) \to \infty$ as $\alpha \to \infty$ and, reindexing if necessary, $I \subseteq \mathbb{N}$. Katz and Sarnak arrived at a heuristical conjecture known as the **Katz-Sarnak philosophy** in terms of families of L-functions:

Conjecture 10.2.1 (Katz-Sarnak Philosophy).

(i) The statistics about collections of eigenphases of a random matrix belonging to a matrix ensemble of symmetry type G, in the limit as the size of the matrix tends to infinity, should model statistics about the nontrivial zeros of a continuous family of L-functions as the heights of the nontrivial zeros tends to infinity.

(ii) The statistics about a single eigenphase of a random matrix belonging to a matrix ensemble of symmetry type G, in the limit as the size of the matrix tends to infinity, should model statistics about a discrete family of L-functions as the size of the conductor tends to infinity.

Determining the symmetry type of a family is generally a difficult task. Below are some well-studied families and their symmetry types (see [CFK$^+$05] for a determination of the symmetry type):

Symmetry Type	Family
Unitary	$\{L(\sigma+it,f) : L(s,f)$ is of Selberg class$\}$ ordered by t $\{L(s,\chi) : \chi$ is a primitive character modulo $q\}$ ordered by q
Orthogonal	$\{L(s,f) : f \in \mathcal{S}_k(N,\chi)$ is a primitive Hecke newform$\}$ ordered by k $\{L(s,f) : f \in \mathcal{S}_k(N,\chi)$ is a primitive Hecke newform$\}$ ordered by N
Symplectic	$\{L(s,\chi_D) : D$ is a fundamental discriminant$\}$ ordered by $\lvert D\rvert$

10.3 Characteristic Polynomials of Unitary Matrices

The Katz-Sarnak philosophy conjectures that the statistics about nontrivial zeros of L-functions are modeled by the statistics of eigenphases of random unitary matrices. However, there is a striking surface level connection between L-functions and the characteristic polynomials of unitary matrices which we now describe. For $A \in U(N)$, let

$$L(s,A) = \det(I - sA) = \prod_{1 \le n \le N} (1 - se^{i\theta_n}),$$

be the characteristic polynomial of A. It will turn out that $L(s,A)$ has strikingly similar properties to an L-function. The product expression for $L(s,A)$ is clearly analogous to the Euler product expression for an L-function. Upon expanding the product, we obtain

$$L(s,A) = \sum_{0 \le n \le N} a_n s^n,$$

for some coefficients a_n. This expression is the analogue to the Dirichlet series representation for an L-function. Of course, as $L(s,A)$ is a polynomial it is analytic on $\mathbb{C}$. Moreover, $L(s,A)$ possesses a functional equation of shape $s \to \frac{1}{s}$. To see this, first observe that multiplicativity of the determinant gives

$$L(s,A) = (-1)^N \det(A) s^N \det(I - s^{-1}A^{-1}).$$

As A is unitary, $L(s,A^{-1}) = L(s,A^*) = L(s,\overline{A})$. So the above equation can be expressed as

$$L(s,A) = (-1)^N \det(A) s^N L\left(\frac{1}{s}, \overline{A}\right).$$

This is the analogue of the functional equation for $L(s, A)$ and it is of shape $s \to \frac{1}{s}$. We identify the analogues of the gamma factor and conductor as 1 and N respectively. Letting $\Lambda(s, A)$ be defined by

$$\Lambda(s, A) = s^{-\frac{N}{2}} L(s, A),$$

the functional equation can be expressed as

$$\Lambda(s, A) = (-1)^N \det(A) \Lambda\left(\frac{1}{s}, \overline{A}\right).$$

From it, the analogue of root number is seen to be $(-1)^N \det(A)$ and $L(s, A)$ has dual $L(s, \overline{A})$. As the transformation $s \to \frac{1}{s}$ leaves the unit circle invariant, the unit circle is the analogue of the critical line. The fixed point of the transformation $s \to \frac{1}{s}$ is $s = 1$ which is the analogue of the central point. Moreover, as the zeros of $L(s, A)$ are precisely the eigenvalues of A which lie on the unit circle, because A is unitary, the analogue of the Riemann hypothesis is true for $L(s, A)$. We also have an analogue of the approximate functional equation. By substituting the polynomial representation of $L(s, A)$ into the functional equation, we obtain

$$\sum_{0 \le n \le N} a_n s^n = (-1)^N \det(A) s^N \sum_{0 \le n \le N} \overline{a_n} s^{-n} = (-1)^N \det(A) \sum_{0 \le n \le N} \overline{a_n} s^{N-n}.$$

Upon comparing coefficients, we find that

$$a_n = (-1)^N \det(A) \overline{a_{N-n}},$$

for $0 \le n \le N$. So for odd N,

$$L(s, A) = \sum_{0 \le n \le \frac{N-1}{2}} a_n s^n + (-1)^N \det(A) s^N \sum_{0 \le n \le \frac{N-1}{2}} \overline{a_n} s^{-n},$$

and for even N,

$$L(s, A) = a_{\frac{N}{2}} s^{\frac{N}{2}} + \sum_{0 \le n \le \frac{N}{2} - 1} a_n s^n + (-1)^N \det(A) s^N \sum_{0 \le n \le \frac{N}{2} - 1} \overline{a_n} s^{-n}.$$

These equations together are the analogue of the approximate functional equation. This similarity between L-functions and the characteristic polynomials of unitary matrices was heavily exploited by Conrey, Farmer, Keating, Rubinstein, and Snaith to make phenomenal conjectures about the moments of L-functions.

Chapter 11

The Theory of Moments of L-functions

The study of moments of L-functions is essentially the method of studying a single L-function, or a family of L-functions, on average. While this approach leads to weaker results for a single L-function, it is often more malleable and sheds light on the general behavior of these objects. In the following we introduce moments of a single L-function, moments of families, and discuss some of their longstanding conjectures.

11.1 Moments of L-functions

One of the longstanding goal in the study of L-functions is to prove the Riemann hypothesis for the Riemann zeta function. More generally, we wish to prove the Riemann hypothesis for different types of L-functions. Another longstanding goal would be to prove the Lindelöf hypothesis for the Riemann zeta function, or more generally, any Selberg class L-function. While both the Riemann and Lindelöf hypotheses remain out of reach for any Selberg class L-function $L(s,f)$, the latter is more tractable because the convexity bound gives the estimate

$$L\left(\frac{1}{2}+it,f\right) \ll_\varepsilon \mathfrak{q}\left(\frac{1}{2}+it,f\right)^{\delta+\varepsilon},$$

with $\delta=\frac{1}{4}$ whereas the Lindelöf hypothesis claims that we can take $\delta=0$. As we have already mentioned, breaking the convexity bound $\delta=\frac{1}{4}$ is a very daunting task requiring extremely modern techniques and has only been successful in a few cases. This approach of breaking convexity to study the Lindelöf hypothesis is quite direct. There is another approach to study the Lindelöf hypothesis that is more indirect by considering averages of the L-function in question. This is the study of moments of L-functions. For any L-function $L(s,f)$ and $k \ge 1$, we define the k-th **moment** of $L(s,f)$, namely $M_k(T,f)$, to be

$$M_k(T,f) = \int_0^T \left|L\left(\frac{1}{2}+it,f\right)\right|^k dt,$$

for any $T>0$. More generally, we can define moments for any continuous or discrete family $\mathcal{F}$ of L-functions. For any $k \ge 1$, we define the k-th **moment** of $\mathcal{F}$, namely $M_k(T,\mathcal{F};\sigma)$ or $M_k(Q,\mathcal{F};s)$, to be

$$M_k(T,\mathcal{F};\sigma) = \int_0^T |L(\sigma+it,f)|^k\,dt \quad \text{or} \quad M_k(Q,\mathcal{F};s) = \sum_{\substack{f\in\mathcal{F}\\ q(f)\le Q}} |L(s,f)|^k,$$

according to if $\mathcal{F}$ is continuous or discrete. We call these moments **continuous** and **discrete** respectively. Note that continuous moments only depend upon σ while discrete moments may depend upon s. Moreover, we suppress the dependence upon σ or s respectively if $s=\frac{1}{2}$. So for a continuous family, $M_k(T,\mathcal{F}) =$

$M_k(T,f)$. Generally speaking, we are usually interested in proving asymptotics for moments in terms of the parameter of the analytic conductor that is approaching infinity for the particular family. For continuous families this parameter is t and for discrete families it is $q(f)$. In the case of $M_k(T,f)$, it is useful to think of this moment as essentially a k-power average version of the Lindelöf hypothesis for $L(s,f)$. There is actually a close connection to the Lindelöf hypothesis since sufficient estimates for $M_{2k}(T,f)$ in T for all $k \geq 1$ is equivalent to the Lindelöf hypothesis as the following proposition shows:

Proposition 11.1.1. *Let $L(s,f)$ be an L-function. The truth of the Lindelöf hypothesis for $L(s,f)$ is equivalent to the estimate*

$$M_{2k}(T,f) \ll_\varepsilon T^{1+\varepsilon},$$

for all $k \geq 1$.

