Quantum Limits, Counting and Landau-type Formulae in Hyperbolic Space

Niko Petter Johannes Laaksonen

UCL

January 27, 2016

Submitted in partial fulfilment of the requirements
for the Research Degree in Mathematics, PhD

I, Niko Petter Johannes Laaksonen confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this thesis we explore a variety of topics in analytic number theory and automorphic forms. In the classical context, we look at the value distribution of two Dirichlet L-functions in the critical strip and prove that for a positive proportion these values are linearly independent over the real numbers. The main ingredient is the application of Landau's formula with Gonek's error term. The remainder of the thesis focuses on automorphic forms and their spectral theory. In this setting we explore three directions. First, we prove a Landau-type formula for an exponential sum over the eigenvalues of the Laplacian in $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2$ by using the Selberg Trace Formula. Next, we look at lattice point problems in three dimensions, namely, the number of points within a given distance from a totally geodesic hyperplane. We prove that the error term in this problem is $O(X^{3/2})$, where $\operatorname{arccosh} X$ is the hyperbolic distance to the hyperplane. An application of large sieve inequalities provides averages for the error term in the radial and spatial aspect. In particular, the spatial average is consistent with the conjecture that the pointwise error term is $O(X^{1+\epsilon})$. The radial average is an improvement on the pointwise bound by $1/6$. Finally, we identify the quantum limit of scattering states for Bianchi groups of class number one. This follows as a consequence of studying the Quantum Unique Ergodicity of Eisenstein series at complex energies.

Contents

Abstract 3

Acknowledgements 8

Notation 9

1 Introduction **10**

1.1 Summary of Results 13

I Value Distribution of Dirichlet L-functions **19**

2 Analytic Theory of $\zeta(s)$ and $L(s,\chi)$ **20**

2.1 The History 20

2.2 Arithmetic Functions 24

2.3 Dirichlet L-functions 27

2.4 Convexity and Subconvexity 29

3 Discrete Mean Values of Dirichlet L-functions **32**

3.1 Proof of Theorem 3.2 35

3.2 Proof of Theorem 3.3 49

II Automorphic Forms 58

4 Spectral Theory in Hyperbolic Space 59

4.1 Historical Motivation 59

4.2 Hyperbolic Plane 61

4.3 Dedekind Zeta Function 72

4.4 Hyperbolic 3-space 74

5 Landau-type Formulae for Exponential Sums 81

5.1 Numerical Discussion 83

5.2 Application of the Selberg Trace Formula 91

6 Lattice Point Problems 96

6.1 The Geometric Setup of the Problem 100

6.2 Spectral Analysis 106

6.3 Applications of the Large Sieve 121

7 Quantum Unique Ergodicity and Eisenstein Series 135

7.1 A Brief Introduction to Quantum Chaos 135

7.2 Proof of Theorems 7.3 and 7.4 143

A Gamma Function 154

B Radial Average of the Standard Lattice Point Problem in $\mathbb{H}^3$ 156

Bibliography 160

List of Figures

5.1 $\Sigma(T,X)$ in terms of X for $X \in [3,10]$. 85

5.2 $\Sigma(T,X)$ in terms of X for $X \in [13,18]$. 86

5.3 $\Sigma(T,X)$ in terms of X for $X \in [21,27]$. 86

5.4 $\mathrm{Re}\,\Sigma(T,X)$ in terms of X for $X \in [100,10\,000]$. 87

5.5 $\Sigma(T,X)$ for $X \in [10,100]$. 88

5.6 $\Sigma(T,X)$ at $X = 61.98$. 89

5.7 $\Sigma(T,X)$ at $X = 49$. 89

5.8 $\Sigma(T,X)$ at different values of X. 90

5.9 Different normalisations of $\widetilde{\Sigma}(T,X)$ at $X = 49$. 90

5.10 $T^{-1}Z(T,x)$ for $x \in [1.5,30]$. 91

6.1 The (x,u,v) coordinates in $\mathscr{H}$. 102

6.2 Fundamental domain of H in $\mathscr{H}$ with the part for $H\backslash\mathscr{P}$ highlighted (for $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$). 105

7.1 Trajectories in an elliptic billiard. 135

7.2 The Bunimovich Stadium. 136

Life, with its rules, its obligations, and its freedoms, is like a sonnet: You're given the form, but you have to write the sonnet yourself.

— Madeleine L'Engle, A Wrinkle In Time

Acknowledgements

First and foremost I wish to thank my parents for their unconditional support and understanding throughout my studies. It has been a tough path at times, but the fact that I could always rely on them has helped me see it through to the conclusion. I hope I can convey at least some of the deep appreciation I have.

I owe many things in my life to my supervisor Yiannis Petridis. Most of my mathematical education has been influenced by him all the way from the first year Analysis 1 lectures, to my master's project and finally the four years of the doctoral research that culminated in this thesis. I am immensely grateful for his advice, both academic and otherwise, and for the substantial time and effort he has always devoted to me. I also have him to thank for inducting me into the world of coffee, pastries and polar bears.

Of course my life in London would not have been the same without all my friends during my undergraduate and the KLB crew in the latter years. The times spent at the East End were a welcome delight to the otherwise sometimes monotonous work of a teaching assistant.

I would especially like to thank Pietro and Ross for putting up with me for eight years. It has been an incredible experience developing from clueless freshers into even more clueless graduate students. I could not have hoped for better friends to share this journey with.

Notation

Throughout this thesis we let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote the set of natural numbers, integers, rational numbers, real numbers and complex numbers, respectively. We write $\mathbb{R}^+$ and $\mathbb{R}_{\geq 0}$ for the set of positive and non-negative real numbers, respectively, and similarly for $\mathbb{Z}$ and $\mathbb{Q}$. We adopt the convention that $0 \notin \mathbb{N}$. For any ring R we write $R^\times$ for the multiplicative group of units of R. For the most part $s \in \mathbb{C}$ is written as $s = \sigma + it$, so $\operatorname{Re} s = \sigma$ and $\operatorname{Im} s = t$. Similarly for $z \in \mathbb{C}$ we usually have $z = x + iy$. For $n, m \in \mathbb{N}$ their greatest common divisor is denoted by (n, m). We denote by $\Gamma(z)$ the Gamma function, whose definition and a list of useful properties are given in Appendix A. The non-trivial zeros of the Riemann zeta function are denoted by $\rho = \beta + i\gamma$. We use the standard Big O notation. Let f and g be real-valued functions. We say that

$$f(x) = O(g(x)),$$

as $x \to \infty$, if there are constants $M \in \mathbb{R}^+$ and $x_0 \in \mathbb{R}$ such that

$$|f(x)| \leq M|g(x)|, \qquad \forall\, x > x_0.$$

We also use the equivalent notation $f(x) \ll g(x)$. Sometimes we indicate that the constant M depends on a parameter, say ϵ, with a subscript $O_\epsilon(\cdot)$ or $\ll_\epsilon$. We also define

$$f(x) = o(g(x)) \qquad \Longleftrightarrow \qquad \lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

Two functions$f$ and g are asymptotically equal, $f(x) \sim g(x)$, if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

It is also useful to define the more general notion of $f(x) \asymp g(x)$ if

$$\exists x_0 \in \mathbb{R}, \exists m, M \in \mathbb{R}^+ \quad \text{such that} \quad m g(x) \leq f(x) \leq M g(x), \quad \forall\, x > x_0.$$

Let $\mathrm{SL}_2(R)$ denote the special linear group of 2×2 matrices of determinant one over a ring R. We define $\mathrm{PSL}_2(R) = \mathrm{SL}_2(R)/\{\pm I\}$ where I is the identity matrix. This is called the projective special linear group.

Chapter 1

Introduction

Analytic number theory is the study of arithmetic objects through the techniques of mathematical analysis. For prime numbers the powerful tools of complex analysis can be used to give beautiful and elementary proofs, such as Newman's short proof of the Prime Number Theorem (PNT) [105]. The PNT says that if $\pi(x)$ is the number of primes less than x, then

$$\pi(x) \sim \mathrm{li}(x), \qquad \text{as } x \to \infty,$$

where $\mathrm{li}(x)$ is the logarithmic integral,

$$\mathrm{li}(x) = \int_2^x \frac{dt}{\log t}.$$

A fundamental question is that of understanding the distribution of primes, or other arithmetic objects, in the encompassing space. The underlying connection between this idea and the work presented in this thesis is provided by the Riemann zeta function. The Riemann zeta function is given for $\mathrm{Re}\, s > 1$ by

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

In the first part of this thesis we investigate questions of value distribution for Dirichlet L-functions. These are given for $\mathrm{Re}\, s > 1$ by

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

where $\chi:(\mathbb{Z}/q\mathbb{Z})^\times \longrightarrow \mathbb{C}^\times$ is a group homomorphism extended to $\mathbb{Z}$ with $\chi(n)=0$ if n and q are not coprime. The Riemann zeta function (and $L(s,\chi)$ by analogy) is inherently connected to the study of prime numbers through the Euler product, which is given for $\operatorname{Re} s>1$ by a product defined over all primes p,

$$\zeta(s)=\prod_p \frac{1}{1-p^{-s}}. \tag{1.1}$$

The Euler product immediately implies that $\zeta(s)$ is non-zero for $\operatorname{Re} s>1$. In fact, the most famous conjecture in analytic number theory concerns the zeros of the Riemann zeta function. This is the Riemann Hypothesis (RH). Apart from the trivial zeros of $\zeta(s)$ in $\operatorname{Re} s<0$, it turns out that there are infinitely many zeros in the *critical strip*, $0\leq \operatorname{Re} s\leq 1$. Hence, let us define

$$\mathscr{Z}=\{s\in\mathbb{C} : \zeta(s)=0,\, 0\leq \operatorname{Re} s\leq 1\}$$

as the set of *non-trivial zeros* of ζ.

Riemann Hypothesis. *All non-trivial zeros of $\zeta(s)$ lie on the critical line* $\operatorname{Re} s=1/2$, *that is,*

$$\mathscr{Z}\subset\left\{s\in\mathbb{C} : \operatorname{Re} s=\tfrac{1}{2}\right\}.$$

It is a highly surprising fact that the Riemann Hypothesis is actually equivalent to having the best possible error term $O(x^{1/2+\epsilon})$ in the Prime Number Theorem (after a slight reformulation [76, Theorem 5.8]). It is therefore of great importance to understand the value distribution of $\zeta(s)$ and $L(s,\chi)$ in the critical strip. This is the topic of Part I.

In Part II we study automorphic forms and their spectral theory. Automorphic forms are generalisations of periodic functions and modular forms in particular. They live on a negatively curved Riemannian manifold M of dimension n, which we call the hyperbolic n-space. Our automorphic forms (Maaß forms and Eisenstein series) are also eigenfunctions of the Laplace–Beltrami operator Δ on M. The spectrum of Δ reveals a lot of information about the geometry of M. For example, Weyl's law [71] for compact M relates the eigenvalues λ_j of Δ to the volume of M by

$$\#\left\{j\geq 0 : \sqrt{\lambda_j}\leq\lambda\right\}\sim\frac{\operatorname{vol}(M)}{(4\pi)^{n/2}\Gamma(\frac{n}{2}+1)}\lambda^n, \tag{1.2}$$

as $\lambda\to\infty$. In two dimensions it is possible to use the complex upper half-plane as a model for M. Define

$$\mathbb{H}^2 = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}.$$

Then M can be identified with $\Gamma\backslash\mathbb{H}^2$, where Γ is a discrete subgroup of the group of orientation preserving isometries of $\mathbb{H}^2$. The lengths of primitive closed geodesics, *prime geodesics*, on M (for any dimension) share many properties with the usual prime numbers. The prime geodesics are in one-to-one correspondence with primitive hyperbolic conjugacy classes (dilations) of Γ, [53, §10.5]. Let $\pi_\Gamma(x)$ denote the number of primitive hyperbolic conjugacy classes of norm less than x in Γ. The Prime Geodesic Theorem for $\Gamma\backslash\mathbb{H}^2$ states that

$$\pi_\Gamma(x) \sim \operatorname{li}(x).$$

The conjecture for the error term is the same as that for the PNT. Of particular interest to us are groups Γ of arithmetic nature, because in this case many computations can be carried out explicitly. A natural example is $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ which acts on $\mathbb{H}^2$ by Möbius transformations. The current best result for the error term in the Prime Geodesic Theorem in this case is $\pi_\Gamma(x) - \operatorname{li}(x) = O(x^{25/36+\epsilon})$, due to Soundararajan and Young [94].

The crux of Chapter 5 is an asymptotic formula that relates an exponential sum, denoted $S(T,X)$, over the eigenvalues of Δ to the lengths of prime geodesics and, surprisingly, to the usual prime numbers. The exponential sum $S(T,X)$ is intricately related to the error term in the Prime Geodesic Theorem. It turns out that $S(T,X)$ also influences the error term in the hyperbolic lattice point problem over local averages. This is the hyperbolic analogue of the classical Gauß Circle Problem, which asks to estimate the number of lattice points $\mathbb{Z}^2$ in a circle of radius X. It is straightforward to see that the leading term approximates the area of the circle, so the focus of the problem is in estimating the error term.

We investigate a certain variant of the hyperbolic lattice point problem in three dimensions on compact $\Gamma\backslash\mathbb{H}^3$, where

$$\mathbb{H}^3 = \{p = z + jy : z \in \mathbb{C}, y > 0\}.$$

We also apply a hyperbolic analogue of the large sieve to get improvements on the error term on average. In Chapter 7 we study the equidistribution of masses of scattering states for three dimensional hyperbolic manifolds $\Gamma\backslash\mathbb{H}^3$, where Γ is now a Bianchi group of class number one. This is done by studying the quantum limits (large eigenvalue limits) of Eisenstein series, the generalised eigenfunctions of Δ, off the critical line. We also identify the correct limit when the measures become equidistributed.

1.1 Summary of Results

We now give a detailed description of the main results of this thesis. For the sake of brevity, we defer full definitions to the pertinent chapters where appropriate.

In Chapter 3 we look at the value distribution of Dirichlet L-functions $L(s,\chi)$ for primitive Dirichlet characters χ. We consider the values of a pair of such functions, with characters of distinct prime moduli, on sample points, which we choose to be the non-trivial zeros of ζ or a horizontal shift of them. Away from the critical line we prove linear independence.

Theorem 1.1. *Assume the Riemann Hypothesis. Let χ_1, χ_2 be two primitive Dirichlet characters modulo distinct primes q and ℓ, respectively. Let $\sigma \in (\frac{1}{2}, 1)$. Then, for a positive proportion of the non-trivial zeros of $\zeta(s)$ with $\gamma > 0$, the values of the Dirichlet L-functions $L(\sigma + i\gamma, \chi_1)$ and $L(\sigma + i\gamma, \chi_2)$ are linearly independent over $\mathbb{R}$.*

The main ingredient of the proof is a uniform version of Landau's formula (3.3). The classical Landau's formula (3.2) says that for a fixed $x > 1$,

$$\sum_{0<\gamma\leq T} x^{\rho} = -\frac{T}{2\pi}\Lambda(x) + O(\log T), \qquad \text{as } T \to \infty, \tag{1.3}$$

where $\Lambda : \mathbb{R} \longrightarrow \mathbb{R}$ is the von Mangoldt function extended to $\mathbb{R}$ by

$$\Lambda(x) = \begin{cases} \log p, & \text{if } x = p^m \text{ for some prime } p \text{ and } m \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

On the critical line we use a different technique, by identifying the sample points as the residues of ζ'/ζ and using contour integration. The result we obtain is unconditional, but we fail to prove a positive proportion.