Proof. First suppose the Lindelöf hypothesis for $L(s,f)$ holds. As $\mathfrak{q}\left(\frac{1}{2}+it,f\right) \ll t^{d_f}$, Lindelöf hypothesis for $L(s,f)$ implies

$$M_{2k}(T,f) \ll \int_0^T t^\varepsilon\,dt \ll T^{1+\varepsilon},$$

as desired. Now suppose that $M_{2k}(T,f) \ll_\varepsilon T^{1+\varepsilon}$ for all $k \geq 1$. If the Lindelöf hypothesis for $L(s,f)$ is false, then there is some $0 < \lambda < 1$ and positive unbounded real sequence $(t_n)_{n\geq 1}$ such that

$$\left|L\left(\frac{1}{2}+it_n,f\right)\right| > C\mathfrak{q}\left(\frac{1}{2}+it_n,f\right)^\lambda,$$

for any $C > 0$. In particular,

$$\left|L\left(\frac{1}{2}+it,f\right)\right| > Ct_n^{d_f\lambda}.$$

Now the convexity bound for $L'(s,f)$ implies that $L\left(\frac{1}{2}+it\right) \leq C't^{d_f}$ for some $C' > 0$ and t bounded away from zero. Then

$$\left|L\left(\frac{1}{2}+it,f\right) - L\left(\frac{1}{2}+it_n,f\right)\right| = \left|\int_{t_n}^t L\left(\frac{1}{2}+ir,f\right)\,dr\right| < C't_n^{d_f}|t-t_n| < \frac{C}{2}t_n^{d_f\lambda},$$

where the last inequality holds provided $|t-t_n| \leq t_n^{d_f(\lambda-1)}$ and sufficiently large n (hence t_n and t are bounded away from zero). Then for such n, the previous two bounds together imply

$$\left|L\left(\frac{1}{2}+it,f\right)\right| > \frac{C}{2}t_n^{d_f\lambda},$$

for $|t-t_n| \leq t_n^{d_f(\lambda-1)}$. Now take $T = \frac{4}{3}t_n$ so that the interval $(t_n - t_n^{d_f(\lambda-1)}, t_n + t_n^{d_f(\lambda-1)})$ is contained in $\left(\frac{T}{2},T\right)$ for sufficiently large n. Then

$$\int_{\frac{T}{2}}^T \left|L\left(\frac{1}{2}+it,f\right)\right|^{2k}\,dt > \int_{t_n-t_n^{d_f(\lambda-1)}}^{t_n+t_n^{d_f(\lambda-1)}} \left|L\left(\frac{1}{2}+it,f\right)\right|^{2k}\,dt > 2\left(\frac{C}{2}\right)^{2k} t_n^{(2k+1)d_f\lambda-d_f},$$

and it follows that

$$M_{2k}(T,f) > 2\left(\frac{C}{2}\right)^{2k}\left(\frac{3}{4}\right)^{(2k+1)d_f\lambda-d_f} T^{(2k+1)d_f\lambda-d_f},$$

which is a contradiction for sufficiently large k. This completes the proof. $\square$

The usefulness of Proposition 11.1.1 is that the difficulty of proving the Lindelöf hypothesis has been transferred to proving asymptotics for the moments $M_{2k}(T,f)$ each of which should, heuristically speaking, be an easier problem to resolve on its own.

11.2 Todo: [The CFKRS Conjectures]

Hardy and Littlewood were the first to introduce moments and they did so in the context of the Riemann zeta function. In 1918, they proved an asymptotic for the second moment (see [HL16] for a proof):

$$M_2(T,\zeta) \sim T\log(T).$$

In 1926, Ingham obtained an asymptotic for the fourth moment (see [Ing28] for a proof):

$$M_4(T,\zeta) \sim \frac{1}{2\pi^2}T\log^4(T).$$

Observe that both of these asymptotics are stronger than those in Proposition 11.1.1. In other words, this is evidence in support of the truth of the Lindelöf hypotheses for $\zeta(s)$. Unfortunately, this is where progress significantly halts. No analogous formula have been obtained for moments of any L-function when $k > 2$. In fact, the problem is so intractable that, until recently, there were not even conjectures about what the asymptotics should be. In 2000, Keating and Snaith used the Katz-Sarnak philosophy to put forth a precise conjecture for the asymptotics of the moments of the Riemann zeta function (see [KS00] for details):

Conjecture 11.2.1. *For all $k \geq 1$,*

$$M_{2k}(T,\zeta) \sim \frac{g_k}{(k^2)!}a_k T\log^{k^2}(T),$$

with

$$g_k = (k^2)!\prod_{0\leq j\leq k-1}\frac{j!}{(j+k)!},$$

and

$$a_k = \prod_p\left((1-p^{-1})^{(k-1)^2}\sum_{0\leq j\leq k-1}\binom{k-1}{j}^2 p^{-j}\right).$$

Conjecture 11.2.1 agrees with the results of Hardy and Littlewood for $k=1$ (note that $g_1 = 1$ and $a_1 = 1$) and Ingham for $k=2$ (note that $g_2 = 2$ and $a_2 = \zeta(2)^{-1} = \frac{6}{\pi^2}$ because it is the sum of the reciprocals of squares). Monumental progress was made in 2005 when Conrey, Farmer, Keating, Rubinstein, and Snaith used random matrix theory to put forth a procedure for deducing conjectured asymptotic formulas, not just moments of the Riemann zeta function, but for many families of L-functions (see [CFK$^+$05]). Moreover, they proved that analogous statistics hold for the associated symmetry types of the families which is in agreement with the Katz-Sarnak philosophy. This procedure is now known as the **CFKRS recipe** for moments and we will describe some of their conjectures. For a unitary example, the CFKRS recipe predicts the refined asymptotic for the Riemann zeta function:

Conjecture 11.2.2. *For all $k \geq 1$, $T \geq 1$, and $\varepsilon > 0$,*

$$M_{2k}(T,\zeta) = TP_k(\log(T)) + O_\varepsilon(T^{\frac{1}{2}+\varepsilon}),$$

where P_k is some polynomial of degree k^2 with leading coefficient $\frac{g_k}{(k^2)!}a_k$ with

$$g_k = (k^2)!\prod_{0\leq j\leq k-1}\frac{j!}{(j+k)!},$$

and

$$a_k = \prod_p\left((1-p^{-1})^{(k-1)^2}\sum_{0\leq j\leq k-1}\binom{k-1}{j}^2 p^{-j}\right).$$

Conjecture 11.2.2 has been proven in the cases $k = 1, 2$ (see [CFK$^+$05] for comments). Moreover, this result is clearly stronger than Conjecture 11.2.1 and it gives an exact error of order $O(T^{\frac{1}{2}+\varepsilon})$. This error is roughly on the order of the square-root of the main term. For an orthogonal example, the CFKRS recipe predicts an asymptotic for a family of Hecke L-functions:

Conjecture 11.2.3. *Let $\mathcal{H}$ be the family of bases of primitive Hecke eigenforms for the space of newforms of weight 2, square-free level q, and ordered by q. For all $k \geq 1$, $Q \geq 1$, and $\varepsilon > 0$,*

$$\mathcal{M}_k(Q, \mathcal{H}) = \frac{1}{3} Q R_k(\log(Q)) + O(Q^{\frac{1}{2}} + \varepsilon),$$

where R_k is some polynomial of degree $\frac{1}{2}k(k-1)$ with leading coefficient $\frac{g_k}{\left(\frac{1}{2}k(k-1)\right)!} a_k$ with

$$g_k = 2^{k-1} \left(\frac{1}{2}k(k-1)\right)! \prod_{1 \leq j \leq k-1} \frac{j!}{(2j)!},$$

and

$$a_k = \prod_{p \nmid q} (1 - p^{-1})^{\frac{1}{2}k(k-1)} \frac{2}{\pi} \int_0^{\pi} \sin^2\left(\theta \left(\frac{e^{i\theta}(1 - e^{i\theta}p^{-\frac{1}{2}})^{-1} - e^{-i\theta}(1 - e^{i\theta}p^{-\frac{1}{2}})^{-1}}{e^{i\theta} - e^{-i\theta}}\right)^k\right) d\theta.$$

The main term in Conjecture 11.2.3 has been proved in the cases $k = 1, 2, 3, 4$ when q is prime (see [CFK$^+$05] for comments). For an symplectic example, the CFKRS recipe predicts an asymptotic for a family of Dirichlet L-functions:

Conjecture 11.2.4. *Let $\mathcal{D}$ be the family of Dirichlet L-functions associated to the quadratic characters χ_d attached to fundamental discriminants d and ordered by $|d|$. For all $k \geq 1$, $D \geq 1$, and $\varepsilon > 0$,*

$$\mathcal{M}_k(D, \mathcal{D}) = \frac{6}{\pi^2} D Q_k(\log(D)) + O(D^{\frac{1}{2}+\varepsilon}),$$

where Q_k is some polynomial of degree $\frac{1}{2}k(k+1)$ with leading coefficient $\frac{g_k}{\left(\frac{1}{2}k(k+1)\right)!} a_k$ with

$$g_k = \left(\frac{1}{2}k(k+1)\right)! \prod_{1 \leq j \leq k-1} \frac{j!}{(2j)!},$$

and

$$a_k = \prod_p \frac{(1 - p^{-1})^{\frac{1}{2}k(k+1)}}{(1 + p^{-1})} \left(\frac{(1 - p^{-\frac{1}{2}})^{-k} + (1 + p^{-\frac{1}{2}})^{-k}}{2} + p^{-1}\right).$$

The main term in Conjecture 11.2.4 has been proved in the cases $k = 1, 2, 3$ (see [CFK$^+$05] for comments).

Part VI

Appendices

Appendix A

Number Theory

A.1 Arithmetic Functions

An arithmetic function f is a function $f : \mathbb{N} \to \mathbb{C}$. That is, it takes the positive integers into the complex numbers. We say that f is **additive** if $f(nm) = f(n) + f(m)$ for all positive integers n and m such that $(n, m) = 1$. If this condition simply holds for all n and m then we say f is **completely additive**. Similarly, we say that f is **multiplicative** if $f(nm) = f(n)f(m)$ for all positive integers n and m such that $(n, m) = 1$. If this condition simply holds for all n and m then we say f is **completely multiplicative**. Many important arithmetic functions are either additive, completely additive, multiplicative, or completely multiplicative. Note that if a f is additive or multiplicative then f is uniquely determined by its values on prime powers and if f is completely additive or completely multiplicative then it is uniquely determined by its values on primes. Moreover, if f is additive or completely additive then $f(1) = 0$ and if f is multiplicative or completely multiplicative then $f(1) = 1$. Below is a list defining the most important arithmetic functions (some of these functions are restrictions of common functions but we define them here as arithmetic functions because their domain being $\mathbb{N}$ is important):

(i) The **constant function**: The function $\mathbf{1}(n)$ restricted to all $n \geq 1$. This function is neither additive or multiplicative.

(ii) The **unit function**: The function $e(n)$ defined by

$$e(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \geq 2. \end{cases}$$

This function is completely multiplicative.

(iii) The **identity function**: The function $\mathrm{id}(n)$ restricted to all $n \geq 1$. This function is completely multiplicative.

(iv) The **logarithm**: The function $\log(n)$ restricted to all $n \geq 1$. This function is completely additive.

(v) The **Möbius function**: The function $\mu(n)$ defined by

$$\mu(n) = \begin{cases} 1 & \text{if } n \text{ is square-free with an even number of prime factors}, \\ -1 & \text{if } n \text{ is square-free with an odd number of prime factors}, \\ 0 & \text{if } n \text{ is not square-free}, \end{cases}$$

for all $n \geq 1$. This function is multiplicative.

(vi) The **characteristic function of square-free integers**: The square of the Möbius function $\mu^2(n)$ for all $n \geq 1$. This function is multiplicative.

(vii) **Liouville's function**: The function $\lambda(n)$ defined by

$$\lambda(n) = \begin{cases} 1 & \text{if } n = 1, \\ (-1)^k & \text{if } n \text{ is composed of } k \text{ not necessarily distinct prime factors,} \end{cases}$$

for all $n \geq 1$. This function is completely multiplicative.

(viii) **Euler's totient function**: The function $\varphi(n)$ defined by

$$\varphi(n) = \sum_{m \pmod n}' 1,$$

for all $n \geq 1$. This function is multiplicative.

(ix) The **divisor function**: The function $\sigma_0(n)$ defined by

$$\sigma_0(n) = \sum_{d|n} 1,$$

for all $n \geq 1$. This function is multiplicative.

(x) The **sum of divisors function**: The function $\sigma_1(n)$ defined by

$$\sigma_1(n) = \sum_{d|n} d,$$

for all $n \geq 1$. This function is multiplicative.

(xi) The **generalized sum of divisors function**: The function $\sigma_s(n)$ defined by

$$\sigma_s(n) = \sum_{d|n} d^s,$$

for all $n \geq 1$ and any complex number s. This function is multiplicative.

(xii) The **number of distinct prime factors function**: The function $\omega(n)$ defined by

$$\omega(n) = \sum_{p|n} 1,$$

for all $n \geq 1$. This function is additive.

(xiii) The **total number of prime divisors function**: The function $\Omega(n)$ defined by

$$\Omega(n) = \sum_{p^m|n} 1,$$

for all $n \geq 1$ and where $m \geq 1$. This function is completely additive.

(xiv) The **von Mangoldt function**: The function $\Lambda(n)$ defined by

$$\Lambda(n) = \begin{cases} 0 & \text{if } n \text{ is not a prime power,} \\ \log(p) & \text{if } n = p^m \text{ for some prime } p \text{ and integer } m \geq 1, \end{cases}$$

for all $n \geq 1$. This function is neither additive or multiplicative.

If f and g are two arithmetic functions, then we can define a new arithmetic function $f * g$ called the **Dirichlet convolution** of f and g defined by

$$(f * g)(n) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right),$$

for all $n \geq 1$. This is especially useful when f and g are multiplicative:

Proposition A.1.1. *If f and g are multiplicative arithmetic functions, then so is their Dirichlet convolution $f * g$.*

A.2 The Möbius Function

Recall that the Möbius function is the arithmetic function μ defined by

$$\mu(n) = \begin{cases} 1 & \text{if } n \text{ is square-free with an even number of prime factors,} \\ -1 & \text{if } n \text{ is square-free with an odd number of prime factors,} \\ 0 & \text{if } n \text{ is not square-free,} \end{cases}$$

and it is multiplicative. It also satisfies an important summation property:

Proposition A.2.1.

$$\sum_{d|n} \mu(d) = \delta_{n,1}.$$

From this property, the important **Möbius inversion formula** can be derived:

Theorem A.2.1 (Möbius inversion formula). *Suppose f and g are arithmetic functions. Then*

$$g(n) = \sum_{d|n} f(d),$$

for all $n \geq 1$, if and only if

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

for all $n \geq 1$.

In terms of Dirichlet convolution, the Möbius inversion formula is equivalent to stating that $g = f * \mathbf{1}$ if and only if $f = g * \mu$. Using Möbius inversion, the following useful formula can also be derived:

Proposition A.2.2. *For $\sigma > 1$,*

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

There is also an important similar statement to the Möbius inversion formula that we will need:

Theorem A.2.2. *Let f be an arithmetic function and let B be the completely multiplicative function defined on primes p by*

$$B(p) = f(p)^2 - f(p^2).$$

Then

$$f(n)f(m) = \sum_{d|(n,m)} B(d) f\left(\frac{nm}{d^2}\right),$$

for all $n, m \geq 1$, if and only if

$$f(nm) = \sum_{d|(n,m)} \mu(d)B(d) f\left(\frac{n}{d}\right) f\left(\frac{m}{d}\right),$$

for all $n, m \geq 1$.

Any arithmetic function f satisfying the conditions of Theorem A.2.2 is said to be **specially multiplicative**.

A.3 The Sum and Generalized Sum of Divisors Functions

It is very useful to know that $\sigma_0(n)$ grows slowly:

Proposition A.3.1.

$$\sigma_0(n) \ll_\varepsilon n^\varepsilon.$$

This is all we really need to know for the sum of divisors function. As for the generalized sum of divisors function, it has the remarkable property that it can be written as a product. To state it, recall that $\mathrm{ord}_p(n)$ is the positive integer satisfying $p^{\mathrm{ord}_p(n)} \mid\mid n$. Then we have the following statement:

Proposition A.3.2. *For $s \neq 0$,*

$$\sigma_s(n) = \prod_{p|n} \frac{p^{(\mathrm{ord}_p(n)+1)s} - 1}{p^s - 1}.$$

A.4 Quadratic Symbols

Let p be an odd prime. We are often interested in when the equation $x^2 = m \pmod{p}$ is solvable for some $m \in \mathbb{Z}$. The **Legendre symbol** $\left(\frac{m}{p}\right)$ keeps track of this:

$$\left(\frac{m}{p}\right) = \begin{cases} 1 & \text{if } x^2 \equiv m \pmod{p} \text{ is solvable,} \\ -1 & \text{if } x^2 \equiv m \pmod{p} \text{ is not solvable,} \\ 0 & \text{if } m \equiv 0 \pmod{p}. \end{cases}$$

Euler's criterion gives an alternative expression for Legendre symbol when m is coprime to p:

Proposition A.4.1 (Euler's criterion). *Let p be an odd prime and suppose $m \in \mathbb{Z}$ with $(m, p) = 1$. Then*

$$\left(\frac{m}{p}\right) = m^{\frac{p-1}{2}} \pmod{p}.$$

From the definition and Euler's criterion is it not difficult to show that the Legendre symbol satisfies the following properties:

Proposition A.4.2. *Let p be an odd prime and let $a, b \in \mathbb{Z}$. Then the following hold:*

(i) If $a \equiv b \pmod p$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(ii) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.