Theorem 1.2. *Two Dirichlet L-functions with primitive characters modulo distinct primes attain different values at cT non-trivial zeros of $\zeta(s)$ up to height T, for some positive constant c.*

In Remark 3.2 we discuss why a positive proportion on the critical line is difficult to prove. Even for sample points with more structure — arithmetic progressions — a positive proportion was not achieved until recently by Li and Radziwiłł [64].

In Chapter 5 we work on $M = \Gamma\backslash\mathbb{H}^2$ with $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Let $\lambda_j = \frac{1}{4} + t_j^2$ be the eigenvalues of Δ. Our theorem in Chapter 5 concerns an exponential sum defined in terms of the spectral parameters t_j. For $X > 1$, let

$$S(T,X) = \sum_{|t_j| \leq T} X^{it_j}.$$

This is the cut-off kernel of the wave equation on M. It is also analogous to the left-hand side of Landau's formula (1.3). Information about the growth of $S(T,X)$ is crucial in proving good estimates on the error term in the Prime Geodesic Theorem as well as the local average in the hyperbolic lattice point problem [77]. Trivially from Weyl's law (1.2) we know that

$$S(T,X) \ll T^2.$$

The conjecture, presented in [77], is square root cancellation in T with only arbitrarily small contribution in terms of X:

$$S(T,X) \ll_\epsilon T^{1+\epsilon} X^\epsilon.$$

To find evidence of the conjecture, we first engaged in a numerical investigation of $S(T,X)$ and found many interesting phenomena regarding its peak points and vanishing. These tests were also carried out for the corresponding sine kernel. Based on these observations we conjectured and later proved an asymptotic expansion for $S(T,X)$ in terms of T for a fixed $X > 1$. This takes a form which is very similar to Landau's formula (1.3) with a hyperbolic von Mangoldt function Λ_Γ. The role of prime numbers is replaced by the lengths of primitive closed geodesics (see (5.1)). The result is the following theorem.

Theorem 1.3. *For a fixed $X > 1$, we have*

$$\begin{aligned} S(T,X) = \frac{\mathrm{vol}(\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2)}{\pi} \frac{\sin(T\log X)}{\log X} T + \frac{T}{\pi}\left(X^{1/2} - X^{-1/2}\right)^{-1} \Lambda_\Gamma(X) \\ + \frac{2T}{\pi} X^{-1/2} \Lambda(X^{1/2}) + O(T/\log T), \end{aligned}$$

as $T \to \infty$.

Notice that the prime numbers also appear in this formula through $\Lambda(X^{1/2})$. This factor is due to the non-compactness of M, and in particular the existence of a continuous spectrum of Δ and the explicit form of the scattering matrix for $\mathrm{PSL}_2(\mathbb{Z})$ given in terms of the Riemann zeta function. The proof of Theorem 1.3 is an application of the Selberg Trace Formula, which relates the spectral trace of Δ to geometric properties of M.

Chapter 6 deals with hyperbolic lattice point problems in three dimensions. These problems ask us to estimate the number of points in an orbit Γp, $p \in \mathbb{H}^3$, which are contained in some domain $D \subset \mathbb{H}^3$. We expect that this number approximates the volume of D in any dimension. In the two dimensional standard lattice point problem one wants to estimate

$$N(z,w,X) = \#\{\gamma \in \Gamma : d(\gamma z, w) \leq \operatorname{arccosh}(X/2)\},$$

where d is the hyperbolic distance on $\mathbb{H}^2$. If we assume that there are no small eigenvalues of Δ (see Section 4.2), the counting function satisfies

$$N(z,w,X) = \frac{\pi}{\operatorname{vol}(\Gamma\backslash\mathbb{H}^2)} X + O(X^{2/3}). \tag{1.4}$$

This was proved by Selberg for cofinite Γ, [53, Theorem 12.1]. The error term has never been improved for any cofinite Γ or any points z, $w \in \mathbb{H}^2$, [53, pg.175], though it is conjectured that the optimal bound is $O(X^{1/2+\epsilon})$.

We concentrate on three dimensional hyperbolic lattice point problems. The upper half-space $\mathbb{H}^3$ can be understood as a subset of the quaternions with the last coordinate set to zero. Denote points $p \in \mathbb{H}^3$ by $p = (x_1, x_2, y)$. Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ be discrete and suppose that $\Gamma\backslash\mathbb{H}^3$ is compact, i.e. Γ is *cocompact*. Instead of a hyperbolic sphere, we consider lattice points in a sector of the hyperbolic space of angle 2Θ emanating from the line $x_2 = y = 0$. This sector is bisected by a totally geodesic hyperplane $\mathscr{P} = \{p \in \mathbb{H}^3 : x_2(p) = 0\}$. The angle between $\mathscr{P}$ and a ray from the origin to a point $p \in \mathbb{H}^3$ is denoted by $v(p)$. Also, let $\{u_j\}_{j\geq 0}$ be a complete orthonormal system in $L^2(\Gamma\backslash\mathbb{H}^3)$ of eigenfunctions of Δ. First we identify the main term in the counting and prove a pointwise bound on the error term.

Theorem 1.4. *Let Γ be a cocompact discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. Set $H = \Gamma \cap \mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{C})}(\mathscr{P})$ and let $\hat{u}_j$ be the period integral of u_j over the fundamental domain of H restricted to $\mathscr{P}$. Define*

$$N(p,X) = \#\{\gamma \in H\backslash\Gamma : (\cos v(\gamma p))^{-1} \leq X\}.$$

Then we have

$$N(p,X) = M(p,X) + E(p,X),$$

where

$$M(p,X) = \frac{\operatorname{vol}(H\backslash\mathscr{P})}{\operatorname{vol}(\Gamma\backslash\mathbb{H}^3)} X^2 + \sum_{1<s_j<2} \frac{2^{s_j-1}}{s_j} \hat{u}_j u_j(p) X^{s_j},$$

and

$$E(p,X) = O(X^{3/2}).$$

The conjecture is, as in two dimensions, that the error term exhibits square root cancellation so that $E(p,X) = O(X^{1+\epsilon})$. The strength of the error term in Theorem 1.4 is the same as in the standard hyperbolic lattice point problem in three dimensions, see Lax and Phillips [63]. We then apply large sieve inequalities in $\mathbb{H}^3$. The first is due to Chamizo [11] and the second we prove in Theorem 6.16. These are used to obtain mean square results for the spatial and radial averages of the error term.

Theorem 1.5. *Let* Γ *be a cocompact discrete subgroup of* $\mathrm{PSL}_2(\mathbb{C})$*. Then, for* $X > 2$,

$$\int_{\Gamma\backslash\mathbb{H}^3} |E(p,X)|^2 \, d\mu(p) \ll X^2 \log^2 X. \tag{1.5}$$

Theorem 1.6. *Let* Γ *be a cocompact discrete subgroup of* $\mathrm{PSL}_2(\mathbb{C})$*. Then, for* $X > 2$,

$$\frac{1}{X}\int_X^{2X} |E(p,x)|^2 \, dx \ll X^{2+2/3} \log X. \tag{1.6}$$

Notice that, while in the spatial aspect (1.5) we obtain the conjecture on average, this is not the case for the radial average (1.6). In Section 6.3.4 we explain why this is the case and why we do not expect any improvements from this method. Moreover, in Appendix B we prove that the same limitation in the radial average applies to the standard hyperbolic lattice point problem in three dimensions.

The final result of this thesis is about a famous conjecture of Rudnick and Sarnak [89] known as the Quantum Unique Ergodicity conjecture, which is a statement about the equidistribution of measures $d\mu_j = |u_j|^2 d\mu$, where u_j are $L^2(M)$ eigenfunctions of Δ on a hyperbolic manifold M and μ is the standard volume measure on M. The conjecture states that for any such M the measures $d\mu_j$ converge to $d\mu$ in the weak-$*$ topology as $j \to \infty$. For some class of hyperbolic manifolds this conjecture has already been solved by Lindenstrauss [65] and Soundararajan [95].

For non-compact $M = \Gamma\backslash\mathbb{H}^2$, where $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ is a general cofinite group, very little is known about the discrete spectrum of Δ. It is not even understood whether it is infinite [84]. Luo [67] shows that, under some multiplicity assumptions on the eigenvalues, Weyl's law fails for deformed congruence groups. Hence the limit of $d\mu_j$ might not be relevant for the discrete spectrum. It is possible to consider instead the scattering states, which are residues of Eisenstein series $E(z,s)$ at the non-physical poles of the scattering matrix. It is known that under small deformations of Γ in the Teichmüller space, cusp forms dissolve into scattering states as characterised by Fermi's Golden Rule [83, 81].

Our last theorem generalises a result of Petridis, Raulf, and Risager [80] from $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2$ to $\Gamma\backslash\mathbb{H}^3$ for $\Gamma = \mathrm{PSL}_2(\mathcal{O}_K)$, where $\mathcal{O}_K$ is the ring of integers of an imaginary quadratic field K of class number one. They consider modified measures $d\mu_j$ with u_j replaced by the scattering states. Their strategy is to evaluate limits of $d\mu_{s(t)} = |E(z,s(t))|^2 d\mu$, as $t \to \infty$, where $s(t) = \sigma_t + it$ is a sequence with $\sigma_t \to \sigma_\infty$. We prove two results on $M = \Gamma\backslash\mathbb{H}^3$ depending on whether we are on the critical line ($\sigma_\infty = 1$ in three dimensions) or not. Let ζ_K be the Dedekind zeta function of K with discriminant d_K, and define $d\mu_{s(t)} = |E(p,s(t))|^2 d\mu$.

Theorem 1.7. *Assume $\sigma_\infty = 1$ and $(\sigma_t - 1)\log t \to 0$. Let A and B be compact Jordan measurable subsets of M. Then*

$$\frac{\mu_{s(t)}(A)}{\mu_{s(t)}(B)} \to \frac{\mu(A)}{\mu(B)},$$

as $t \to \infty$. In fact, we have

$$\mu_{s(t)}(A) \sim \mu(A) \frac{2(2\pi)^2}{|\mathcal{O}_K^\times||d_K|\zeta_K(2)} \log t.$$

On the other hand, off the critical line the measures do not become equidistributed.

Theorem 1.8. *Assume $\sigma_\infty > 1$. Let A be a compact Jordan measurable subset of M. Then*

$$\mu_{s(t)}(A) \to \int_A E(p, 2\sigma_\infty) d\mu(p),$$

as $t \to \infty$.

In Section 7.1 we show how to apply this to quantum limits of the scattering states v_{ρ_n} of $\Gamma\backslash\mathbb{H}^3$, where ρ_n is a sequence of the non-trivial zeros of ζ_K.

Theorem 1.9. *Let A be a compact Jordan measurable subset of M. Then*

$$\int_A |v_{\rho_n}(p)|^2 d\mu(p) \to \int_A E(p, 4 - 2\gamma_\infty) d\mu(p),$$

as $n \to \infty$.

In the above γ_∞ is the limit of the real parts of the non-trivial zeros of the Dedekind zeta function corresponding to the number field K. Under the Generalised Riemann Hypothesis we would have $\gamma_\infty = 1/2$ so that the limit would be $E(p,3)d\mu$.

Part I

Value Distribution of Dirichlet L-functions

Chapter 2

Analytic Theory of the Riemann Zeta Function and Dirichlet L-functions

2.1 The History

The discussion in this introduction is heavily based on the book *An introduction to the theory of the Riemann Zeta-Function* by Patterson [76].

Definition 2.1. For $\operatorname{Re} s > 1$, the Riemann zeta function $\zeta(s)$ is given by the absolutely convergent infinite series

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

It is not difficult to show that this series also has an expression as an infinite product over primes.

Lemma 2.1 ([76, §1.1]). *For* $\operatorname{Re} s > 1$, $\zeta(s)$ *satisfies the infinite product*

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}}, \tag{2.1}$$

where the product is taken over all prime numbers p.

This was first observed by Euler already in 1749 and as such (2.1) is called the *Euler product*. He also computed values for $\zeta(2n)$, $n \in \mathbb{N}$, such as $\zeta(2) = \pi^2/6$. No explicit values for $\zeta(2n+1)$ are known.

Definition 2.2. The von Mangoldt function $\Lambda(n)$ on $\mathbb{N}$ is given by

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^m \text{ for } p \text{ a prime and } m \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

A short computation reveals a simple form for the logarithmic derivative of ζ:

$$\frac{\zeta'}{\zeta}(s) = -\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}, \tag{2.2}$$

which is valid for $\operatorname{Re} s > 1$. The Prime Number Theorem is often formulated in terms of the von Mangoldt function. The following form of the error term was proved by de la Vallée Poussin in 1899.

Theorem 2.2 (Prime Number Theorem [54, (2.37)]). *There exists a constant $c > 0$, such that, as $X \to \infty$,*

$$\sum_{n \le X} \Lambda(n) = X + O(X e^{-c(\log X)^{1/2}}).$$

It is not difficult to pass between the above and the traditional asymptotic form of the PNT with $\pi(x)$ (see [76, §4.5]).

Riemann's major contribution to number theory was to realise that there is a strong connection between the zeros of $\zeta(s)$ and the prime numbers. This was done by showing that $\zeta(s)$ has an analytic continuation to all of $\mathbb{C}$ with a simple pole at $s = 1$. He also proved that the Riemann zeta function satisfies a *functional equation*, which we will need later on. We will outline one of Riemann's original proofs of these facts as it will serve as a useful point of comparison in Part II.

Lemma 2.3. *The Riemann zeta function $\zeta(s)$ has an analytic continuation as a meromorphic function to all of $\mathbb{C}$ with a simple pole of residue 1 at $s = 1$. Moreover, let*

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

Then $\zeta(s)$ satisfies the functional equation

$$\zeta(1-s) = \gamma(s)\zeta(s), \tag{2.3}$$

for any $s \in \mathbb{C}$.

Proof. This proof appears in Titchmarsh [100, §2.6] and Gelbart [31, pg. 187]. Let

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

By the definition of the Gamma function as a Mellin transform of e^{-s} (A.1) we have

$$\pi^{-s}\Gamma(s)\zeta(2s)=\sum_{n=1}^{\infty}\int_0^{\infty} e^{-n^2\pi x}x^{s-1}\,dx=\int_0^{\infty}\frac{\theta(ix)-1}{2}x^{s-1}\,dx,$$

for $\operatorname{Re} s>1/2$. The theta function satisfies the functional equation

$$\theta\left(\tfrac{i}{x}\right)=\sqrt{x}\,\theta(ix).$$

With this we can split the integration and write

$$\begin{aligned}\pi^{-s}\Gamma(s)\zeta(2s)&=\int_0^{1}\frac{\theta(ix)}{2}x^{s-1}\,dx-\frac{1}{2s}+\int_1^{\infty}\frac{\theta(ix)-1}{2}x^{s-1}\,dx\\&=\int_1^{\infty}(x^{s-1}+x^{1/2-(s+1)})\frac{\theta(ix)-1}{2}\,dx-\frac{1}{2s}-\frac{1}{1-2s}.\end{aligned}\tag{2.4}$$

The integral in (2.4) is absolutely convergent for any $s\in\mathbb{C}$, so the analytic continuation follows. The right-hand side is invariant under $s\mapsto 1/2-s$, which proves the functional equation. □

Definition 2.3. The completed Riemann zeta function $\xi(s)$ is defined as

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

From the functional equation it follows that $\xi(s)=\xi(1-s)$ for any $s\in\mathbb{C}$. The Euler product (2.1) immediately implies that $\zeta(s)$ cannot have any zeros in the region of absolute convergence. On the other hand, from the functional equation with $s=2n+1$, $n\in\mathbb{N}$, we can see that ζ has zeros at negative even integers coming from the poles of the Gamma function. These are called the *trivial zeros* of the Riemann zeta function. In his memoir [87], Riemann conjectured that $\zeta(s)$ has infinitely many zeros in the region $0\le\operatorname{Re} s\le 1$, and in particular, that it has a Hadamard-type product in terms of these zeros. First, let

$$\mathscr{Z}=\{\rho\in\mathbb{C}\,:\,\zeta(\rho)=0,\,0\le\operatorname{Re}\rho\le 1\}$$

be the set of *non-trivial zeros* of $\zeta(s)$ with multiplicities. Also, define

$$\mathscr{Z}_+ = \{\rho \in \mathscr{Z} : \operatorname{Im}\rho > 0\}.$$

Lemma 2.4 ([76, §3.1]). *For all $s \in \mathbb{C}$ we have the absolutely convergent product*

$$-s(1-s)\pi^{-s/2}\Gamma(\tfrac{s}{2})\zeta(s) = \prod_{\rho\in\mathscr{Z}_+}\left(1-\frac{s}{\rho}\right)\left(1-\frac{s}{1-\rho}\right). \tag{2.5}$$

A comparison of the products (2.1) and (2.5) revealed to Riemann a deep connection between the primes and the non-trivial zeros $\mathscr{Z}$.