From Proposition A.4.2, to compute the Legendre symbol in general it suffices to know how to compute $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$, and $\left(\frac{q}{p}\right)$ where q is another odd prime. The **supplemental laws of quadratic reciprocity** are formulas for the first two symbols:

Proposition A.4.3 (Supplemental laws of quadratic reciprocity). *Let p be an odd prime. Then the following formulas hold:*

(i)

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4, \\ -1 & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

(ii)

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv 1, 7 \pmod 8, \\ -1 & \text{if } p \equiv 3, 5 \pmod 8. \end{cases}$$

The **law of quadratic reciprocity** handles the last symbol by relating $\left(\frac{q}{p}\right)$ to $\left(\frac{p}{q}\right)$:

Theorem A.4.1 (Law of quadratic reciprocity). *Let p and q be odd primes. Then*

$$\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\left(\frac{q}{p}\right).$$

We can generalize the Jacobi symbol further by making it multiplicative in the denominator. Let n be a positive odd integer with prime factorization $n = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$ and let $m \in \mathbb{Z}$. The **Jacobi symbol** $\left(\frac{m}{n}\right)$ is defined by

$$\left(\frac{m}{n}\right) = \prod_{1 \le i \le k} \left(\frac{m}{p_i}\right)^{r_i}.$$

When $n = p$ is prime, the Jacobi symbol reduces to the Legendre symbol and the Jacobi symbol is precisely the unique multiplicative extension of the Legendre symbol to all positive odd integers. Accordingly, the Jacobi symbol has the following properties:

Proposition A.4.4. *Let m and n be positive odd integers and let $a, b \in \mathbb{Z}$. Then the following hold:*

(i) If $a \equiv b \pmod p$, then $\left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$.

(ii) $\left(\frac{a}{mn}\right) = \left(\frac{a}{m}\right)\left(\frac{a}{n}\right)$.

(iii) $\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right)$.

There is also an associated reciprocity law:

Proposition A.4.5. *Let m and n be positive odd integers. Then*

$$\left(\frac{m}{n}\right) = (-1)^{\frac{m-1}{2}\frac{n-1}{2}} \left(\frac{n}{m}\right).$$

We can further generalize the Jacobi symbol so that it is valid for all integers. Let $m, n \in \mathbb{Z}$ where $n = up_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$ is the prime factorization of n with $u = \pm 1$. The **Kronecker symbol** $\left(\frac{m}{n}\right)$ is defined by

$$\left(\frac{m}{n}\right) = \left(\frac{m}{u}\right) \prod_{1 \le i \le k} \left(\frac{m}{p_i}\right)^{r_i},$$

where we set

$$\left(\frac{m}{1}\right) = 1, \quad \left(\frac{m}{-1}\right) = \begin{cases} 1 & \text{if } m \ge 0, \\ -1 & \text{if } m < 0, \end{cases} \quad \left(\frac{m}{0}\right) = \begin{cases} 1 & \text{if } m = \pm 1, \\ 0 & \text{if } m \ne \pm 1, \end{cases}$$

and

$$\left(\frac{m}{2}\right) = \begin{cases} 1 & \text{if } m \equiv 1, 7 \pmod 8, \\ -1 & \text{if } m \equiv 3, 5 \pmod 8, \\ 0 & \text{if } m \equiv 0 \pmod 2). \end{cases}$$

When n is a positive odd integer, the Kronecker symbol reduces to the Jacobi symbol. The Kronecker symbol also satisfies a reciprocity law:

Proposition A.4.6. *Let m and n be integers. Then*

$$\left(\frac{m}{n}\right) = (-1)^{\frac{m^{(2)}-1}{2}\frac{n^{(2)}-1}{2}} \left(\frac{n}{|m|}\right),$$

where $m^{(2)}$ and $n^{(2)}$ are the parts of m and n relatively prime to 2 respectively.

Appendix B

Analysis

B.1 Local Absolute Uniform Convergence

Throughout, all functions are understood to complex-valued and regions are understood to be those of $\mathbb{R}^n$ or $\mathbb{C}^n$. Often, we are interested in some series

$$\sum_{n\geq 1} f_n(\mathbf{x}),$$

where of analytic functions $f_n(\mathbf{x})$ on some region D. We say that the series above is **locally absolutely uniformly convergent** if

$$\sum_{n\geq 1} |f_n(\mathbf{x})|,$$

converges uniformly on compact subsets of D. A common method to determine if series is absolutely uniformly convergent is to use the Weierstrass M-test:

Theorem B.1.1 (Weierstrass M-test). *Suppose $(f_n(\mathbf{x}))_{n\geq 1}$ is a sequence of functions on a region D and there is a nonnegative sequence $(M_n)_{n\geq 1}$ such that $|f_n(\mathbf{x})| \leq M_n$ for all $n \geq 1$ and*

$$\sum_{n\geq 1} M_n,$$

converges. Then

$$\sum_{n\geq 1} f_n(\mathbf{x}),$$

is absolutely uniformly convergent on D.

Applying the Weierstrass M-test to arbitrary compact subsets on a region shows that the resulting series is locally absolutely uniformly convergent. This mode of convergence is very useful because it is enough to guarantee the series is analytic on D:

Theorem B.1.2. *Suppose $(f_n(\mathbf{x}))_{n\geq 1}$ is a sequence of analytic functions on a region D. Then if*

$$\sum_{n\geq 1} f_n(\mathbf{x}),$$

is locally absolutely uniformly convergent, it is analytic on D.

We can also apply this idea in the case of integrals. Suppose we have an integral

$$\int_D f(\mathbf{x}, \mathbf{y})\, d\mathbf{y},$$

where $f(\mathbf{x}, \mathbf{y})$ is a complex-valued function on some region $D \times S$ and sis analytic in $\mathbf{x}$ for every $\mathbf{y}$. The integral is a function of $\mathbf{x}$, and we say that the integral is **locally absolutely uniformly convergent** if

$$\int_S |f(\mathbf{x}, \mathbf{y})|\, d\mathbf{y},$$

is uniformly bounded on compact subsets of D. Similar to the series case, this mode of convergence is very useful because it guarantees the integral is analytic on D:

Theorem B.1.3. *Suppose $f(\mathbf{x}, \mathbf{y})$ is a complex-valued function on some region $D \times S$ and is analytic in $\mathbf{x}$ for every $\mathbf{y}$. Then if*

$$\int_S |f(\mathbf{x}, \mathbf{y})|\, d\mathbf{y},$$

is locally absolutely uniformly convergent, it is analytic on D.

B.2 Interchange of Integrals, Sums & Derivatives

Often, we would like to interchange a limit and a integral. This process is not always allowed, but in many instances it is. The **dominated convergence theorem** (DCT) covers the most well-known sufficient condition:

Theorem B.2.1 (Dominated convergence theorem). *Let $(f_n(\mathbf{x}))_{n\geq 1}$ be a sequence of continuous and integrable functions on some region D. Suppose that that the sequence converges pointwise to a function $f(\mathbf{x})$, and that there is some integrable function $g(\mathbf{x})$ on D such that*

$$|f_n(\mathbf{x})| \leq g(\mathbf{x})$$

for all $n \geq 1$ and all $\mathbf{x} \in D$. Then $f(\mathbf{x})$ is integrable on D and

$$\lim_{n\to\infty} \int_D f_n(\mathbf{x})\, d\mathbf{x} = \int_D f(\mathbf{x})\, d\mathbf{x}.$$

This theorem is often employed when the underlying sequence is a sequence of partial sums of an absolutely convergent series (the dominating function $g(\mathbf{x})$ will be the absolute series). In this case we have the following result:

Corollary B.2.1. *Suppose $(f_n(\mathbf{x}))_{n\geq 1}$ is a sequence of continuous functions that are integrable on some region D and*

$$\sum_{n\geq 1} f_n(\mathbf{x}),$$

is absolutely convergent. Then

$$\sum_{n\geq 1} \int_D f_n(\mathbf{x})\, d\mathbf{x} = \int_D \sum_{n\geq 1} f_n(\mathbf{x})\, d\mathbf{x}.$$

More commonly, we would like to interchange two sums, two integrals, or a sum and an integral assuming that one expression is absolutely converges. This is sufficient as given by **Fubini's theorem** (FT):

Theorem B.2.2 (Fubini's theorem). *The following hold:*

(i) If $(f_{n,m}(\mathbf{x}))_{n,m\geq 1}$ is a sequence of continuous functions, then

$$\sum_{(n,m)\in\mathbb{Z}} f_{n,m}(\mathbf{x}) = \sum_{n\geq 1}\sum_{m\geq 1} f_{n,m}(\mathbf{x}) = \sum_{m\geq 1}\sum_{n\geq 1} f_{n,m}(\mathbf{x}),$$

provided any of the expressions are absolutely convergent.