Lemma 2.5 (The Explicit Formula of ζ, [76, §3.9]). *For $X > 1$, we have*

$$\sum_{n\le X}\Lambda(n) = X - \sum_{\rho\in\mathscr{Z}}\frac{X^\rho}{\rho} - \log 2\pi - \tfrac{1}{2}\log(1-X^{-2}),$$

where the last term in the summation over n is taken with weight $1/2$ if $n = X$.

It is therefore crucial to understand the set $\mathscr{Z}$. From the simple observation that $\overline{\zeta(s)} = \zeta(\bar{s})$ and the functional equation (2.3) it follows that

$$\rho \in \mathscr{Z} \implies \bar{\rho},\, 1-\rho,\, 1-\bar{\rho} \in \mathscr{Z}.$$

Riemann was led to suspect that this symmetry becomes in fact the "strongest possible", that is, all the non-trivial zeros would lie on the line $\operatorname{Re}s = 1/2$. This conjecture became known as the Riemann Hypothesis, which remains open to this day.

Riemann Hypothesis. *The non-trivial zeros of $\zeta(s)$ satisfy*

$$\mathscr{Z} \subset \left\{s \in \mathbb{C} : \operatorname{Re}s = \tfrac{1}{2}\right\}.$$

Remark 2.1. Even proving that $\zeta(\sigma + it) \neq 0$ for $\sigma = 1$ and $t \neq 0$ takes some effort. The PNT is in fact equivalent to this fact [76, §4.3]. Any improvements on σ define what are called *zero-free regions* for ζ and they correspond exactly to better error terms in the Prime Number Theorem [76, §5.8]. While we will not use them in the case of ζ, the corresponding zero-free regions for Dirichlet L-functions will be vital for our mean value estimates.

In order to study the distribution of the zeta zeros it is convenient to define the following counting function.

Definition 2.4. The counting function $N(T)$ of the non-trivial zeros of ζ is defined for $T > 0$ by

$$N(T) = \#\{\rho \in \mathscr{Z} : 0 < \operatorname{Im} \rho < T\}. \tag{2.6}$$

Riemann found (without proof) the asymptotic expansion of this function. It was later proved by von Mangoldt.

Lemma 2.6 ([76, §4.9])**.** *As $T \to \infty$, we have*

$$N(T) = \frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T). \tag{2.7}$$

While the series representation of $\zeta(s)$ is not valid in the critical strip, it is still possible to approximate the zeta function by truncated Dirichlet series. These are called *approximate functional equations*. Even though we do not need an approximate functional equation for ζ, we prefer to include one for comparison with the corresponding formula for Dirichlet L-functions.

Theorem 2.7 (Approximate functional equation for ζ, [76, §6.1])**.** *Let $X, Y \geq 1$ and $\alpha < 1$. Suppose $\sigma + it = s \in \mathbb{C}$ satisfies $1 - \alpha < \sigma < \alpha$ and $|t| = 2\pi XY$. Then*

$$\zeta(s) = \sum_{n \leq X} n^{-s} + \pi^{s-1/2} \frac{\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})} \sum_{m \leq Y} m^{s-1} + O(X^{1/2-\sigma}(X^{-1/2} + Y^{-1/2}) \log XY),$$

where the implied constant depends only on α.

2.2 Arithmetic Functions

As a brief interlude, we provide some basic definitions of multiplicative number theory.

Definition 2.5. A function $f : \mathbb{N} \longrightarrow \mathbb{C}$ is called an *arithmetic function*. We use the notations $f(n)$ and f_n interchangeably. We say that f is *additive* if

$$f(mn) = f(m) + f(n) \qquad \text{for all } m, n \in \mathbb{N} \text{ with } (m,n) = 1.$$

On the other hand, f is *multiplicative* if

$$f(mn) = f(m)f(n) \qquad \text{for all } m, n \in \mathbb{N} \text{ with } (m,n) = 1. \tag{2.8}$$

Moreover, if (2.8) is true for any $m, n \in \mathbb{N}$, then f is called *completely multiplicative*.

Suppose that a_n and b_n are two arithmetic functions. Then the function c_n defined by $\sum_n a_n n^{-s} \sum_n b_n n^{-s} = \sum_n c_n n^{-s}$ is also arithmetic.

Definition 2.6. The *Dirichlet convolution* of two arithmetic functions f and g is given by

$$(f * g)(n) = \sum_{d|n} f(d) g(\tfrac{n}{d}).$$

In particular, if f and g are multiplicative, then so is $f * g$, see [54, pg. 13]. We list some common arithmetic functions that we will need in this thesis.

Definition 2.7. The *Möbius function* $\mu(n)$ is defined by

$$\mu(n) = \begin{cases} (-1)^r, & \text{if } n \text{ is a product of } r \text{ distinct primes,} \\ 0, & \text{otherwise.} \end{cases}$$

Definition 2.8. The *divisor function* $d(n)$ is given by

$$d(n) = \sum_{d|n} 1.$$

Definition 2.9. We define the *Euler totient function* $\varphi(n)$ by

$$\varphi(n) = \#\{\text{residue classes } (a \bmod n) : (a,n) = 1\}.$$

All of $\mu(n)$, $d(n)$ and $\varphi(n)$ are multiplicative [54, pp. 12–15]. The von Mangoldt function is also an arithmetic function, but it is not multiplicative. Finally, we look at arithmetic functions defined in terms of characters of finite abelian groups G. These are group homomorphisms $\chi : G \longrightarrow \mathbb{C}^\times$. In particular, we are interested in the case $G = (\mathbb{Z}/q\mathbb{Z})^\times$.

Definition 2.10. Let $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \longrightarrow \mathbb{C}^\times$ be a character defined on the residue classes coprime to q. We extend χ to a function on all of $\mathbb{Z}$ by setting $\chi(n) = 0$ if $(n,q) > 1$. The resulting function is called a *Dirichlet character modulo* q.

Dirichlet characters are completely multiplicative and periodic modulo q. The *principal character modulo* q, χ_0, is the Dirichlet character corresponding to the trivial group character. To each Dirichlet character we associate its *conductor*, that is, the smallest divisor $\tilde{q} \geq 1$ of q such that $\chi = \chi_0 \tilde{\chi}$, where $\tilde{\chi}$ is a Dirichlet character modulo $\tilde{q}$. In this case, we say that χ is induced by $\tilde{\chi}$. If $\tilde{q} = q$, χ is called *primitive*.

Notice that if q is a prime, then all Dirichlet characters modulo q are primitive. If χ only assumes values ± 1, then χ is a *real* Dirichlet character, otherwise it is *complex*. Furthermore, if $\chi(-1) = 1$ then χ is *even* and if $\chi(-1) = -1$ then χ is *odd*. The character orthogonality relations for Dirichlet characters take the following form.

Lemma 2.8 ([54, pg. 45]). *Let χ be a Dirichlet character modulo q. Then*

$$\sum_{a \bmod q} \chi(a) = \begin{cases} \varphi(q), & \text{if } \chi = \chi_0, \\ 0, & \text{otherwise}, \end{cases} \tag{2.9}$$

and

$$\sum_{\chi \bmod q} \chi(a) = \begin{cases} \varphi(q), & \text{if } a \equiv 1 \bmod q, \\ 0, & \text{otherwise.} \end{cases} \tag{2.10}$$

We will make extensive use of equation (2.10) to decompose sums defined over residue classes. We finish this section with a discussion on sums of Dirichlet characters $\chi(n)$ over n.

Definition 2.11. The *Gauß sum*, $G(k,\chi)$, for a Dirichlet character χ modulo q is given by

$$G(k,\chi) = \sum_{a=1}^{q} \chi(a) e^{2\pi i a k/q}. \tag{2.11}$$

We sometimes write $G(1,\chi) = G(\chi)$.

Of course trivially we have that $|G(k,\chi)| \leq q$. We can write down the Gauß sum of an induced character χ in terms of the Gauß sum of its factorisation.

Lemma 2.9 ([54, Lemma 3.1]). *Let χ be a Dirichlet character modulo q induced by the primitive character $\tilde{\chi}$ modulo $\tilde{q}$. Then*

$$G(\chi) = \mu\left(\frac{q}{\tilde{q}}\right) \tilde{\chi}\left(\frac{q}{\tilde{q}}\right) G(\tilde{\chi}).$$

If χ is primitive, then

$$|G(\chi)|^2 = q. \tag{2.12}$$

The last equation (2.12) can also be written as

$$G(1,\overline{\chi})G(-1,\chi) = q.$$

Moreover, we also have [54, (3.15)]

$$G(\chi)G(\overline{\chi}) = \chi(-1)q. \tag{2.13}$$

Usually we can express the more general Gauß sums $G(k,\chi)$ in terms of $G(\chi)$.

Lemma 2.10 ([54, Lemma 3.2]). *Let χ be a Dirichlet character modulo q. Then, if $(k,q)=1$, we have*

$$G(k,\chi) = \overline{\chi}(k)G(\chi). \tag{2.14}$$

If q is a prime then (2.14) is valid for all $k \in \mathbb{Z}$ as $\chi(k) = 0$ for $(k,q) > 1$.

2.3 Dirichlet L-functions

We now use Dirichlet characters to define a generalisation of the Riemann zeta function. In Part II we need a generalisation of ζ to number fields, the Dedekind zeta function, but we delay the discussion to Chapter 7. Dirichlet L-functions were introduced by Dirichlet already in 1837 — 22 years before Riemann's memoir — as functions of a real variable, mind you, to prove his theorem on primes in arithmetic progressions [22, §1].

Definition 2.12. Let χ be a Dirichlet character modulo q. The Dirichlet L-function associated with χ is defined for $\operatorname{Re} s > 1$ by

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

It is not surprising that many properties of $L(s,\chi)$ can be deduced from those of ζ either from the simple observation that $|\chi(n)| \le 1$ or by summation by parts and properties of Gauß sums. Clearly $L(s,\chi)$ has an Euler product

$$L(s,\chi) = \prod_p (1-\chi(p)p^{-s})^{-1}.$$

We first observe from the Euler product that if χ is induced by $\tilde{\chi}$, that is $\chi = \tilde{\chi}\chi_0$, then [22, pg. 37]

$$L(s,\chi) = L(s,\tilde{\chi}) \prod_{p|q} (1-\tilde{\chi}(p)p^{-s}). \tag{2.15}$$

Hence it is sufficient to concentrate on primitive χ. The Dirichlet L-function $L(s,\chi)$ has an analytic continuation to all of $\mathbb{C}$ as an entire function unless $\chi = \chi_0$ is principal modulo q for some q in which case it has a simple pole at $s=1$ with residue

$$\operatorname*{res}_{s=1} L(s,\chi_0)=\frac{\varphi(q)}{q}. \tag{2.16}$$

The functional equation varies depending on whether $\chi(-1)=\pm 1$.

Lemma 2.11. *Let χ be a primitive Dirichlet character modulo q. Define*

$$\Delta(s,\chi)=\left(\frac{2\pi}{q}\right)^{-s}\frac{G(\overline{\chi})}{q}\Gamma(s)\left(e^{-\pi is/2}+\overline{\chi}(-1)e^{\pi is/2}\right).$$

Then,

$$L(1-s,\overline{\chi})=\Delta(s,\chi)L(s,\chi). \tag{2.17}$$

Proof. This follows from [22, §9] after separating the cases $\chi(-1)=1$ and $\chi(-1)=-1$ and applying the reflection formulae (A.4) and (A.3), respectively, followed by the duplication formula (A.5) and equation (2.13). □

From the functional equation we can see that the zero set of $L(s,\chi)$ is more complicated than that of $\zeta(s)$. First of all, the location of the trivial zeros (zeros s with $\operatorname{Re} s<0$) depends on whether χ is even or odd. Suppose now that χ is primitive. Then, if $\chi(-1)=1$ then $L(-2n,\chi)=0$ for $n\in\mathbb{N}$. Notice that also $L(0,\chi)=0$. On the other hand, if $\chi(-1)=-1$ then $L(1-2n,\chi)=0$. Naturally $L(s,\chi)\neq 0$ for $\operatorname{Re} s>1$.

In the critical strip there are again infinitely many complex zeros, which we call the non-trivial zeros of $L(s,\chi)$. The Generalised Riemann Hypothesis (GRH) states that all of the non-trivial zeros lie on the line $\operatorname{Re} s=1/2$. *A priori*, it is possible for $L(s,\chi)$ to have an “exceptional” zero on the real line close to 1. This is called a *Siegel zero* (although mathematicians were aware of it even before his work [22, §21]). The point is that for any modulus q there is at most one character, which is real, and can potentially admit a Siegel zero. Moreover, the distance of the zero from $s=1$ depends on q. Since we are working with characters of fixed modulus, this will not be an issue for us.

Approximate functional equations for $L(s,\chi)$ are more intricate than those of ζ. This is because it is possible to take advantage of the cancellation in the coefficients of the Dirichlet series, while for ζ such savings are not available. Hence for $L(s,\chi)$ it is often beneficial to work with a smooth cut-off in the approximate functional equation. We decide to use a sharp cut-off, although it is plausible that our results in this part could be improved slightly by using a smoothed out version of the approximate functional equation. An example of a smoothed out approximate functional equation can be found in [69, pp. 443–446]. We use the following sharp version.

Theorem 2.12 (Lavrik [62]). *Let χ be a primitive character mod q. For $s=\sigma+it$ with $0<\sigma<1$, $t>0$, and $x=\Delta\sqrt{\frac{qt}{2\pi}}$, $y=\Delta^{-1}\sqrt{\frac{qt}{2\pi}}$, and $\Delta\geq 1$, $\Delta\in\mathbb{N}$, we have*

$$L(s,\chi)=\sum_{n\leq x}\frac{\chi(n)}{n^s}+\varepsilon(\chi)\left(\frac{q}{\pi}\right)^{\frac{1}{2}-s}\frac{\Gamma\left(\frac{1-s+\mathfrak{a}}{2}\right)}{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}\sum_{n\leq y}\frac{\overline{\chi}(n)}{n^{1-s}}+R_{xy}, \tag{2.18}$$

with

$$R_{xy}\ll\sqrt{q}\left(y^{-\sigma}+x^{\sigma-1}(qt)^{1/2-\sigma}\right)\log 2t,$$

and in particular, for $x=y$,

$$R\ll x^{-\sigma}\sqrt{q}\log 2t.$$

Here $\varepsilon(\chi)=q^{-1/2}i^{\mathfrak{a}}G(1,\chi)$ and $\mathfrak{a}=(1-\chi(-1))/2$.

It is not hard to see that this formula is, in fact, valid for all real $\Delta\geq 1$. Moreover, (approximate) functional equations for imprimitive characters do also exist, but they are more complicated (see [54, (9.69) and Lemma 10.3]). Therefore we restrict our attention to primitive characters.

2.4 Convexity and Subconvexity

One of the important questions about the value distribution of ζ in the critical strip is the order of growth of $|\zeta(s)|$, for $s=\sigma+it$, as $|t|\to\infty$. The natural first estimate is the *convexity bound*, which follows from trivial estimates on ζ.

Proposition 2.13. *Define*

$$\mu_0(\sigma)=\begin{cases}0, & \text{if } \sigma>1,\\ \frac{1-\sigma}{2}, & \text{if } 0<\sigma<1,\\ \frac{1}{2}-\sigma, & \text{if } \sigma<0.\end{cases} \tag{2.19}$$

Then for $|t|>1$ and $\sigma\in[a,b]$ we have

$$\zeta(s)\ll_{a,b,\epsilon}|t|^{\mu_0(\sigma)+\epsilon}. \tag{2.20}$$

Proof. Due to absolute (and uniform) convergence $|\zeta(s)|$ is bounded for any $\sigma > \sigma_0$ with a fixed $\sigma_0 > 1$ and any $t \in \mathbb{R}$. On the left half-plane, $\sigma < 0$, we have by the functional equation (2.3) and Stirling asymptotics (A.6) that $\zeta(s) = O(t^{1/2-\sigma+\epsilon})$. We can now apply the Phragmén–Lindelöf Theorem (see [76, A5.8]) to conclude that μ_0 has to be a convex function. The result follows. See [76, §2.12] for details. □

The bound off the critical strip is of course sharp. Hence, any improvements on $\mu_0(\sigma)$ for $\sigma \in (0,1)$ are called *subconvex estimates*. To this end, let us define

$$\mu(\sigma) = \inf\{a \in \mathbb{R} : \zeta(\sigma + it) = O(|t|^a) \text{ as } t \to \infty\}.$$

Notice that by the functional equation we have

$$\mu(\sigma) - \mu(1-\sigma) = \frac{1}{2} - \sigma. \tag{2.21}$$

In light of (2.21), the conjecture is that the convexity is the sharpest possible.

Lindelöf Hypothesis. *For any $\epsilon > 0$, we have as $t \to \infty$*

$$\zeta(\tfrac{1}{2} + it) = O(|t|^\epsilon). \tag{2.22}$$

In other words,

$$\mu(\tfrac{1}{2}) = 0.$$

It should be noted that Ernst Lindelöf (1870–1946) was a fellow Finn, who worked on topology and complex analysis. His doctoral supervisor, Hjalmar Mellin (1854–1933), also appears in this thesis in the form of the Mellin transform. It turns out that Riemann Hypothesis actually implies the Lindelöf Hypothesis, but not the other way around (as far as we know) [76, §5.5]. It is fairly painless to obtain $\mu(\frac{1}{2}) \le 1/4$. The first improvement on this was made by Weyl who showed that

$$\zeta(\tfrac{1}{2} + it) \ll t^{1/6} \log^{3/2} t$$

through a method which requires estimates on exponential sums. This can be extended to include σ other than $1/2$, [100, Theorem 5.8]. The current best known result on the critical line is due to Bourgain [8]. He shows that

$$\zeta(\tfrac{1}{2} + it) \ll_\epsilon t^{53/342+\epsilon}.$$

For L-functions subconvexity is a more subtle matter. Apart from bounds in t we also have to consider the q-aspect. In our work the modulus is always fixed so we will not deal with the q-aspect further. In the t-aspect subconvexity of degree two L-functions is a crucial component of the QUE conjecture. This will be discussed further in Chapter 7.

Chapter 3

Discrete Mean Values of Dirichlet L-functions

Mean values play an important role in analytic number theory. This is because often enough knowledge about mean values allows one to deduce concrete pointwise results about the value distribution of the corresponding function. For example, the Lindelöf Hypothesis is equivalent to estimating the integral moments of ζ on the critical line.

Lemma 3.1 ([100, (13.2.1)]). *Lindelöf Hypothesis is equivalent to having the estimates*

$$\frac{1}{T}\int_1^T |\zeta(\tfrac{1}{2}+it)|^{2k}\,dt = O(T^{\epsilon}),$$

for all $k \in \mathbb{N}$ and $\epsilon > 0$.

This is a difficult problem and the required bound is currently established only for $k=1$ and $k=2$ by Hardy and Littlewood [38] and Ingham [51], respectively. Gonek [32, Theorem 2] proved the following mean value for $\alpha \in \mathbb{R}$, $|\alpha| \le (\log T)/2\pi$,

$$\sum_{0<\gamma\le T} |\zeta(\tfrac{1}{2}+i(\gamma+2\pi\alpha/\log T)|^2 = \left(1-\left(\frac{\sin\pi\alpha}{\pi\alpha}\right)^2\right)\frac{T}{2\pi}\log^2 T + O(T\log^{7/4} T), \quad (3.1)$$

from which we can get an estimate on large gaps between non-trivial zeros of ζ [70],

$$\limsup_{\gamma} \frac{(\gamma'-\gamma)\log\gamma}{2\pi} > 1.9,$$

where γ' and $\gamma > 0$ denote consecutive ordinates of non-trivial zeros.

The value distribution of $\zeta(s)$ and $L(s,\chi)$ has a long history and has recently attracted attention in for example [29], [33], [97]. A typical focus for investigation for these functions is the distribution of their zeros. In 1976 Fujii [27] showed that a positive proportion of zeros of $L(s,\psi)L(s,\chi)$ are distinct, where the characters are primitive and not necessarily of distinct moduli. A zero of the product is said to be distinct if it is a zero of only one of the two, or if it is a zero of both then it occurs with different multiplicities for each function.

It is, in fact, believed that all zeros of Dirichlet L-functions to primitive characters are simple, and that two L-functions with distinct primitive characters do not share any non-trivial zeros at all. This comes from the Grand Simplicity Hypothesis (GSH), see [88]. The hypothesis is that the set

$$\{\gamma \geq 0 : L(\tfrac{1}{2}+i\gamma,\chi)=0 \text{ and } \chi \text{ is primitive}\}$$

is linearly independent over $\mathbb{Q}$. Since we are counting with multiplicities, it is implicit in the statement of the GSH that all zeros of Dirichlet L-functions are simple, and that $\gamma \neq 0$, i.e. $L(\frac{1}{2},\chi)\neq 0$. We know unconditionally that a positive proportion (35%) of the zeros of $L(s,\chi)$ are simple [2]. Conrey, Iwaniec, and Soundararajan [20] prove that 56% of all zeros of $L(s,\chi)$ are simple when averaged over the family of all Dirichlet characters χ in a suitable way. Under the GRH it was shown by Chandee et al. [13] that the proportion of simple zeros for one $L(s,\chi)$ can be strengthened to 91% if χ is primitive.

A similar result is expected for an even bigger class of functions. Murty and Murty [72] proved that two functions of the Selberg class $\mathscr{S}$ cannot share too many zeros (counted with multiplicity). They show that if F, $G\in\mathscr{S}$ then $F=G$ provided that

$$|Z_F(T)\Delta Z_G(T)| = o(T),$$

where $Z_F(T)$ denotes the set of zeros of $F(s)$ in the region $\operatorname{Re} s \geq 1/2$ and $|\operatorname{Im} s| \leq T$, and Δ is the symmetric difference. Booker [5] proves unconditionally that degree two L-functions associated to classical holomorphic newforms have infinitely many simple zeros.

Apart from looking at the zeros, there has also been investigation into the a-values of ζ and $L(s,\chi)$, that is, the distribution of s such that $\zeta(s)=a$ (or $L(s,\chi)=a$) for some fixed $a\in\mathbb{C}$. Garunkštis and Steuding [30] prove a discrete average for ζ' over the a-values of ζ, which implies that there are infinitely many simple a-points in the critical strip. On the critical line, however, we do not even know whether there are infinitely many a-points. For further results on the distribution of simple a-points see [34]. On the other hand, we can also look at points where $\zeta(s)$ (or $L(s,\chi)$) has a specific fixed argument $\varphi \in (-\pi,\pi]$. In [56] the authors prove that ζ takes arbitrarily large values with argument φ, i.e.

$$\max_{\substack{0<t\leq T\\ \mathrm{Arg}(\zeta(1/2+it))=\varphi}} |\zeta(\tfrac{1}{2}+it)| \gg (\log T)^{5/4}.$$

In this work we compare the values or the arguments of two Dirichlet L-functions at a specific set of sample points. We choose these points to be either the Riemann zeros, $\beta+i\gamma$, or a horizontal shift of them. We will prove two results in this direction depending on whether we are on the critical line or not.

Theorem 3.2. *Assume the Riemann Hypothesis, i.e. $\beta=\frac{1}{2}$. Let χ_1, χ_2 be two primitive Dirichlet characters modulo distinct primes q and ℓ, respectively. Let $\sigma\in(\frac{1}{2},1)$. Then, for a positive proportion of the non-trivial zeros of $\zeta(s)$ with $\gamma>0$, the values of the Dirichlet L-functions $L(\sigma+i\gamma,\chi_1)$ and $L(\sigma+i\gamma,\chi_2)$ are linearly independent over $\mathbb{R}$.*

Remark 3.1. If the values $L(\sigma+i\gamma,\chi_1)$ and $L(\sigma+i\gamma,\chi_2)$ are linearly independent over $\mathbb{R}$, then in particular their arguments are different.

Theorem 3.3. *Two Dirichlet L-functions with primitive characters modulo distinct primes attain different values at cT non-trivial zeros of $\zeta(s)$ up to height T, for some positive constant c.*

Remark 3.2. In Theorem 3.3 we fail to obtain a positive proportion and we expect this to be a limitation of the method used. Garunkštis and Kalpokas [28] look at the mean square of a single Dirichlet L-function at the zeros of another, and show that it is non-zero for at least cT of the zeros for some explicit $c>0$. On the other hand, attempting to introduce a mollifier to overcome this limitation does not seem hopeful either. Martin and Ng [68] evaluate the mollified first and second moments of $L(s,\chi)$ in arithmetic progressions on the critical line and prove that at least $T(\log T)^{-1}$ of the values are non-zero, which misses a positive proportion by a logarithm. This was extended to a positive proportion by Li and Radziwiłł [64]. However, their method relies on the strong rigidity of the arithmetic progression and fails when the sequence is slightly perturbed.

Remark 3.3. We assume that the conductors of χ_1 and χ_2 are primes in order to make the notation simpler. It should be possible to generalise our results to the case when the conductors are coprime or have distinct prime factors.

A main ingredient in the proofs is the Gonek–Landau formula and results derived from it. In 1911 Landau [60] proved that for a fixed $x > 1$,

$$\sum_{0<\gamma\le T} x^{\rho} = -\frac{T}{2\pi}\Lambda(x) + O(\log T), \tag{3.2}$$

as $T \to \infty$. Here the sum runs over the positive imaginary parts of the Riemann zeros. What is striking in this formula is that the right-hand side grows by a factor of T only if x is a prime power. This version of Landau's formula is of limited practical use since the estimate is not uniform in x. Gonek [32] proved a version of Landau's formula which is uniform in both x and T:

Lemma 3.4 (Gonek–Landau Formula)**.** *Let $x, T > 1$. Then*

$$\begin{aligned}\sum_{0<\gamma\le T} x^{\rho} &= -\frac{T}{2\pi}\Lambda(x) + O(x\log 2xT \log\log 3x)\\ &+ O\left(\log x \min\left(T, \frac{x}{\langle x\rangle}\right)\right) + O\left(\log 2T \min\left(T, \frac{1}{\log x}\right)\right),\end{aligned} \tag{3.3}$$

where $\langle x\rangle$ denotes the distance from x to the nearest prime power other than x itself.

If one fixes x then this reduces to the original result of Landau as $T \to \infty$. As an application of Lemma 3.4 Gonek proves (under the RH) the mean value (3.1).

For Theorem 3.3 we need a modified version of the Gonek–Landau formula for integrals, see [29].

Lemma 3.5 (Modified Gonek Lemma)**.** *Suppose that $\sum_{n=1}^{\infty} a(n)n^{-s}$ converges for $\sigma > 1$ and $a(n) = O(n^{\epsilon})$. Let $a = 1 + \log^{-1} T$. Then*

$$\frac{1}{2\pi i}\int_{a+i}^{a+iT}\left(\frac{m}{2\pi}\right)^{s}\Gamma(s)\exp\left(\delta\frac{\pi i s}{2}\right)\sum_{n=1}^{\infty}\frac{a(n)}{n^{s}}\,ds$$
$$= \begin{cases} \displaystyle\sum_{n\le\frac{Tm}{2\pi}} a(n)\exp\left(-2\pi i\frac{n}{m}\right) + O(m^{a}T^{1/2+\epsilon}), & \text{if } \delta = -1,\\ O(m^{a}), & \text{if } \delta = +1.\end{cases}$$

3.1 Proof of Theorem 3.2

The proof will follow the steps of [32, Theorem 2] and [78, Theorem 1.9] for the Riemann zeta function and GL_2 L-functions.

Two non-zero complex numbers z and w are linearly independent over the reals is equivalent to the quotient z/w being non-real, or that $|z\overline{w}-\overline{z}w|>0$. For us z and w are values of Dirichlet L-functions. Instead of looking at these functions at a single point, we will average over multiple points with a fixed real part $\sigma\in(\frac{1}{2},1)$ and the imaginary part at the height of the Riemann zeros.

We are assuming the RH purely because it makes the proof simpler as expressions of the form x^ρ become easier to deal with if we know the real part explicitly. On the other hand, the distribution of these specific points does not seem to have any impact on the rest of the proof. We suspect that the RH is not an essential requirement. In fact, following [33], it might be possible to obtain the result without the RH by integrating

$$\frac{\zeta'}{\zeta}(s-\sigma)B(s,P)L(s,\chi_1)\overline{L(s,\chi_2)}$$

over a suitable contour. This picks the desired points as residues of the integrand yielding the required sum. This idea is also used in the proof of Theorem 3.3.

The proof will be divided into three propositions after which the main result follows easily. In the first proposition we want to calculate discrete mean values of sums of terms of the type $L(\sigma+i\gamma,\chi_1)\overline{L(\sigma+i\gamma,\chi_2)}$ and its complex conjugate. If we subtract one of these mean values from the other then each term is non-zero precisely when the two numbers are linearly independent over the reals. Hence we need to prove that the two mean values are not equal, which is the content of Proposition 3.8. Finally, we get the main result by applying the Cauchy–Schwarz inequality to the difference of the mean values. Because of this we also need to estimate a sum of squares of the absolute values of the above quantities, that is,

$$|L(\sigma+i\gamma,\chi_1)\overline{L(\sigma+i\gamma,\chi_2)}-\overline{L(\sigma+i\gamma,\chi_1)}L(\sigma+i\gamma,\chi_2)|^2.$$

This is done in Proposition 3.7.

The first obstacle in our proof is that the mean values are complex conjugates. To show that the difference is non-zero leads to determining whether $\operatorname{Im}L(2\sigma,\chi_1\overline{\chi}_2)\neq 0$, which does not always hold. Thus we need to introduce some kind of weighting in order to shove these sums off balance. We do this by multiplying by a finite Dirichlet polynomial, $B(s,P)$, which cancels some terms from either of the L-functions, depending on which mean value we are considering. We define

$$B(s,P)=\prod_{p\leq P}(1-\chi_1(p)p^{-s})(1-\chi_2(p)p^{-s}) \tag{3.4}$$

for some fixed prime P, depending only on q and ℓ, to be determined in Proposition 3.8. Let us also assume that this Dirichlet polynomial has the expansion

$$B(s,P) = \sum_{n\leq R} c_n n^{-s},$$

for some R depending on P. Since $|c_p| \leq 2$ for any prime p, we have for all n that

$$|c_n| \leq 2^P. \tag{3.5}$$

Remark 3.4. It should be possible to simplify the proof by replacing $B(s,P)$ with $(1-\chi_1(\ell)\ell^{-s})(1-\chi_2(q)q^{-s})$.