(ii) If $(f_n(\mathbf{x},\mathbf{y}))_n$ is a sequence of continuous and integrable functions on some region $D\times S$, then

$$\int_{D\times S} f(\mathbf{x},\mathbf{y})\,d(\mathbf{x}\times\mathbf{y}) = \int_S\int_D f(\mathbf{x},\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} = \int_D\int_S f(\mathbf{x},\mathbf{y})\,d\mathbf{y}\,d\mathbf{x},$$

provided any of the expressions are absolutely convergent.

(iii) If $(f_n(\mathbf{x}))_{n\geq 1}$ is a sequence of continuous and integrable functions on some region D, then

$$\sum_{n\geq 1}\int_D f_n(\mathbf{x})\,d\mathbf{x} = \int_D \sum_{n\geq 1} f_n(\mathbf{x})\,d\mathbf{x},$$

provided either side is absolutely convergent.

Let $(f_n(\mathbf{x}))_{n\geq 1}$
be a sequence of continuous functions that are integrable on some region D, then

$$\sum_{n\geq 1}\int_D f_n(\mathbf{x})\,d\mathbf{x} = \int_D \sum_{n\geq 1} f_n(\mathbf{x})\,d\mathbf{x},$$

provided either side is absolutely convergent.

Other times we would also like to interchange a derivative and an integral. The **Leibniz integral rule** tells us when this is allowed:

Theorem B.2.3 (Leibniz integral rule). *Suppose $f(\mathbf{x},t)$ is a function on some region $D\times[a(\mathbf{x}),b(\mathbf{x})]$, for some real-valued functions $a(\mathbf{x})$ and $b(\mathbf{x})$, and such that both $f(\mathbf{x},t)$ and its partial derivative $\frac{\partial}{\partial x_i}f(\mathbf{x},t)$ are continuous in $\mathbf{x}$ and t. Also suppose that $a(\mathbf{x})$ and $b(\mathbf{x})$ are continuous with continuous partial derivatives $\frac{\partial}{\partial x_i}a(\mathbf{x})$ and $\frac{\partial}{\partial x_i}b(\mathbf{x})$ for $\mathbf{x}\in D$. Then*

$$\frac{\partial}{\partial x_i}\left(\int_{a(\mathbf{x})}^{b(\mathbf{x})} f(\mathbf{x},t)\,dt\right) = f(\mathbf{x},b(\mathbf{x}))\frac{\partial}{\partial x_i}b(\mathbf{x}) - f(\mathbf{x},a(\mathbf{x}))\frac{\partial}{\partial x_i}a(\mathbf{x}) + \int_{a(\mathbf{x})}^{b(\mathbf{x})}\frac{\partial}{\partial x_i}f(\mathbf{x},t)\,dt,$$

on D.

The Leibniz integral rule is sometimes applied in the case when $a(\mathbf{x}) = a$ and $b(\mathbf{x}) = b$ are constant. In this case, we get the following corollary:

Corollary B.2.2. *Suppose $f(\mathbf{x},t)$ is a function on some region $D\times[a,b]$, for some $a<b$, and such that both $f(\mathbf{x},t)$ and its partial derivative $\frac{\partial}{\partial x_i}f(\mathbf{x},t)$ are continuous in $\mathbf{x}$ and t. Then*

$$\frac{\partial}{\partial x_i}\left(\int_a^b f(\mathbf{x},t)\,dt\right) = \int_a^b \frac{\partial}{\partial x_i}f(\mathbf{x},t)\,dt,$$

on D.

B.3 Summation Formulas

The most well-known summation formula is **partial summation**:

Theorem B.3.1 (Partial summation). *Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be two sequences of complex numbers. Then for any positive integers N and M with $1 \leq M < N$ we have*

$$\sum_{M\leq k\leq N} a_k(b_{k+1} - b_k) = (a_N b_{N+1} - a_M b_M) - \sum_{M+1\leq k\leq N} b_k(a_k - a_{k-1}).$$

There is a more useful summation formula as it lets one estimate discrete sums by integrals. For this we need some notation. If $(a_n)_{n\geq 1}$ is a sequence of complex numbers, for every $X > 0$ set

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

Then **Abel's summation formula** is the following:

Theorem B.3.2 (Abel's summation formula). *Let $(a_n)_{n\geq 1}$ be a sequence of complex numbers. For every X and Y with $0 \leq X < Y$ and continuously differentiable function $\phi : [X, Y] \to \mathbb{C}$, we have*

$$\sum_{X\leq n\leq Y} a_n\phi(n) = A(Y)\phi(Y) - A(X)\phi(X) - \int_X^Y A(u)\phi'(u)\,du.$$

There are also some useful corollaries. For example, if we take the limit as $Y \to \infty$ we obtain:

Corollary B.3.1. *Let $(a_n)_{n\geq 1}$ be a sequence of complex numbers. For every $X \geq 0$ and continuously differentiable function $\phi(y)$, we have*

$$\sum_{n\geq X} a_n\phi(n) = \lim_{Y\to\infty} A(Y)\phi(Y) - A(X)\phi(X) - \int_X^\infty A(u)\phi'(u)\,du.$$

We can take this corollary further by letting $X < 1$ so that $A(X) = 0$ to get the following:

Corollary B.3.2. *Let $(a_n)_{n\geq 1}$ be a sequence of complex numbers. For every continuously differentiable function $\phi(y)$, we have*

$$\sum_{n\geq 1} a_n\phi(n) = \lim_{Y\to\infty} A(Y)\phi(Y) - \int_1^\infty A(u)\phi'(u)\,du.$$

B.4 Factorizations, Order & Rank

The **elementary factors**, also referred to as **primary factors**, are the entire functions $E_n(z)$ defined by

$$E_n(z) = \begin{cases} 1 - z & \text{if } n = 0, \\ (1-z)e^{z+\frac{z^2}{2}+\cdots+\frac{z^n}{n}} & \text{if } n \neq 0. \end{cases}$$

If $f(z)$ is an entire function, then it admits a factorization in terms of its zeros and the elementary factors. This is called the **Weierstrass factorization** of $f(z)$:

Theorem B.4.1 (Weierstrass factorization). *Let $f(z)$ be an entire function with $\{a_n\}_{n\geq 1}$ the nonzero zeros of $f(z)$ counted with multiplicity. Also suppose that $f(z)$ has a zero of order m at $z = 0$ where it is understood that if $m = 0$ we mean $f(0) \neq 0$ and if $m < 0$ we mean $f(z)$ has a pole of order $|m|$ at $z = 0$. Then there exists an entire function $g(z)$ and sequence of nonnegative integers $(p_n)_{n\geq 1}$ such that*

$$f(z) = z^m e^{g(z)} \prod_{n\geq 1} E_{p_n}\left(\frac{z}{a_n}\right).$$

The Weierstrass factorization of $f(z)$ can be strengthened if $f(z)$ does not grow too fast. We say $f(z)$ is of **finite order** if there exists a $\rho_0 > 0$ such that

$$f(z) \ll e^{|z|^{\rho_0}},$$

for all $z \in \mathbb{C}$. The **order** ρ of $f(z)$ is the infimum of the ρ_0. Let $q = \lfloor \rho \rfloor$. If there is no such ρ_0, $f(z)$ is said to be of **infinite order** and we set $\rho = q = \infty$. Let $\{a_n\}_{n\geq 1}$ be the nonzero zeros of $f(z)$ that are not zero and ordered such that $a_n \to \infty$ as $n \to \infty$ if there are infinitely many zeros. Then we define the **rank** of $f(z)$ to be the smallest positive integer p such that the series

$$\sum_{n\geq 1} \frac{1}{|a_n|^{p+1}},$$

converges. If there is no such integer we set $p = \infty$ and if there are finitely many zeros we set $p = 0$. We set $g = \max(p, q)$ and call g the **genus** of $f(z)$. We can now state the **Hadamard factorization** of $f(z)$:

Theorem B.4.2 (Hadamard factorization). *Let $f(z)$ be an entire function of finite order ρ. If p is the rank and g is the genus, then $g \leq \rho$. Moreover, let $\{a_n\}_{n\geq 1}$ be the nonzero zeros of $f(z)$ counted with multiplicity and suppose that $f(z)$ has a zero of order m at $z = 0$ where it is understood that if $m = 0$ we mean $f(0) \neq 0$ and if $m < 0$ we mean $f(z)$ has a pole of order $|m|$ at $z = 0$. Then there exists a polynomial $Q(z)$ of degree at most q such that*

$$f(z) = z^m e^{Q(z)} \prod_{n\geq 1} E_p\left(\frac{z}{a_n}\right).$$

Moreover, the sum

$$\sum_{n\geq 1} \frac{1}{|a_n|^{\rho+\varepsilon}},$$

converges.