Let $N(T)$ be the counting function defined in (2.6). We prove the following propositions.

Proposition 3.6. *Assume the Riemann Hypothesis. Fix $\sigma \in (\frac{1}{2}, 1)$. Then, with $s = \sigma + i\gamma$, we have*

$$\sum_{0<\gamma\leq T} B(s,P)L(s,\chi_1)\overline{L(s,\chi_2)} \sim N(T) \sum_{n=1}^{\infty} \frac{d_n \overline{\chi}_2(n)}{n^{2\sigma}}, \tag{3.6}$$

and

$$\sum_{0<\gamma\leq T} B(s,P)\overline{L(s,\chi_1)}L(s,\chi_2) \sim N(T) \sum_{n=1}^{\infty} \frac{e_n \overline{\chi}_1(n)}{n^{2\sigma}}, \tag{3.7}$$

where

$$B(s,P)L(s,\chi_1) = \sum_{n=1}^{\infty} \frac{d_n}{n^s}, \quad B(s,P)L(s,\chi_2) = \sum_{n=1}^{\infty} \frac{e_n}{n^s},$$

and γ runs through the ordinates of the non-trivial zeros of ζ.

Proposition 3.7. *Fix $\sigma \in (\frac{1}{2}, 1)$. Suppose $s = \sigma + i\gamma$ and let*

$$A(\gamma) = B(s,P)\big(L(s,\chi_1)\overline{L(s,\chi_2)} - \overline{L(s,\chi_1)}L(s,\chi_2)\big).$$

Then, under the Riemann Hypothesis,

$$\sum_{0<\gamma\leq T} |A(\gamma)|^2 \ll N(T), \tag{3.8}$$

where γ runs through the ordinates of the non-trivial zeros of ζ.

Proposition 3.8. *Fix $\sigma \in (\frac{1}{2}, 1)$. Under the Riemann Hypothesis we can find a prime P such that*

$$\sum_{0<\gamma\leq T} A(\gamma) \sim C \cdot N(T), \tag{3.9}$$

for some non-zero constant C. Here γ runs through the ordinates of the non-trivial zeros of ζ.

Proof of Theorem 3.2. By the Cauchy–Schwarz inequality and Propositions 3.7 and 3.8 we have

$$\sum_{\substack{0<\gamma\leq T\\ A(\gamma)\neq 0}} 1 \geq \frac{|\sum_{0<\gamma\leq T} A(\gamma)|^2}{\sum_{0<\gamma\leq T} |A(\gamma)|^2} \gg \frac{|C|^2 N(T)^2}{N(T)} = |C|^2 N(T). \tag{3.10}$$

This proves that a positive proportion of the $A(\gamma)$'s are non-zero; in particular, for the same γ's, $L(s,\chi_1)$ and $L(s,\chi_2)$ are linearly independent over the reals. □

3.1.1 Proof of Proposition 3.6

As the d_n's contain only products of characters, we have $d_n = O(1)$ (see proof of Proposition 3.8 for the calculation). In particular they define a multiplicative arithmetic function. We define, for a fixed t,

$$B(s,P) \sum_{n\leq\sqrt{\frac{q\ell t}{2\pi}}} \chi_1(n) n^{-s} = \sum_{n\leq R\sqrt{\frac{q\ell t}{2\pi}}} d'_n n^{-s}.$$

We have

$$d_n = \sum_{n=km} c_k \chi_1(m), \tag{3.11}$$

and hence for $n \leq R\sqrt{\frac{q\ell t}{2\pi}}$

$$d'_n = \sum_{\substack{n=km\\ k\leq R}} c_k \chi_1(m). \tag{3.12}$$

From this it follows that $d_n = d'_n$ for $n \leq \sqrt{\frac{q\ell t}{2\pi}}$. We also need to show that $d'_n \ll 1$. Let $p_1,\ldots,p_h$, for some $h>1$, denote all the primes below P in an increasing order. Define $\widetilde{P} = p_1\cdots p_h P$. From the product representation of $B(s,P)$, equation (3.4), we see that $c_n = 0$ for $n>1$, if n contains any prime factors greater than P. Thus, write $n = p_1^{\alpha_1}\cdots p_h^{\alpha_h} P^{\alpha_0}\nu = n'\nu$ for some $\alpha_i \geq 0$. Then

$$d'_n = \sum_{\substack{k|n'\\ k<R}} c_k \chi_1\left(\frac{n'}{k}\nu\right) = \chi_1(\nu) d'_{n'}.$$

Thus it suffices to consider n with prime factors only up to P. Since $B(s,P)$ has a finite Euler product of degree two we have $c_{p^j} = 0$ for any prime p and $j \geq 3$. So we can suppose that $n = p_1^{\alpha_1} \cdots p_h^{\alpha_h} P^{\alpha_0}$, where $0 \le \alpha_i \le 2$ for all $i \le h$. The number of summands in (3.12) is then at most 3^{h+1}. By (3.5), we find that $|d_n'| \le 2^P 3^{h+1}$. In particular, $d_n' \ll 1$ as required.

The approximate functional equation (2.18) for χ_1 with $\Delta = \sqrt{\ell}$ gives

$$L(s,\chi_1) = \sum_{n \le \sqrt{\frac{q\ell t}{2\pi}}} \chi_1(n) n^{-s} + X(s,\chi_1) \sum_{n \le \sqrt{\frac{qt}{2\pi\ell}}} \overline{\chi}_1(n) n^{s-1} + O(t^{-\sigma/2}\log t + t^{-1/4}),$$

where

$$X(s,\chi) = \varepsilon(\chi)\left(\frac{q}{\pi}\right)^{1/2-s} \frac{\Gamma\left(\frac{1-s+\mathfrak{a}}{2}\right)}{\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)}.$$

Similarly for χ_2 with $\Delta = \sqrt{q}R$ we get

$$L(s,\chi_2) = \sum_{n \le R\sqrt{\frac{q\ell t}{2\pi}}} \chi_2(n) n^{-s} + X(s,\chi_2) \sum_{n \le \frac{1}{R}\sqrt{\frac{\ell t}{2\pi q}}} \overline{\chi}_2(n) n^{s-1} + O(t^{-\sigma/2}\log t + t^{-1/4}).$$

We can now expand the left-hand side in (3.6) to

$$\begin{aligned}
&\sum_{0<\gamma\le T} B(s,P) \\
&\times\Bigg(\sum_{n \le \sqrt{\frac{q\ell\gamma}{2\pi}}} \chi_1(n) n^{-s} + X(s,\chi_1) \sum_{n \le \sqrt{\frac{q\gamma}{2\pi\ell}}} \overline{\chi}_1(n) n^{s-1} + O(\gamma^{-\sigma/2}\log\gamma + \gamma^{-1/4})\Bigg) \\
&\times\Bigg(\sum_{n \le R\sqrt{\frac{q\ell\gamma}{2\pi}}} \overline{\chi}_2(n) n^{-\overline{s}} + \overline{X(s,\chi_2)} \sum_{n \le \frac{1}{R}\sqrt{\frac{\ell\gamma}{2\pi q}}} \chi_2(n) n^{\overline{s}-1} + O(\gamma^{-\sigma/2}\log\gamma + \gamma^{-1/4})\Bigg).
\end{aligned} \tag{3.13}$$

Denote the sum with χ_1 and $\overline{\chi}_2$ by $M(T)$. We will take care of the other sums at the end of the proof. The main term comes from the diagonal entries of $M(T)$. First, write

$$\begin{aligned}
M(T) &= \sum_{0<\gamma\le T} B(s,P) \sum_{n \le \sqrt{\frac{q\ell\gamma}{2\pi}}} \chi_1(n) n^{-s} \sum_{m \le R\sqrt{\frac{q\ell\gamma}{2\pi}}} \overline{\chi}_2(m) m^{-\overline{s}} \\
&= \sum_{0<\gamma\le T} \sum_{n \le R\sqrt{\frac{q\ell\gamma}{2\pi}}} d_n' n^{-\sigma-i\gamma} \sum_{m \le R\sqrt{\frac{q\ell\gamma}{2\pi}}} \overline{\chi}_2(m) m^{-\sigma+i\gamma}.
\end{aligned}$$

Then we separate the diagonal terms

$$M(T) = \sum_{0<\gamma\le T}\Bigg(\sum_{n\le R\sqrt{\frac{q\ell\gamma}{2\pi}}}\frac{d'_n\overline{\chi}_2(n)}{n^{2\sigma}} + \sum_{n\neq m}^{R\sqrt{\frac{q\ell\gamma}{2\pi}}}\frac{d'_m\overline{\chi}_2(n)}{(nm)^{\sigma}}\left(\frac{n}{m}\right)^{i\gamma}\Bigg)$$
$$= Z_1 + Z_2. \tag{3.14}$$

The asymptotics in (3.6) come from Z_1. We have

$$Z_1 = \sum_{0<\gamma\le T}\Bigg(\sum_{n=1}^{\infty}\frac{d_n\overline{\chi}_2(n)}{n^{2\sigma}} - \sum_{n>R\sqrt{\frac{q\ell\gamma}{2\pi}}}\frac{d_n\overline{\chi}_2(n)}{n^{2\sigma}} + \sum_{n\le R\sqrt{\frac{q\ell\gamma}{2\pi}}}\frac{(d'_n-d_n)\overline{\chi}_2(n)}{n^{2\sigma}}\Bigg)$$
$$= N(T)\sum_{n=1}^{\infty}\frac{d_n\overline{\chi}_2(n)}{n^{2\sigma}} + C_1 + C_2.$$

We need to estimate C_1 and C_2. For C_1 we have

$$C_1 \ll \sum_{0<\gamma\le T}\sum_{n>\sqrt{\gamma}} n^{-2\sigma} \ll \sum_{0<\gamma\le T}\gamma^{1/2-\sigma} = o(N(T)).$$

Similarly,

$$C_2 \ll \sum_{0<\gamma\le T}\sum_{n>\sqrt{\frac{q\ell\gamma}{2\pi}}} n^{-2\sigma} = o(N(T)).$$

To estimate Z_2 we wish to exchange the order of summation and apply the Gonek–Landau formula (3.3). Splitting and rewriting Z_2 in terms of the zeros of ζ we get

$$Z_2 = \sum_{0<\gamma\le T}\sum_{n\le R\sqrt{\frac{q\ell\gamma}{2\pi}}}\sum_{m<n}\Bigg(\frac{d'_m\overline{\chi}_2(n)}{n^{\sigma+1/2}m^{\sigma-1/2}}\left(\frac{n}{m}\right)^{1/2+i\gamma} + \frac{d'_n\overline{\chi}_2(m)}{n^{\sigma+1/2}m^{\sigma-1/2}}\overline{\left(\frac{n}{m}\right)^{1/2+i\gamma}}\Bigg)$$
$$= \sum_{n\le R\sqrt{\frac{q\ell T}{2\pi}}}\sum_{m<n}\sum_{\frac{2\pi n^2}{q\ell R^2}\le\gamma\le T}\Bigg(\frac{d'_m\overline{\chi}_2(n)}{n^{\sigma+1/2}m^{\sigma-1/2}}\left(\frac{n}{m}\right)^{\rho} + \frac{d'_n\overline{\chi}_2(m)}{n^{\sigma+1/2}m^{\sigma-1/2}}\overline{\left(\frac{n}{m}\right)^{\rho}}\Bigg).$$

To apply the Gonek–Landau formula we split the innermost sum to $0<\gamma\le T$ and $0<\gamma\le 2\pi n^2/q\ell R^2$. Hence, we can write

$$Z_2 = Z_{21} + Z_{22} + Z_{23} + Z_{24} + Z_{25},$$

with

$$Z_{21} = -\frac{T}{2\pi}\sum_{n\le R\sqrt{\frac{q\ell T}{2\pi}}}\sum_{m<n}\frac{d'_m\overline{\chi}_2(n) + d'_n\overline{\chi}_2(m)}{n^{\sigma+1/2}m^{\sigma-1/2}}\Lambda\left(\frac{n}{m}\right),$$

$$Z_{22} \ll \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m<n} \frac{n^2 \Lambda(n/m)}{n^{\sigma+1/2} m^{\sigma-1/2}},$$

$$Z_{23} \ll \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m<n} \frac{1}{n^{\sigma+1/2} m^{\sigma-1/2}} \frac{n}{m} \log \frac{2nT}{m} \log\log \frac{3n}{m},$$

$$Z_{24} \ll \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m<n} \frac{1}{n^{\sigma+1/2} m^{\sigma-1/2}} \log \frac{n}{m} \min\left(T, \frac{n/m}{\langle n/m \rangle}\right),$$

and

$$Z_{25} \ll \log T \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m<n} \frac{1}{n^{\sigma+1/2} m^{\sigma-1/2}} \min\left(T, \frac{1}{\log(n/m)}\right).$$

We begin by estimating Z_{21}. The only non-vanishing terms are with $m|n$. Thus we write $n = km$ and obtain

$$Z_{21} \ll \frac{T}{2\pi} \sum_{k \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m < \frac{R}{k}\sqrt{\frac{q\ell T}{2\pi}}} \frac{\Lambda(k)}{k^{\sigma+1/2} m^{2\sigma}} \ll \frac{T}{2\pi} \sum_{k \le R\sqrt{\frac{q\ell T}{2\pi}}} k^{\epsilon-\sigma-1/2} \sum_{m < \frac{R}{k}\sqrt{\frac{q\ell T}{2\pi}}} m^{-2\sigma},$$

since $\Lambda(k) \ll k^{\epsilon}$ for any $\epsilon > 0$. Since both sums are partial sums of convergent series we get $Z_{21} = O(T)$. Working similarly with Z_{22} gives

$$\begin{aligned} Z_{22} &\ll \sum_{k \le R\sqrt{\frac{q\ell T}{2\pi}}} \sum_{m < \frac{R}{k}\sqrt{\frac{q\ell T}{2\pi}}} \frac{\Lambda(k)}{k^{\sigma-3/2} m^{2\sigma-2}} \ll \sum_{k \le R\sqrt{\frac{q\ell T}{2\pi}}} k^{3/2-\sigma+\epsilon} \sum_{m < \frac{R}{k}\sqrt{\frac{q\ell T}{2\pi}}} m^{2-2\sigma} \\ &\ll \sum_{k \le R\sqrt{\frac{q\ell T}{2\pi}}} k^{3/2-\sigma+\epsilon} \left(\left(\frac{T^{1/2}}{k} \right)^{3-2\sigma} + 1 \right) \ll T^{\frac{3-2\sigma}{2}} \sum_{k \ll T^{1/2}} k^{\sigma-3/2+\epsilon} = O(T). \end{aligned}$$

For Z_{23} we get

$$\begin{aligned} Z_{23} &\ll \log T \log\log T \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \frac{1}{n^{\sigma-1/2}} \sum_{m<n} \frac{1}{m^{\sigma+1/2}} \\ &\ll \log T \log\log T \sum_{n \le R\sqrt{\frac{q\ell T}{2\pi}}} \frac{1}{n^{\sigma-1/2}} = o(N(T)). \end{aligned}$$

In order to estimate Z_{24} we write $n = um + r$, where $-m/2 < r \le m/2$. Hence

$$\left\langle u + \frac{r}{m} \right\rangle = \begin{cases} \frac{|r|}{m}, & \text{if } u \text{ is a prime power and } r \neq 0, \\ \ge \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{3.15}$$

Let $c = R\sqrt{q\ell/2\pi}$ then $n/m \le n \le c\sqrt{T}$, and so

$$\begin{aligned} Z_{24} &\ll \log T \sum_{n\le cT^{1/2}} \sum_{m<n} \frac{1}{n^{\sigma+1/2}m^{\sigma-1/2}} \frac{n}{m} \frac{1}{\langle n/m\rangle} \\ &\ll \log T \sum_{m\le cT^{1/2}} \sum_{u\le \lfloor cT^{1/2}/m\rfloor+1} \sum_{-\frac{m}{2}<r\le\frac{m}{2}} \frac{1}{m^{\sigma+1/2}(um+r)^{\sigma-1/2}} \frac{1}{\langle u+\frac{r}{m}\rangle}, \end{aligned}$$

and then evaluate the sum over r depending on whether u is a prime power or not to get

$$\begin{aligned} &\ll \log T \sum_{m\le cT^{1/2}} \sum_{u\le \lfloor cT^{1/2}/m\rfloor+1} \left(\Lambda(u)\frac{m\log m}{m^{\sigma+1/2}(um)^{\sigma-1/2}} + \frac{m}{m^{\sigma+1/2}(um)^{\sigma-1/2}}\right) \\ &\ll \log T \sum_{m\le cT^{1/2}} \frac{\log m}{m^{2\sigma-1}} \sum_{u\ll cT^{1/2}/m} \frac{u^{\epsilon}}{u^{\sigma-1/2}} = O(T). \end{aligned}$$

Finally, for Z_{25} set $m = n-r$, $1\le r\le n-1$. So in particular

$$\log\frac{n}{m} > -\log\left(1-\frac{r}{n}\right) > \frac{r}{n}.$$

Hence,

$$\begin{aligned} Z_{25} &\ll \log T \sum_{n\le cT^{1/2}} \sum_{1\le r<n} \frac{1}{n^{\sigma+1/2}(n-r)^{\sigma-1/2}} \frac{n}{r} \\ &\ll \log T \sum_{n\le cT^{1/2}} \frac{1}{n^{\sigma-1/2}} \sum_{r\le n-1} \frac{1}{r} = O(T). \end{aligned}$$

It remains to estimate all the other terms in (3.13). By repeating the analysis done for Z_1 and Z_2 we obtain the following estimates

$$\sum_{0<\gamma\le T} \left| \sum_{n\le R\sqrt{\frac{q\ell\gamma}{2\pi}}} d'_n n^{-\sigma-i\gamma} \right|^2 \ll N(T), \tag{3.16}$$

and

$$\sum_{0<\gamma\le T} \left| \sum_{n\le R\sqrt{\frac{q\ell\gamma}{2\pi}}} \overline{\chi}_2(n) n^{-\sigma+i\gamma} \right|^2 \ll N(T). \tag{3.17}$$

With trivial changes to the above argument we get,

$$\sum_{0<\gamma\le T} \left| \sum_{n\le \sqrt{\frac{q\gamma}{2\pi\ell}}} \chi_1(n) n^{\sigma-1+i\gamma} \right|^2 \ll T^{\sigma-1/2} N(T), \tag{3.18}$$

and

$$\sum_{0<\gamma\leq T}\left|\sum_{n\leq\frac{1}{R}\sqrt{\frac{\ell\gamma}{2\pi q}}}\overline{\chi}_2(n)n^{\sigma-1-i\gamma}\right|^2\ll T^{\sigma-1/2}N(T). \tag{3.19}$$

We also need to estimate the order of growth of the derivative in t of $|X(s,\chi)|^2$. First, notice that $|\varepsilon(\chi)|=1$, so

$$|X(s,\chi)|=\left(\frac{q}{\pi}\right)^{1/2-\sigma}\left|\Gamma\left(\frac{1-s+\mathfrak{a}}{2}\right)\right|\left|\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)\right|^{-1}.$$

By Stirling asymptotics

$$|X(s,\chi)|^2\sim A\gamma^{1-2\sigma}\left(\frac{q}{\pi}\right)^{1-2\sigma}, \tag{3.20}$$

as $\overline{\Gamma(z)}=\Gamma(\overline{z})$, where A is some non-zero constant. Thus, with $\psi=\Gamma'/\Gamma$,

$$\begin{aligned}\frac{d}{d\gamma}|X(s,\chi)|^2&=|X(s,\chi)|^2\frac{i}{2}\left(\psi\left(\overline{\tfrac{1-s+\mathfrak{a}}{2}}\right)-\psi\left(\tfrac{1-s+\mathfrak{a}}{2}\right)+\psi\left(\overline{\tfrac{s+\mathfrak{a}}{2}}\right)-\psi\left(\tfrac{s+\mathfrak{a}}{2}\right)\right)\\&=|X(s,\chi)|^2\frac{i}{2}\left(2i\left(\arg\left(\overline{\tfrac{1-s+\mathfrak{a}}{2}}\right)-\arg\left(\tfrac{s+\mathfrak{a}}{2}\right)\right)+O(\gamma^{-2})\right)\\&\ll\gamma^{1-2\sigma}\Big(O(\gamma^{-1})+O(\gamma^{-2})\Big)\ll\gamma^{-2\sigma},\end{aligned} \tag{3.21}$$

by a standard estimate on ψ (A.7) and the Taylor expansion of arccot. Let

$$S(T)=\sum_{0<\gamma\leq T}|X(\sigma+i\gamma,\chi_1)|^2\left|\sum_{n\leq R}c_n n^{-\sigma-i\gamma}\right|^2\left|\sum_{n\leq\sqrt{\frac{q\gamma}{2\pi\ell}}}\chi_1(n)n^{\sigma-1+i\gamma}\right|^2.$$

We use summation by parts, (3.21), (3.18), and (3.19) to see that

$$\begin{aligned}S(T)=|X(\sigma+iT,\chi_1)|^2&\sum_{0<\gamma\leq T}\left|\sum_{n\leq\sqrt{\frac{q\gamma}{2\pi\ell}}}\chi_1(n)n^{\sigma-1+i\gamma}\right|^2\\&-\int_1^T\sum_{0<\gamma\leq t}\left|\sum_{n\leq\sqrt{\frac{q\gamma}{2\pi\ell}}}\chi_1(n)n^{\sigma-1+i\gamma}\right|^2\frac{d}{dt}|X(\sigma+it,\chi_1)|^2\,dt,\end{aligned}$$

which simplifies to

$$S(T)\ll T^{1-2\sigma}T^{\sigma-1/2}N(T)+\int_1^T t^{\sigma-1/2}N(t)t^{-2\sigma}\,dt.$$

The first term is clearly $o(N(T))$. For the integral we use the fact that $N(t)=O(t\log t)$

to estimate it as

$$\int_1^T t^{1/2-\sigma}\log t\,dt \ll T^{3/2-\sigma+\epsilon}.$$

Hence we have that $S(T)=o(N(T))$, and similarly

$$\sum_{0<\gamma\le T}|X(\sigma+i\gamma,\overline{\chi}_2)|^2\left|\sum_{n\le\frac{1}{R}\sqrt{\frac{\ell\gamma}{2\pi q}}}\overline{\chi}_2(n)n^{\sigma-1-i\gamma}\right|^2 \ll T^{3/2-\sigma+\epsilon}=o(N(T)).$$

Finally we use the Cauchy–Schwarz inequality, (3.16), (3.17), and the above two equations to estimate all other terms in (3.13) as $o(N(T))$. □

3.1.2 Proof of Proposition 3.7

Since $B(s,P)$ is a finite Dirichlet polynomial it is bounded independently of T. Thus, to estimate $\sum_{\gamma\le T}|A(\gamma)|^2$, it suffices to estimate

$$\sum_{0<\gamma\le T}|L(s,\chi_1)|^2|L(s,\chi_2)|^2=O(N(T)). \tag{3.22}$$

The approximate functional equation for $L(s,\chi_1)$, as in the proof of Proposition 3.6, gives

$$\begin{aligned}L(s,\chi_1)&=\sum_{n\le\sqrt{\frac{q\ell t}{2\pi}}}\chi_1(n)n^{-s}+X(s,\chi_1)\sum_{n\le\sqrt{\frac{qt}{2\pi\ell}}}\overline{\chi}_1(n)n^{s-1}+O(t^{-\sigma/2}\log t+t^{-1/4})\\&=W_1+X(s,\chi_1)W_2+O(t^{-\sigma/2}\log t)+O(t^{-1/4}).\end{aligned}$$

Similarly,

$$L(s,\chi_2)=Y_1+X(s,\chi_2)Y_2+O(t^{-\sigma/2}\log t)+O(t^{-1/4}),$$

where

$$Y_1=\sum_{n\le\sqrt{\frac{q\ell t}{2\pi}}}\chi_2(n)n^{-s},\quad Y_2=\sum_{n\le\sqrt{\frac{\ell t}{2\pi q}}}\overline{\chi}_2(n)n^{s-1}.$$

We have

$$\sum_{0<\gamma\le T}Y_1\overline{Y}_1W_1\overline{W}_1=\sum_{0<\gamma\le T}\sum_{m,n,\mu,\nu\le\sqrt{\frac{q\ell\gamma}{2\pi}}}\frac{\chi_1(m)\chi_2(n)\overline{\chi}_1(\mu)\overline{\chi}_2(\nu)}{(mn\mu\nu)^{\sigma}}\left(\frac{\mu\nu}{mn}\right)^{i\gamma}. \tag{3.23}$$

Again, we consider the diagonal terms separately from the rest of the sum. The number of solutions to $mn=\mu\nu=r$ is at most the square of the number of divisors of r, $d(r)^2$. Thus

$$\sum_{0<\gamma\leq T}\sum_{mn=\mu\nu}^{(q\ell\gamma/2\pi)^{1/2}}\frac{\chi_1(m)\chi_2(n)\overline{\chi}_1(\mu)\overline{\chi}_2(\nu)}{(mn)^{2\sigma}}\ll\sum_{0<\gamma\leq T}\sum_{r=1}^{\infty}\frac{d(r)^2}{r^{2\sigma}}\ll N(T), \tag{3.24}$$

since the inner series converges. For the off-diagonal terms set $\mu\nu=r$ and $mn=s$. We can treat the cases $s<r$ and $s>r$ separately. In the following analysis we assume $m,\, n,\, \mu,\, \nu\leq(q\ell T/2\pi)^{1/2}$. Consider first the terms with $s<r$ in (3.23). We have that

$$Z_2=\sum_{r\leq q\ell T/2\pi}\sum_{s<r}\sum_{m|s,\,\mu|r}\frac{\chi_1(m)\chi_2(s/m)\overline{\chi}_1(\mu)\overline{\chi}_2(r/\mu)}{r^{\sigma}s^{\sigma}}\sum_{K\leq\gamma\leq T}\left(\frac{r}{s}\right)^{i\gamma}, \tag{3.25}$$

where $K=\min(T,(2\pi/q\ell)\max(m^2,s^2/m^2,\mu^2,r^2/\mu^2))$. Applying Gonek–Landau Formula (3.4) to Z_2 gives

$$\begin{aligned}Z_2&=\sum_{r\ll T}\sum_{s<r}\sum_{m|s,\,\mu|r}\frac{\chi_1(m)\chi_2(s/m)\overline{\chi}_1(\mu)\overline{\chi}_2(r/\mu)}{r^{\sigma+1/2}s^{\sigma-1/2}}\left(\sum_{0<\gamma\leq T}\left(\frac{r}{s}\right)^{\rho}-\sum_{0<\gamma<K}\left(\frac{r}{s}\right)^{\rho}\right)\\&=Z_{21,2}+Z_{23}+Z_{24}+Z_{25},\end{aligned}$$

with

$$\begin{aligned}Z_{21,2}&=\sum_{r\ll T}\sum_{s<r}\sum_{m|s,\,\mu|r}\frac{\chi_1(m)\chi_2(s/m)\overline{\chi}_1(\mu)\overline{\chi}_2(r/\mu)}{r^{\sigma+1/2}s^{\sigma-1/2}}\frac{K-T}{2\pi}\Lambda\left(\frac{r}{s}\right),\\Z_{23}&\ll\sum_{r\ll T}\sum_{s<r}\sum_{m|s,\,\mu|r}\frac{1}{r^{\sigma+1/2}s^{\sigma-1/2}}\frac{r}{s}\log\frac{2Tr}{s}\log\log\frac{3r}{s},\\Z_{24}&\ll\sum_{r\leq cT}\sum_{s<r}\sum_{m|s,\mu|r}\frac{\chi_1(m)\chi_2(s/m)\overline{\chi}_1(\mu)\overline{\chi}_2(r/\mu)}{r^{\sigma+1/2}s^{\sigma-1/2}}\log\frac{r}{s}\min\left(T,\frac{r/s}{\langle r/s\rangle}\right),\end{aligned}$$

and

$$Z_{25}\ll\sum_{r\ll T}\sum_{s<r}\sum_{m|s,\,\mu|r}\frac{1}{r^{\sigma+1/2}s^{\sigma-1/2}}\log 2T\min\left(T,\frac{1}{\log(r/s)}\right).$$

For $Z_{21,2}$ we set $r=sk$. Since $d(x)\ll x^{\epsilon}$ and $K\leq T$, we get

$$Z_{21,2}\ll T\sum_{k\ll T}\sum_{s\ll T/k}\frac{\Lambda(k)k^{\epsilon}s^{\epsilon}}{k^{\sigma+1/2}s^{2\sigma}}=O(T).$$

We also have

$$
\begin{aligned}
Z_{23} &\ll \log T \log\log T \sum_{r \ll T} \frac{r^{\epsilon}}{r^{\sigma-1/2}} \sum_{s<r} \frac{s^{\epsilon}}{s^{\sigma+1/2}} \\
&\ll \log T \log\log T \sum_{r \ll T} \frac{r^{\epsilon}}{r^{\sigma-1/2}} = o(N(T)).
\end{aligned}
$$

We can rewrite Z_{24} as

$$
\sum_{r \leq cT} \frac{\overline{(\chi_1 * \chi_2)}(r)}{r^{\sigma+1/2}} \sum_{s<r} \frac{(\chi_1 * \chi_2)(s)}{s^{\sigma-1/2}} \log \frac{r}{s} \min\left(T, \frac{r/s}{\langle r/s \rangle}\right), \tag{3.26}
$$

where $*$ denotes the Dirichlet convolution. Let $r = us + t$, where $-s/2 < t \leq s/2$, and separate the terms where u is not a prime power to $Z_{24,1}$, and denote the remaining terms by $Z_{24,2}$. We use (3.15) to see that

$$
Z_{24,1} \ll \sum_{s \leq cT} \sum_{u \ll cT/s+1} \sum_{|t|<s/2} \frac{s^{\epsilon}(us+t)^{\epsilon}}{s^{\sigma+1/2}(us+t)^{\sigma-1/2}} \log\left(u + \frac{t}{s}\right). \tag{3.27}
$$

Rewriting yields

$$
Z_{24,1} \ll \log T \sum_{s \leq cT} \frac{s^{2\epsilon}}{s^{2\sigma}} \sum_{u \ll cT/s+1} \sum_{|t|<s/2} \left(u + \frac{t}{s}\right)^{1/2-\sigma+\epsilon}.
$$

The terms in u can be bound from above by $(u-1)^{1/2-\sigma+\epsilon}$. Thus

$$
\begin{aligned}
Z_{24,1} &\ll \log T \sum_{s \leq cT} s^{1+2\epsilon-2\sigma} \left(\frac{cT}{s}\right)^{3/2-\sigma+\epsilon} \\
&\ll T^{3/2-\sigma+\epsilon} T^{1/2-\sigma+\epsilon} \log T = O(N(T)).
\end{aligned}
$$