B.5 The Phragmén-Lindelöf Convexity Principle

The **Phragmén-Lindelöf convexity principle** is a generic name for extending the maximum modulus principle to unbounded regions. The **Phragmén-Lindelöf convexity principle** for vertical strips is the case when the unbounded region is the vertical strip $a < \sigma < b$:

Theorem B.5.1 (Phragmén-Lindelöf convexity principle, vertical strip version). *Let $f(s)$ be a holomorphic function on an open neighborhood of the vertical strip $a < \sigma < b$ such that $f(s) \ll e^{|s|^A}$ for some $A \geq 0$. Then the following hold:*

(i) If $|f(s)| \leq M$ for $\sigma = a, b$, that is on the boundary edges of the strip, then $|f(s)| \leq M$ for all s in the strip.

(ii) Assume that there is a continuous function $g(t)$ such that

$$f(a+it) \ll g(t)^{\alpha} \quad \text{and} \quad f(b+it) \ll g(t)^{\beta},$$

for all $t \in \mathbb{R}$. Then

$$f(s) \ll g(t)^{\alpha\ell(\sigma)+\beta(1-\ell(\sigma))},$$

where ℓ is the linear function such that $\ell(a) = 1$ and $\ell(b) = 0$.

We will also need a variant. The **Phragmén-Lindelöf convexity principle** for vertical half-strips is the case when the unbounded region is the vertical half-strip $a < \sigma < b$ with $t > c$:

Theorem B.5.2 (Phragmén-Lindelöf convexity principle, vertical half-strip version). *Let $f(s)$ be a holomorphic function on an open neighborhood of the vertical strip $a < \sigma < b$ with $t > c$ such that $f(s) \ll e^{|s|^A}$ for some $A \geq 0$. Then the following hold:*

(i) If $|f(s)| \leq M$ for $\sigma = a, b$ with $t \geq c$ and $t = c$ with $a \leq \sigma \leq b$, that is on the boundary edges of the half-strip, then $|f(s)| \leq M$ for all s in the strip.

(ii) Assume that there is a continuous function $g(t)$ such that

$$f(a+it) \ll g(t)^{\alpha} \quad \text{and} \quad f(b+it) \ll g(t)^{\beta},$$

for all $t \geq c$. Then

$$f(s) \ll g(t)^{\alpha\ell(\sigma)+\beta(1-\ell(\sigma))},$$

where ℓ is the linear function such that $\ell(a) = 1$ and $\ell(b) = 0$.

B.6 Bessel Functions

For any $\nu \in \mathbb{C}$, the **Bessel equation** is the ODE

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} + (z^2 - \nu^2)w = 0.$$

There are two linearly independent solutions to this equation. One solution is the **Bessel function of the first kind** $J_\nu(z)$ defined by

$$J_\nu(z) = \sum_{n \geq 0} \frac{(-1)^n}{n!\Gamma(n+\nu+1)}\left(\frac{z}{2}\right)^{2n+\nu}.$$

For integers n, $J_n(z)$ is entire and we have

$$J_n(z) = (-1)^n J_{-n}(z).$$

Otherwise, $J_\nu(z)$ has a pole at $z = 0$ and $J_\nu(z)$ and $J_{-\nu}(z)$ are linearly independent solutions to the Bessel equation. The other solution is the **Bessel function of the second kind** $Y_\nu(z)$ defined by

$$Y_\nu(z) = \frac{J_\nu(z)\cos(\nu\pi) - J_{-\nu}(z)}{\sin(\nu\pi)},$$

for non-integers ν, and for integers n is

$$Y_n(z) = \lim_{\nu \to n} Y_\nu(z).$$

For any integer n, we also have

$$Y_n(z) = (-1)^n Y_{-n}(z).$$

For the J-Bessel function there is also an important integral representation called the **Schläflin integral representation**:

Proposition B.6.1 (Schläflin integral representation for the J-Bessel function). *For any $\nu \in \mathbb{C}$,*

$$J_\nu(z) = \frac{1}{2\pi i}\left(\frac{z}{2}\right)^\nu \int_{-\infty}^{(0^+)} t^{-(\nu+1)} e^{t-\frac{z^2}{4t}}\, dt,$$

provided $|\arg(z)| < \frac{\pi}{2}$ and where the contour is the Hankel contour about the negative real axis.

The **modified Bessel equation** is the ODE

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} - (z^2+\nu^2)w = 0.$$

Like the Bessel equation, there are two linearly independent solutions. One solution is the **modified Bessel function of the first kind** $I_\nu(x)$ given by

$$I_\nu(z) = i^{-\nu}J_\nu(iz) = \sum_{n\geq 0}\frac{1}{n!\Gamma(n+\nu+1)}\left(\frac{z}{2}\right)^{2n+\nu}.$$

For integers n, this solution is symmetric in n. That is,

$$I_n(z) = I_{-n}(z).$$

We also have a useful integral representation in a half-plane:

Proposition B.6.2. *For any $\nu \in \mathbb{C}$,*

$$I_\nu(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos(t)}\cos(\nu t)\, dt - \frac{\sin(\nu\pi)}{\pi}\int_0^\infty e^{-z\cosh(t)-\nu t}\, dt,$$

provided $|\arg(z)| < \frac{\pi}{2}$.

From this integral representation we can show the following asymptotic:

Lemma B.6.1. *For any $\nu \in \mathbb{C}$,*

$$I_\nu(z) \sim_\varepsilon \sqrt{\frac{1}{2\pi z}}e^z,$$

provided $|\arg(z)| \leq \frac{\pi}{2} - \varepsilon$ for some $\varepsilon > 0$.

The other solution is the **modified Bessel function of the second kind** $K_\nu(z)$ defined by

$$K_\nu(z) = \frac{\pi}{2}\frac{I_{-\nu}(z) - I_\nu(z)}{\sin(\nu\pi)},$$

for non-integers ν, and for integers n is

$$K_n(z) = \lim_{\nu \to n} K_\nu(z).$$

This is one of the more important types of Bessel functions as they appear in the Fourier coefficients of certain Eisenstein series. This function is symmetric in ν even when ν is an integer. That is,

$$K_\nu(z) = K_{-\nu}(z),$$

for all ν. We also have a very useful integral representation in a half-plane:

Proposition B.6.3. *For any $\nu \in \mathbb{C}$,*

$$K_\nu(z) = \int_0^\infty e^{-z\cosh(t)} \cosh(\nu t)\, dt,$$

provided $|\arg(z)| < \frac{\pi}{2}$.

From this integral representation it does not take much to show the following asymptotic:

Lemma B.6.2. *For any $\nu \in \mathbb{C}$,*

$$K_\nu(z) \sim_\varepsilon \sqrt{\frac{\pi}{2z}} e^{-z},$$

provided $|\arg(z)| \leq \frac{\pi}{2} - \varepsilon$ for some $\varepsilon > 0$.

B.7 Whittaker Functions

For $\kappa, \mu \in \mathbb{C}$, the **Whittaker equation** is the ODE

$$\frac{dw}{dz} + \left(\frac{1}{4} - \frac{\kappa}{z} - \frac{\frac{1}{4} - \mu^2}{z^2}\right) w = 0.$$

There are two linearly independent solutions to this equation provided $-2\mu \notin \mathbb{Z}_{\geq 1}$. If we additionally assume that $w(z) = o(e^{2\pi y})$ as $y \to \infty$, then there is only one linearly independent solution. This solution is the **Whittaker function** $W_{\kappa,\mu}(z)$. It can be expressed in the form

$$W_{\kappa,\mu}(z) = z^{\mu+\frac{1}{2}} e^{-\frac{z}{2}} U\left(\mu - \kappa + \frac{1}{2}, 1 + 2\mu, z\right),$$

where $U(\alpha, \beta, z)$ is the **confluent hypergeometric function** initially defined by

$$U(\alpha, \beta, z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-zu} u^{\alpha-1} (1+u)^{\beta-\alpha-1}\, du,$$

for $\mathrm{Re}(\alpha) > 0$ and $x > 0$ and then analytically continued to $\mathbb{C}^3$. From this integral representation we can show the following asymptotic:

Lemma B.7.1. *For any $\kappa, \mu \in \mathbb{C}$ with $-2\mu \notin \mathbb{Z}_{\geq 1}$,*

$$W_{\kappa,\mu}(z) \sim_\varepsilon z^\kappa e^{-\frac{z}{2}},$$

provided $|\arg(z)| \leq \frac{\pi}{2} - \varepsilon$ for some $\varepsilon > 0$.