For $Z_{24,2}$ let $'$ in summation denote that the sum extends only over prime powers. We need to estimate

$$
\sum_{s \leq cT} \sum_{u \leq \lfloor \frac{cT}{s} \rfloor + 1}' \sum_{0 \neq |t| < s/2} \frac{\overline{(\chi_1 * \chi_2)}(us+t)(\chi_1 * \chi_2)(s)}{(us+t)^{\sigma+1/2} s^{\sigma-1/2}} \log\left(u + \frac{t}{s}\right) \min\left(T, \frac{us+t}{|t|}\right)
$$

as $O(N(T))$. This can be rewritten as

$$
\sum_{s \leq cT} \frac{(\chi_1 * \chi_2)(s)}{s^{2\sigma}} \sum_{u \leq \lfloor \frac{cT}{s} \rfloor + 1}' \sum_{0 \neq |t| < s/2} \frac{\overline{(\chi_1 * \chi_2)}(us+t)}{(u + \frac{t}{s})^{\sigma+1/2}} \log\left(u + \frac{t}{s}\right) \min\left(T, \frac{us+t}{|t|}\right).
$$

By taking absolute values and using the triangle inequality we find that

$$\begin{aligned} Z_{24,2} &\ll \log^2 T \sum_{s \ll T} s^{2\epsilon - 2\sigma + 1} \sum_{u \ll T/s} u^{1/2 - \sigma + \epsilon} \\ &\ll T^{3/2 - \sigma + \epsilon} \log^2 T \sum_{s \ll T} s^{\epsilon - \sigma - 1/2} = O(N(T)), \end{aligned}$$

as required. It remains to estimate Z_{25}. We use the same method as in Proposition 3.6. Let $s = r - k$, and $1 \le k < r$ to get

$$\begin{aligned} Z_{25} &\ll \log T \sum_{r \ll T} \sum_{k < r} \frac{1}{r^{\sigma + 1/2 - \epsilon} (r - k)^{\sigma - 1/2 - \epsilon}} \frac{r}{k} \\ &\ll \log T \sum_{r \ll T} \frac{1}{r^{\sigma - 1/2 - \epsilon}} \sum_{k < r} \frac{1}{k} = o(N(T)). \end{aligned}$$

Finally, if $s > r$ we can consider the complex conjugate of (3.23) to obtain the same estimate. The rest of the proof proceeds in the same way as in Proposition 3.6. We obtain trivially the estimates

$$\sum_{0 < \gamma \le T} |W_1|^4 \ll N(T), \tag{3.28}$$

$$\sum_{0 < \gamma \le T} |Y_1|^4 \ll N(T). \tag{3.29}$$

Also, by modifying the argument slightly we find that

$$\sum_{0 < \gamma \le T} |W_2|^4 \ll T^{2\sigma - 1 + \epsilon} N(T), \tag{3.30}$$

$$\sum_{0 < \gamma \le T} |Y_2|^4 \ll T^{2\sigma - 1 + \epsilon} N(T). \tag{3.31}$$

We also need to estimate the derivative of $|X(s, \chi)|^4$. By estimate (3.21) from Proposition 3.6 we get

$$\frac{d}{d\gamma} |X(s, \chi)|^4 \ll \gamma^{1 - 2\sigma} \gamma^{-2\sigma} \ll \gamma^{1 - 4\sigma}.$$

The rest of the proof now follows from estimating

$$\sum_{0 < \gamma \le T} |X(s, \chi_1)|^4 |W_2|^4 = O(T^{1 - 2\sigma + \epsilon} N(T)),$$

and similarly for Y_2, and applying the Cauchy–Schwarz inequality to the remaining terms in the expansion of the product in (3.22). □

3.1.3 Proof of Proposition 3.8

Let

$$D=\sum_{n=1}^{\infty}\frac{d_n\overline{\chi}_2(n)}{n^{2\sigma}},\quad E=\sum_{n=1}^{\infty}\frac{e_n\overline{\chi}_1(n)}{n^{2\sigma}}.$$

By Proposition 3.7 it is sufficient to show that $D-E\neq 0$. First, we need to compute the d_n and e_n's explicitly. Let us denote the set of primes smaller than P by $\mathscr{P}=\{p_1,p_2,\ldots,p_h,P\}$. Suppose P is large enough so that $q,\ \ell\in\mathscr{P}$. The coefficients d_n are defined by the Euler product

$$\prod_{p\le P}(1-\chi_2(p)p^{-s})\times\prod_{p>P}\sum_{n=0}^{\infty}\frac{\chi_1(p^n)}{p^{ns}}.$$

If $p^2\mid n$, $p\in\mathscr{P}$, then n disappears from the expansion, i.e. $d_n=0$. If n has no prime factors from the set $\mathscr{P}$, then we just get the usual coefficient from $L(s,\chi_1)$. On the other hand, if some prime $p\in\mathscr{P}$ divides n exactly once then it contributes $-\chi_2(p)$. Hence

$$d_n=\begin{cases}\chi_1(n), & \text{if } p\nmid n \text{ for all } p\in\mathscr{P},\\ (-1)^k\chi_1\big(\frac{n}{p_{i_1}\cdots p_{i_k}}\big)\chi_2(p_{i_1}\cdots p_{i_k}), & \text{if } p_{i_j}\parallel n \text{ for } p_{i_j}\in\mathscr{P} \text{ for all } j,\\ 0 & \text{otherwise.}\end{cases}$$

Similarly for e_n. Here $p^\alpha\parallel n$ denotes that p^α is the largest power of p dividing n. Hence for $p>P$ the Euler factors of D are of the form

$$(1-\chi_1(p)\overline{\chi}_2(p)p^{-2\sigma})^{-1},$$

while for E one obtains the complex conjugate. On the other hand, for $p\le P$ we have

$$1+d_p\overline{\chi}_2(p)p^{-2\sigma}+d_{p^2}\overline{\chi}_2(p^2)p^{-4\sigma}+\cdots=1-\chi_2\overline{\chi}_2(p)p^{-2\sigma}=1-p^{-2\sigma},$$

unless $p=\ell$, and similarly for the second series. Now, suppose that $D=E$, then

$$\prod_{\substack{p\le P\\ p\neq\ell}}(1-p^{-2\sigma})\prod_{p>P}(1-(\chi_1\overline{\chi}_2)(p)p^{-2\sigma})^{-1}=\prod_{\substack{p\le P\\ p\neq q}}(1-p^{-2\sigma})\prod_{p>P}(1-(\overline{\chi}_1\chi_2)(p)p^{-2\sigma})^{-1}.$$

We cancel out the common terms in the product over $p<P$, which yields

$$(1-q^{-2\sigma})z=(1-\ell^{-2\sigma})\overline{z},$$

where

$$z = \prod_{p>P} (1-(\chi_1\overline{\chi}_2)(p)p^{-2\sigma})^{-1}.$$

Hence,

$$\frac{1-q^{-2\sigma}}{1-\ell^{-2\sigma}} = \frac{\overline{z}}{z}.$$

Taking absolute values yields

$$\frac{1-q^{-2\sigma}}{1-\ell^{-2\sigma}} = 1,$$

which is a contradiction. □

3.2 Proof of Theorem 3.3

We now sample the values of $L(s,\chi)$ at precisely the non-trivial zeros of ζ. In this case we do not assume the RH. Off the critical line we used the method of Gonek–Landau to prove linear independence. On the critical line, however, this becomes very difficult. This is mainly because of the corresponding Z_{24} term in the first proposition. We get

$$Z_{24} = \sum_{n\leq X} \sum_{m<n} \frac{\chi(m/n)}{n} \log\left(\frac{n}{m}\right) \min\left(T, \frac{n/m}{\langle n/m \rangle}\right),$$

where $X = qT/2\pi\sqrt{\log T}$. This should be $o(N(T)\log T)$, which seems to be very difficult to prove. In the proof of Conrey, Ghosh, and Gonek [21] they make a reduction to the discrete mean values of one L-function at a time. We have been unable to find such a reduction in our case. Garunkštis, Kalpokas, and Steuding [29] presented a more suitable method through contour integration and a modified Gonek Lemma (see Lemma 3.5).

Denote the characters in Theorem 3.3 by χ_1 and χ_2 with distinct prime moduli q and ℓ. For any Dirichlet character ω modulo n, we denote the principal character modulo n by ω_0. Moreover, put $B(s,p) = p^s$ for some prime p to be determined later. Then,

$$A(\gamma) := B(\rho,p)(L(\rho,\chi_1) - L(\rho,\chi_2))$$

is non-zero precisely when the two L-functions assume distinct values. Let $G(k,\chi)$ be the generalised Gauß sum as defined in (2.11).

Proposition 3.9. *Let $\mathfrak{C}$ be the rectangular contour with vertices at $a+i$, $a+iT$, $1-a+iT$,* *and* $1-a+i$ *with positive orientation, where* $a=1+(\log T)^{-1}$*. Then we have*

$$\frac{1}{2\pi i}\int_{\mathfrak{C}}\frac{\zeta'}{\zeta}(s)B(s,p)L(s,\chi_1)\,ds\sim\frac{\overline{C}_{\chi_1}T}{2\pi}\log\frac{T}{2\pi},\tag{3.32}$$

where

$$C_{\chi_1}=\frac{G(1,\overline{\chi}_1)G(-p,\chi_1)}{q},$$

and similarly for χ_2*.*

Then, by the residue theorem, we get

$$\sum_{0<\gamma\le T}A(\gamma)=\frac{1}{2\pi i}\int_{\mathfrak{C}}\frac{\zeta'}{\zeta}(s)B(s,p)(L(s,\chi_1)-L(s,\chi_2))\,ds.$$

Proposition 3.10. *With the same contour as in Proposition 3.9, we have for* $j,j'\in\{1,2\}$ *that*

$$\frac{1}{2\pi i}\int_{\mathfrak{C}}\frac{\zeta'}{\zeta}(s)L(s,\chi_j)L(1-s,\overline{\chi}_{j'})\,ds\ll T\log^2 T.$$

This proposition gives us estimates for all the terms in $\sum|A(\gamma)|^2$, since $B(s,p)$ can be bound independently of T. Finally, we have to prove that the difference coming from Proposition 3.9 is non-zero.

Proposition 3.11. *There is a prime* p*, different from* q *and* ℓ*, such that* $C_{\chi_1}-C_{\chi_2}\neq 0$*.*

With these propositions we can prove Theorem 3.3 in the same way as in (3.10).

3.2.1 Proof of Proposition 3.9

We prove the proposition for χ_1 as the case of χ_2 is identical. Denote the integral in (3.32) by $\mathscr{I}$. Then

$$\begin{aligned}\mathscr{I}&=\left(\int_{a+i}^{a+iT}+\int_{a+iT}^{1-a+iT}+\int_{1-a+iT}^{1-a+i}+\int_{1-a+i}^{a+i}\right)\frac{\zeta'}{\zeta}(s)B(s,p)L(s,\chi_1)\,ds\\&=\mathscr{I}_1+\mathscr{I}_2+\mathscr{I}_3+\mathscr{I}_4.\end{aligned}$$

We can evaluate $\mathscr{I}_1$ explicitly to get

$$\begin{aligned}\mathscr{I}_1 &= \int_{a+i}^{a+iT} \frac{\zeta'}{\zeta}(s)L(s,\chi_1)p^s\,ds\\ &= -i\sum_{n,m}\frac{\Lambda(n)\chi_1(m)}{(mnp^{-1})^a}\int_1^T\left(\frac{p}{mn}\right)^{it}dt\\ &\ll \frac{\zeta'}{\zeta}(a)\zeta(a)+T = O(T),\end{aligned}$$

where the second term comes from the case $mn=p$. For $\mathscr{I}_2$ we use the following bounds (see [22, pg. 108]):

$$\frac{\zeta'}{\zeta}(\sigma+iT)\ll\log^2 T,\quad \text{if } -1\leq\sigma\leq 2,\quad |T|\geq 1, \tag{3.33}$$

and

$$L(\sigma+iT,\chi_1)\ll|T|^{1/2}\log|T+2|,\quad \text{if } 1-a\leq\sigma\leq a,\quad |T|\geq 1. \tag{3.34}$$

These yield $\mathscr{I}_2=O(T^{1/2}\log^3 T)$. Next we consider $\mathscr{I}_3$. Changing variables $s\mapsto 1-\overline{s}$ gives

$$\mathscr{I}_3=\frac{-1}{2\pi i}\int_{a+i}^{a+iT}\frac{\zeta'}{\zeta}(1-\overline{s})L(1-\overline{s},\chi_1)p^{1-\overline{s}}\,ds.$$

Conjugating and applying the functional equation (2.3) of ζ and (2.17) of $L(s,\chi_1)$ yields

$$\overline{\mathscr{I}}_3=\frac{p}{2\pi i}\int_{a+i}^{a+iT}\left(\frac{\zeta'}{\zeta}(s)+\frac{\gamma'}{\gamma}(s)\right)L(s,\chi_1)\Delta(s,\chi_1)p^{-s}\,ds,$$

where

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

and

$$\Delta(s,\chi_1)=\left(\frac{q}{2\pi}\right)^s\frac{1}{q}G(1,\overline{\chi}_1)\Gamma(s)\left(e^{-\pi is/2}+\overline{\chi}_1(-1)e^{\pi is/2}\right).$$

Using the definition of Δ to expand the above we find that

$$\overline{\mathscr{I}}_3=p\,(\mathscr{F}_1+\cdots+\mathscr{F}_4),$$

where

$$\mathscr{F}_1 = \frac{G(1,\overline{\chi}_1)}{q}\frac{1}{2\pi i}\int_{a+i}^{a+iT}\frac{\gamma'}{\gamma}(s)\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds,$$
$$\mathscr{F}_2 = \frac{\overline{\chi}_1(-1)G(1,\overline{\chi}_1)}{q}\frac{1}{2\pi i}\int_{a+i}^{a+iT}\frac{\gamma'}{\gamma}(s)\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(+\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds,$$
$$\mathscr{F}_3 = \frac{G(1,\overline{\chi}_1)}{q}\frac{1}{2\pi i}\int_{a+i}^{a+iT}\frac{\zeta'}{\zeta}(s)\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds,$$
$$\mathscr{F}_4 = \frac{\overline{\chi}_1(-1)G(1,\overline{\chi}_1)}{q}\frac{1}{2\pi i}\int_{a+i}^{a+iT}\frac{\zeta'}{\zeta}(s)\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(+\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds.$$

We rewrite $\mathscr{F}_1$ in the following way

$$\mathscr{F}_1 = \frac{G(1,\overline{\chi}_1)}{q}\int_1^T\frac{\gamma'}{\gamma}(a+i\tau) \times d\left(\frac{1}{2\pi i}\int_{a+i}^{a+i\tau}\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds\right). \tag{3.35}$$

By Lemma 3.5,

$$\frac{1}{2\pi i}\int_{a+i}^{a+i\tau}\left(\frac{q}{2\pi p}\right)^s\Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)L(s,\chi_1)\,ds = \sum_{n\le\frac{\tau q}{2\pi p}}\chi_1(n)\exp\left(-2\pi i\frac{np}{q}\right)+O(\tau^{1/2+\epsilon}).$$

We can separate the periods to write the sum as

$$\sum_{a=1}^{q}\chi_1(a)\exp\left(-2\pi i\frac{ap}{q}\right)\sum_{\substack{n\le\frac{\tau q}{2\pi p}\\ n\equiv a \bmod q}}1 = \frac{\tau}{2\pi p}G(-p,\chi_1)+O(1).$$

We integrate by parts in (3.35) and use the standard estimate

$$\frac{\gamma'}{\gamma}(s) = \log\frac{|t|}{2\pi}+O(|t|^{-1}),\quad |t|\ge 1,$$

which follows from Stirling asymptotics. Thus

$$\begin{aligned}\mathscr{F}_1 &= \frac{C_{\chi_1}}{2\pi p}\int_1^T\left(\log\frac{\tau}{2\pi}+O(\tau^{-1})\right)d\left(\tau+O(\tau^{1/2+\epsilon})\right)\\ &= \frac{C_{\chi_1}T}{2\pi p}\log\frac{T}{2\pi}+O(T^{1/2+\epsilon}).\end{aligned}$$