The Whittaker function also satisfies the important symmetry properties

$$W_{\kappa,\mu}(z) = W_{\kappa,-\mu}(z) \quad \text{and} \quad \overline{W_{\kappa,\mu}(z)} = W_{\kappa,-\mu}(\overline{z}).$$

In particular, $W_{\kappa,\mu}(z)$ is conjugate symmetric if z is real. The Whittaker function also has a simplified form in special cases:

Theorem B.7.1. *For any $\nu, \alpha \in \mathbb{C}$ with $-2\nu \notin \mathbb{Z}_{\geq 1}$,*

$$W_{0,\nu}(z) = \left(\frac{z}{\pi}\right)^{\frac{1}{2}} K_\nu\left(\frac{z}{2}\right) \quad \text{and} \quad W_{\alpha,\alpha-\frac{1}{2}}(z) = z^\alpha e^{-\frac{z}{2}}.$$

B.8 Lattice Sums

Consider $\mathbb{R}^d$ with the standard inner product $\langle \cdot, \cdot \rangle$ and norm $|| \cdot ||$. We are often interested in series that are obtained by summing over the lattice $\mathbb{Z}^d \subset \mathbb{R}^d$. In particular, we have the following general result:

Theorem B.8.1. *Let $d \geq 1$ be an integer. Then*

$$\sum_{\mathbf{a} \in \mathbb{Z}^d - \{\mathbf{0}\}} \frac{1}{||\mathbf{a}||^s},$$

is locally absolutely uniformly convergent in the region $\sigma > d$.

In a practical setting, we usually restrict to the case $d = 2$. In this setting, with a little more work can show a more useful result:

Proposition B.8.1. *Let $z \in \mathbb{H}$. Then*

$$\sum_{(n,m) \in \mathbb{Z}^2 - \{\mathbf{0}\}} \frac{1}{|nz + m|^s},$$

is locally absolutely uniformly convergent in the region $\sigma > 2$. In addition, it is locally absolutely uniformly convergent as a function of z provided $\sigma > 2$.

Appendix C

Algebra

C.1 Finitely Generated Abelian Groups

Let V be a finitely generated abelian group. The following result is the **structure theorem for finitely generated abelian groups**

Theorem C.1.1 (Structure theorem for finitely generated abelian groups). *Let V be a finitely generated abelian group. Then*

$$V \cong \mathbb{Z}^n \oplus \mathbb{Z}/k_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/k_t\mathbb{Z},$$

where k_i are integers such that $k_i \mid k_{i+1}$ for $1 \leq i \leq t-1$.

Let V be a free abelian group of rank n and let $W \leq V$ be a subgroup that is also a free abelian group of rank n. Let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_n\}$ be bases for V and W respectively so that

$$V = \mathbb{Z}v_1 + \cdots + \mathbb{Z}v_n \quad \text{and} \quad W = \mathbb{Z}w_1 + \cdots + \mathbb{Z}w_n.$$

Write

$$w_i = \sum_{1 \leq j \leq n} w_{i,j} v_j,$$

so that $B = (W_{i,j})_{i,j}$ is the base change matrix from $\{v_1, \dots, v_n\}$ to $\{w_1, \dots, w_n\}$. A consequence of the Smith normal form is that the quotient group is given by the absolute value of the determinant:

Proposition C.1.1. *Let V be a free abelian group of rank n and let $W \leq V$ be a subgroup that is also a free abelian group of rank n. Then $|V/W|$ is finite and*

$$|V/W| = |\det(B)|,$$

for any base change matrix B from a basis of V to a basis of W.

C.2 Character Groups

For any finite abelian group G, a **character** φ is a homomorphism $\varphi : G \to \mathbb{C}$. They form a group, denoted $\widehat{G}$, under multiplication called the **character group** of G. If G is an additive group, we say that any $\varphi \in \Gamma$ is a **additive character**. Similarly, if G is a multiplicative group, we say that any $\varphi \in \Gamma$ is a **multiplicative character**. In any case, if $|G| = n$ then $\varphi(g)^n = \varphi(g^n) = 1$ so that φ takes values in the n-th roots of unity. Moreover, to every character φ there is its **conjugate character** $\overline{\varphi}$ defined by

$\overline{\varphi}(g) = \overline{\varphi(a)}$. Clearly the conjugate character is also a character. Since φ takes its value in the roots of unity, $\overline{\varphi(a)} = \varphi(a)^{-1}$ so that $\overline{\varphi} = \varphi^{-1}$. One of the central theorems about characters is that the character group of G is isomorphic to G:

Proposition C.2.1. *Any finite abelian group G is isomorphic to its character group. That is,*

$$G \cong \widehat{G}.$$

The characters also satisfy certain **orthogonality relations**:

Proposition C.2.2 (Orthogonality relations). *Let G be a finite abelian group.*

(i) For any two characters χ and ψ of G,

$$\frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\psi}(g) = \delta_{\chi,\psi}.$$

(ii) For any $g, h \in G$,

$$\frac{1}{|G|}\sum_{\chi\in\widehat{G}}\chi(g)\overline{\chi}(h) = \delta_{g,h}.$$

C.3 Representation Theory

Let G be a group and V be a vector space over a field $\mathbb{F}$. A **representation** (ρ, V), or just ρ if the underlying vector space V is clear, of G on V is a map

$$\rho : G \times V \to V \qquad (g, v) \mapsto \rho(g, v) = g \cdot v,$$

such that the following properties are satisfied:

1. For any $g \in G$, the map
$$\rho : V \to V \qquad v \mapsto g \cdot v,$$
is linear.

2. For any $g, h \in G$ and $v \in V$,
$$1 \cdot v = v \quad \text{and} \quad g \cdot (h \cdot v) = (gh) \cdot v.$$

Therefore ρ defines an action of G on V. An equivalent definition of a representation of G on V is a homomorphism from G into $\mathrm{Aut}(V)$. By abuse of notation, we also denote this homomorphism by ρ. If the dimension of V is n, then (ρ, V) is sai to be an **n-dimensional**. We say that (ρ, W) is a **subrepresentation** of (V, ρ) if $W \subseteq V$ is a G-invariant subspace. In particular, (ρ, W) is a representation itself. Lastly, if (ρ_1, V_1) and (ρ_2, V_2) are two representations, we can form the **direct sum representation** $(\rho_1 \oplus \rho_2, V_1 \oplus V_2)$ where $\rho_1 \oplus \rho_2$ acts diagonally on $V_1 \oplus V_2$. A natural question to ask is how representation can be decomposed as a direct sum of other representations. We say (ρ, V) is **irreducible** if it contains no proper G-invariant subspaces and is **completely irreducible** if it decomposes as a direct sum of irreducible subrepresentations.

We will only need one very useful theorem about representations when G is a finite abelian group and V is a vector space over $\mathbb{C}$. In this case G has a group of characters $\widehat{G}$, and the underlying vector space V is completely reducible with respect to the characters of G:

Theorem C.3.1. *Let V be a vector space over $\mathbb{C}$ and let Φ be a representation of a group G on V. If G is a finite abelian group, then*

$$V = \bigoplus_{\chi \in \widehat{G}} V_\chi,$$

where

$$V_\chi = \{v \in V : g \cdot v = \chi(g)v \text{ for all } g \in G\}.$$

In particular, V is completely reducible and every irreducible subrepresentation is 1-dimensional.

Appendix D

Topology

D.1 Fundamental Domains

Let G be a group acting on a connected Hausdorff space X by automorphisms. Then we can consider the quotient space X/G of X by the action of G. We would like the quotient X/G to inherit properties of G. For this, we require G to satisfy two properties. The first is the G is a discrete group. For the second, we say G acts **properly discontinuously** if for every $x \in X$ there exists a neighborhood U_x of x such that the intersection $gU_x \cap U_x$ is empty for all $g \in G$ unless g is the identity. In the case G is discrete and acts properly discontinuously, the quotient X/G inherits nice topological properties:

Proposition D.1.1. *Let G be a group acting on a a connected Hausdorff space X by automorphisms. If G is discrete and acts properly discontinuously, then X/G is connected Hausdorff.*

In the case of Proposition D.1.1, we often want to consider a useful set of representatives of X/G. A **fundamental domain** for X/G is a closed subset $D \subseteq X$ satisfying the following conditions:

(i) Any point in X is G-equivalent to a point in D.

(ii) If two points in D are G-equivalent via a non-identity element, then they lie on the boundary of D.

(iii) The interior of D is a domain.

In other words, D is a complete set of representatives (possibly with overlap on the boundary) for X/G that has a nice topological structure with respect to X. Note that if D is a fundamental domain then so is gD for any $g \in G$ and moreover

$$X = \bigcup_{g \in G} gD.$$

In particular, the choice of fundamental domain is not unique. Intuitively, a fundamental domain is a geometric realization of X/G which is often more fruitful than thinking of X/G as an abstract set of equivalence classes. Indeed, if X is a subset of $\mathbb{R}^n$ or $\mathbb{C}^n$ then property (iii) ensures that we can integrate over D.

Appendix E

Miscellaneous

E.1 Special Integrals

Below is a table of well-known integrals that are used throughout the text:

Reference	Assumptions	Integral
Gaussian		$\displaystyle\int_{\mathbb{R}} e^{-\pi x^2}\, dx = 1$
[Gol06]	$s, \nu \in \mathbb{C}, \mathrm{Re}(s+\nu) > -1$	$\displaystyle\int_0^\infty K_\nu(y) y^s \,\frac{dy}{y} = 2^{s-2}\Gamma\left(\frac{s+\nu}{2}\right)\Gamma\left(\frac{s-\nu}{2}\right)$
[Gol06]	$n \in \mathbb{Z},\ s \in \mathbb{C},\ y > 0$	$\displaystyle\int_{\mathbb{R}} \frac{e^{-2\pi i n x y}}{(x^2+1)^s}\, dx = \begin{cases} \frac{\sqrt{\pi}\Gamma\left(s-\frac{1}{2}\right)}{\Gamma(s)} & \text{if } n = 0, \\ \frac{2\pi^s \lvert n\rvert^{s-\frac{1}{2}} y^{s-\frac{1}{2}}}{\Gamma(s)} K_{s-\frac{1}{2}}(2\pi\lvert n\rvert y) & \text{if } n \neq 0. \end{cases}$
[Dav80]	$c > 0$	$\displaystyle\frac{1}{2\pi i}\int_{(c)} y^s \,\frac{ds}{s} = \begin{cases} 0 & \text{if } 0 < y < 1, \\ \frac{1}{2} & \text{if } y = 1, \\ 1 & \text{if } y > 1. \end{cases}$

Index

Bibliography

[BB13] Valentin Blomer and Farrell Brumley. The role of the ramanujan conjecture in analytic number theory. *Bulletin of the American Mathematical Society*, 50(2):267–320, 2013.

[Bum97] Daniel Bump. *Automorphic Forms and Representations*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1997. Citation: Bump1997automorphic.

[CFK+05] J Brian Conrey, David W Farmer, Jon P Keating, Michael O Rubinstein, and Nina C Snaith. Integral moments of l-functions. *Proceedings of the London Mathematical Society*, 91(1):33–104, 2005.

[Dav80] Harold Davenport. *Multiplicative Number Theory*, volume 74 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1980.

[DB15] Lokenath Debnath and Dambaru Bhatta. *Integral transforms and their applications*. CRC Press/Taylor & Francis Group, Boca Raton, third edition edition, 2015.

[Del71] Pierre Deligne. Formes modulaires et représentations e-adiques. In *Séminaire Bourbaki vol. 1968/69 Exposés 347-363*, pages 139–172, Berlin, Heidelberg, 1971. Springer.

[Del74] Pierre Deligne. La conjecture de Weil. I. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 43(1):273–307, December 1974.

[DFI02] William Duke, John B Friedlander, and Henryk Iwaniec. The subconvexity problem for artin l–functions. *Inventiones mathematicae*, 149:489–577, 2002.

[DS05] Fred Diamond and Jerry Shurman. *A First Course in Modular Forms*, volume 228 of *Graduate Texts in Mathematics*. Springer, New York, NY, 2005.

[Dys62] Freeman J Dyson. Statistical theory of the energy levels of complex systems. i. *Journal of Mathematical Physics*, 3(1):140–156, 1962.

[Eul44] Leonhard Euler. Variae observationes circa series infinitas. *Commentarii academiae scientiarum Petropolitanae*, pages 160–188, January 1744.

[Eva22] Lawrence C. Evans. *Partial differential equations*. Number 19 in Graduate studies in mathematics. American Mathematical Society, Providence, Rhode Island, second edition edition, 2022.

[G+08] Loukas Grafakos et al. *Classical fourier analysis*, volume 2. Springer, 2008.

[Gau08] Carl Friedrich Gauss. Summatio quarumdam serierum singularium. *Gottingae (Dieterich)*, 1808.

[Gol06] Dorian Goldfeld. *Automorphic Forms and L-Functions for the Group* $\mathrm{GL}(n, \mathbb{R})$. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2006.

[Gra03] J. P. Gram. Note sur les zéros de la fonction $\zeta(s)$ de Riemann. *Acta Mathematica*, 27(0):289–304, 1903.

[Had96] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, 24:199–220, 1896.

[HL16] G. H. Hardy and J. E. Littlewood. Contributions to the theory of the riemann zeta-function and the theory of the distribution of primes. *Acta Mathematica*, 41(none):119–196, January 1916.

[HL21] G. Hardy and J. E. Littlewood. The zeros of riemann's zeta-function on the critical line. *Mathematische Zeitschrift*, 10:283–317, 1921. Citation: Hardy1921zeros.

[HL23] G. H. Hardy and J. E. Littlewood. The approximate functional equation in the theory of the zeta-function, with applications to the divisor-problems of dirichlet and piltz. *Proceedings of the London Mathematical Society*, s2-21(1):39–74, 1923. Citation: Hardy1923approximate.

[HL29] G. H. Hardy and J. E. Littlewood. The approximate functional equations for $\zeta(s)$ and $\zeta^2(s)$. *Proceedings of the London Mathematical Society*, s2-29(1):81–97, 1929. Citation: Hardy1929approximate.

[Ing28] Albert Edward Ingham. Mean-value theorems in the theory of the riemann zeta-function. *Proceedings of the London Mathematical Society*, 2(1):273–300, 1928.

[Iwa02] Henryk Iwaniec. *Spectral Methods of Automorphic Forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, November 2002.

[Kil15] L. J. P. Kilford. *Modular forms: a classical and computational introduction*. Imperial College Press, Hackensack, NJ, 2nd edition edition, 2015.

[Koc01] Helge Koch. Sur la distribution des nombres premiers. *Acta Mathematica*, 24(0):159–182, 1901.

[KRS03] Henry H. Kim, Dinakar Ramakrishnan, and Peter Sarnak. Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2. *Journal of the American Mathematical Society*, 16(1):139–183, 2003.

[KS00] Jon P Keating and Nina C Snaith. Random matrix theory and ζ (1/2+ it). *Communications in Mathematical Physics*, 214:57–89, 2000.

[KS23] Nicholas M Katz and Peter Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45. American Mathematical Society, 2023.

[Lan94] Serge Lang. *Algebraic Number Theory*, volume 110 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1994.

[Leg98] A. M. (Adrien Marie) Legendre. *Essai sur la théorie des nombres*. Paris, Duprat, 1798.

[Leg08] Adrien Marie Legendre. *Essai sur la theorie des nombres; par A.M. Legendre, membre de l'Institut et de la Legion d'Honneur ..* chez Courcier, imprimeur-libraire pour les mathematiques, quai des Augustins, n° 57, 1808.

[Odl92] Andrew M Odlyzko. The 1020-th zero of the riemann zeta function and 175 million of its neighbors. *preprint*, 512, 1992.

[Pou97] Charles Jean de La Vallée Poussin. *Recherches analytiques sur la théorie des nombres premiers*. Hayez, 1897.

[Ram16] Srinivasa Ramanujan. On certain arithmetical functions. *Transactions of the Cambridge Philosophical Society*, 22(9):159 – 184, 1916.

[Rem98] Reinhold Remmert. *Classical Topics in Complex Function Theory*, volume 172 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1998.

[Rie59] Bernhard Riemann. Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse. *Monatsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, pages 671–680, 1859.

[RS62] J. Barkley Rosser and Lowell Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6(1):64–94, March 1962.

[SSS03] Elias M. Stein, Rami Shakarchi, and Elias M. Stein. *Complex analysis*. Number 2 in Princeton lectures in analysis / Elias M. Stein & Rami Shakarchi. Princeton University Press, Princeton Oxford, 2003.

[Wei48] Andre Weil. On Some Exponential Sums. *Proceedings of the National Academy of Sciences of the United States of America*, 34(5):204–207, 1948.