Similarly by Lemma 3.5, $\mathscr{F}_2$ is $O(\log T)$, while $\mathscr{F}_4 = O(1)$. For $\mathscr{F}_3$ we have

$$\begin{aligned}\mathscr{F}_3 &= \frac{G(1,\overline{\chi}_1)}{q}\frac{1}{2\pi i}\int_{a+i}^{a+iT}\left(\frac{q}{2\pi p}\right)^s \Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)\frac{\zeta'}{\zeta}(s)L(s,\chi_1)\,ds \\ &= \frac{-G(1,\overline{\chi}_1)}{q}\sum_{mn\le\frac{Tq}{2\pi p}}\Lambda(m)\chi_1(n)\exp\left(-2\pi i\frac{mnp}{q}\right) + O(T^{1/2+\epsilon}).\end{aligned}$$

Looking at the summation we decompose it as

$$\sum_{m\le\frac{Tq}{2\pi p}}\Lambda(m)\sum_{n\le\frac{Tq}{2\pi pm}}\chi_1(n)\exp\left(-2\pi i\frac{mnp}{q}\right).$$

We separate the periods in the same way as for $\mathscr{F}_1$ and write the above sum as

$$\sum_{m\le\frac{Tq}{2\pi p}}\Lambda(m)G(-mp,\chi_1)\frac{T}{2\pi pm} + O(T).$$

We will show that the summation over m in fact converges. This means that we have $\mathscr{F}_3 = O(T)$. To do this it suffices to consider

$$\sum_{m\le X}\frac{\Lambda(m)\overline{\chi}_1(m)}{m}.$$

Let

$$\psi(X,\overline{\chi}_1) = \sum_{m\le X}\Lambda(m)\overline{\chi}_1(m).$$

Then, by [22, pg. 123 (8)],

$$\psi(X,\overline{\chi}_1) = -\frac{X^\beta}{\beta} + O(X\exp(-c(\log X)^{1/2})),$$

where the term with β comes from the Siegel zero of χ_1 and c is some positive absolute constant. However, since our q is fixed, we know that β is bounded away from 1. Hence, with summation by parts we obtain

$$\sum_{m\le X}\frac{\Lambda(m)\overline{\chi}_1(m)}{m} = \frac{\psi(X,\overline{\chi}_1)}{X} + \int_1^X\frac{\psi(t,\overline{\chi}_1)}{t^2}\,dt = O(1)$$

as required. Finally $\mathscr{I}_4 = O(1)$ as the integrand is analytic in a neighbourhood of the line of integration. □

3.2.2 Proof of Proposition 3.10

We prove the case $j = j' = 1$ as the other cases are either similar or easier. Now, denote the integral by $\mathscr{I}$, i.e.

$$\mathscr{I} = \frac{1}{2\pi i}\int_{\mathfrak{C}} \frac{\zeta'}{\zeta}(s)L(s,\chi_1)L(1-s,\overline{\chi}_1)\,ds,$$

and split it in the same way as in the proof of Proposition 3.9, so that

$$\mathscr{I} = \mathscr{I}_1 + \cdots + \mathscr{I}_4.$$

We can write $\mathscr{I}_1$ as

$$\begin{aligned}\mathscr{I}_1 &= \frac{1}{2\pi i}\int_{a+i}^{a+iT} \frac{\zeta'}{\zeta}(s)L(s,\chi_1)^2\Delta(s,\chi_1)\,ds\\ &= \frac{G(\overline{\chi}_1)}{2\pi i q}\int_{a+i}^{a+iT}\left(\frac{q}{2\pi}\right)^s\Gamma(s)\left(\exp\left(\frac{-\pi i s}{2}\right)+\overline{\chi}_1(-1)\exp\left(\frac{\pi i s}{2}\right)\right)\frac{\zeta'}{\zeta}(s)L(s,\chi_1)^2\,ds\\ &= \mathscr{E}_1 + \mathscr{E}_2.\end{aligned}$$

By Lemma 3.5, $\mathscr{E}_2 = O(1)$. Let us now estimate $\mathscr{E}_1$. We have

$$\mathscr{E}_1 = -\frac{G(\overline{\chi}_1)}{q}\sum_{mn\le\frac{Tq}{2\pi}} d(n)\chi_1(n)\Lambda(m)\exp\left(-2\pi i\frac{nm}{q}\right) + O(T^{1/2+\epsilon}).$$

Denote the sum over m and n by S. As before, we first separate the periods

$$S = \sum_{a,b=1}^{q}\chi_1(a)\exp\left(-2\pi i\frac{ab}{q}\right)\sum_{\substack{mn\le\frac{Tq}{2\pi}\\ n\equiv a \bmod q\\ m\equiv b \bmod q}} d(n)\Lambda(m).$$

Now sum over characters η of modulus q to get

$$\begin{aligned}S &= \frac{1}{\varphi(q)}\sum_{\eta \bmod q}\sum_{a,b=1}^{q}\chi_1(a)\overline{\eta}(a)\exp\left(-2\pi i\frac{ab}{q}\right)\sum_{\substack{mn\le\frac{Tq}{2\pi}\\ m\equiv b \bmod q}} d(n)\eta(n)\Lambda(m)\\ &= \frac{1}{\varphi(q)}\sum_{\eta \bmod q}\sum_{b=1}^{q} G(-b,\chi_1\overline{\eta})\sum_{\substack{mn\le\frac{Tq}{2\pi}\\ m\equiv b \bmod q}} d(n)\eta(n)\Lambda(m),\end{aligned}$$

and as before

$$=\frac{1}{\varphi(q)^2}\sum_{\eta,\omega \bmod q} G(-1,\chi_1\overline{\eta})\left(\sum_{mn\le\frac{Tq}{2\pi}} d(n)\eta(n)\Lambda(m)\omega(m)\right)\sum_{b=1}^{q}\overline{\chi}_1(b)\eta(b)\overline{\omega}(b).$$

The sum over b is non-zero if and only if $\omega=\omega_0$ and $\eta=\chi_1$; or $\omega=\overline{\chi_1}$ and $\eta=\eta_0$; or $\omega\neq\omega_0$ and $\eta=\chi_1\omega$. By Perron's formula (see [76, A2.5])

$$-\sum_{mn\le\frac{Tq}{2\pi}} d(n)\eta(n)\Lambda(m)\omega(m)=\frac{1}{2\pi i}\int_{a-iU}^{a+iU}\frac{L'}{L}(s,\omega)L(s,\eta)^2\left(\frac{Tq}{2\pi}\right)^s\frac{ds}{s}+O\left(\frac{T\log^3 T}{U}\right)$$

for some U with $|U|\le T$. Since our characters are fixed, we can use a Vinogradov-type zero-free region [29, pg. 296]. That is, let $b_1=1-c_1/(\log t)^{3/4+\epsilon}$ (in fact, any power smaller than 1 would do), then $L(\sigma+it,\chi_1)$ has no zeros in the region $\sigma\ge b_1$. Here c_1 is some positive absolute constant. By the approximate functional equation (2.18) and Stirling asymptotics (A.6) we have uniformly for $0<\sigma<1$ and $|t|>1$ that

$$L(\sigma+it,\chi_1)\ll|t|^{\frac{1-\sigma}{2}}\log(|t|+1). \tag{3.36}$$

Then, by shifting the contour we get

$$\begin{aligned}-\sum_{mn\le\frac{Tq}{2\pi}} d(n)\eta(n)\Lambda(m)\omega(m)&=\operatorname*{res}_{s=1}\frac{L'}{L}(s,\omega)L(s,\eta)^2\left(\frac{Tq}{2\pi}\right)^s\frac{1}{s}\\&-\frac{1}{2\pi i}\left(\int_{a+iU}^{b_1+iU}+\int_{b_1+iU}^{b_1-iU}+\int_{b_1-iU}^{a-iU}\right)\frac{L'}{L}(s,\omega)L(s,\eta)^2\left(\frac{Tq}{2\pi}\right)^s\frac{ds}{s}+O\left(\frac{T\log^3 T}{U}\right).\end{aligned} \tag{3.37}$$

We need to find the residues in each of the three cases.

$$\begin{aligned}&\operatorname*{res}_{s=1}\frac{L'}{L}(s,\omega_0)L(s,\chi_1)^2\left(\frac{Tq}{2\pi}\right)^s\frac{1}{s}=-L(1,\chi_1)^2\frac{Tq}{2\pi},\\&\operatorname*{res}_{s=1}\frac{L'}{L}(s,\overline{\chi}_1)L(s,\eta_0)^2\left(\frac{Tq}{2\pi}\right)^s\frac{1}{s}=\frac{Tq}{2\pi}\left(\frac{\varphi(q)}{q}\right)^2\frac{L'}{L}(1,\overline{\chi}_1)\log\frac{Tq}{2\pi}+O(T),\\&\operatorname*{res}_{s=1}\frac{L'}{L}(s,\omega)L(s,\chi_1\omega)^2\left(\frac{Tq}{2\pi}\right)^s\frac{1}{s}=0.\end{aligned}$$

It remains to estimate the integrals on the right-hand side of (3.37). By (3.33) and (3.36) we see that the first and third integrals yield $O(T^aU^{-b_1}\log^4 U)$. We split the second integral and estimate it as

$$\left(\int_{b_1+iU}^{b_1+i} + \int_{b_1+i}^{b_1-i} + \int_{b_1-i}^{b_1-iU}\right) \frac{L'}{L}(s,\overline{\chi}_1)L(s,\eta_0)^2 \left(\frac{Tq}{2\pi}\right)^s \frac{ds}{s}$$
$$= O(T^{b_1}U^{1-b_1}\log^4 U + T^{b_1}|b_1-1|^{-2}),$$

where the second error term comes from the integral over the constant segment. It suffices to choose $U = T^{1/2}$ as then

$$T^a U^{-b_1}\log^4 U \ll Te^{-\frac{1}{2}\log T + \frac{c_1}{2}(\log T)^{1/4-\epsilon}+4\log\log T},$$
$$T^{b_1}U^{1-b_1}\log^4 U \ll Te^{-\frac{c_1}{2}(\log T)^{1/4-\epsilon}+4\log\log T},$$

and

$$T^{b_1}|b_1-1|^{-2} \ll Te^{-c_1(\log T)^{1/4-\epsilon}+(3/2+2\epsilon)\log\log T},$$

which are all $O(T)$. Therefore we conclude that $\mathscr{I}_1 = O(T\log T)$.

Next up is $\mathscr{I}_2$. We use the convexity bound (2.20) adapted for Dirichlet L-functions, which yields

$$L(\sigma+it,\chi_1) \ll_\epsilon |t|^{\mu_0(\sigma)+\epsilon},$$

where $\epsilon > 0$, $-1 < \sigma < 2$ (say) and $|t| > 1$. With this we can write

$$\mathscr{I}_2 = \frac{1}{2\pi i}\int_{a+iT}^{1-a+iT} \frac{\zeta'}{\zeta}(s)L(s,\chi_1)L(1-s,\overline{\chi}_1)\,ds$$
$$\ll \left(\int_{1-a}^{0} + \int_{0}^{1} + \int_{1}^{a}\right)\log^2 T\, T^{\mu_0(\sigma)+\epsilon}T^{\mu_0(1-\sigma)+\epsilon}\,d\sigma.$$

Keeping in mind that $\sigma \le a$ we get

$$\mathscr{I}_2 \ll T^{a-1/2+\epsilon}\log^2 T + T^{1/2+\epsilon}\log^2 T = O(T).$$

For $\mathscr{I}_3$ we do the usual trick of mapping $s \mapsto 1-\overline{s}$. Taking complex conjugates leads to

$$\overline{\mathscr{I}}_3 = \frac{1}{2\pi i}\int_{a+i}^{a+iT}\left(\frac{\zeta'}{\zeta}(s)+\frac{\gamma'}{\gamma}(s)\right)L(s,\chi_1)^2\Delta(s,\chi_1)\,ds.$$

As in Proposition 3.9 we split this up into $\mathscr{F}_1,\ldots,\mathscr{F}_4$. Adding up $\mathscr{F}_3$ and $\mathscr{F}_4$ gives $\mathscr{I}_1$, which is $O(T\log T)$. As before, $\mathscr{F}_2$ does not contribute. So we have to estimate $\mathscr{F}_1$, that is,

$$\mathscr{F}_1 = \frac{G(1,\overline{\chi}_1)}{q}\int_1^T \left(\log\frac{\tau}{2\pi} + O(\tau^{-1})\right) d\left(\frac{1}{2\pi i}\int_{a+i}^{a+i\tau} L(s,\chi_1)^2\Gamma(s)\exp\left(-\frac{\pi i s}{2}\right)ds\right).$$

Working as in Proposition 3.9 we can write the inner integral (plus an error term) as

$$\sum_{n \le \frac{\tau q}{2\pi}} \chi_1(n) d(n) \exp\left(-2\pi i \frac{n}{q}\right).$$

This is $O(\tau \log \tau)$, which gives $\mathscr{I}_3 = O(T \log^2 T)$. It is not difficult to extend this to an asymptotic estimate, but for our purposes the upper bound is sufficient. Trivially we also have that $\mathscr{I}_4 = O(1)$. Hence $\mathscr{I} = O(T \log^2 T)$. □

3.2.3 Proof of Proposition 3.11

By (2.14) and (2.12) we see that $C_{\chi_1} = C_{\chi_2}$ if and only if $\chi_1(p) = \chi_2(p)$. By the Chinese Remainder Theorem and Dirichlet's theorem for primes in arithmetic progressions we can find a prime p different from q and ℓ that satisfies $p \equiv 1 \bmod q$ and $p \equiv a \bmod \ell$, as well as $\chi_2(p) = \chi_2(a) \neq 1$, since χ_2 is non-principal. This gives $1 = \chi_1(p) = \chi_2(p) \neq 1$, which is a contradiction. □

Part II

Automorphic Forms

Chapter 4

Spectral Theory in Hyperbolic Space

In this chapter we give an introduction and the statement of main results in the spectral theory of automorphic forms in hyperbolic spaces. We split the treatment of hyperbolic spaces in two and three dimensions to Sections 4.2 and 4.4, respectively. Since discrete groups acting on the hyperbolic 3-space arise from the ring of integers of imaginary quadratic fields, we give an introduction to this theory in Section 4.3.

4.1 Historical Motivation

Let us first motivate the introduction of non-holomorphic automorphic forms with some historical remarks. Non-holomorphic automorphic forms, in particular Maaß forms and Eisenstein series, are generalisations of holomorphic modular forms, where the analyticity condition is replaced with the requirement of being an eigenfunction of the Laplace operator. It turns out that certain modular (and automorphic) forms are deeply connected to L-functions. Hecke [43] [4, Theorem 2.1] showed that given a Dirichlet series

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

with polynomially bounded coefficients and a function $f:\mathbb{H}^2\longrightarrow\mathbb{C}$ with Fourier series

$$f(z)=a_0+\sum_{n=1}^{\infty}a_n e^{2\pi i n z/q},$$

then f is a modular form of weight k and level $q\geq 2$ for the group generated by $z\mapsto -1/z$ and $z\mapsto z+q$ if and only if for $\Phi(s)=(2\pi/q)^{-s}\Gamma(s)L(f,s)$ we have that

H1 $\Phi(s)+a_0/s+\gamma/(k-s)$ has an analytic continuation to $\mathbb{C}$ as an entire function, where $\gamma=\pm1$, and it is bounded in vertical strips;

H2 Φ satisfies a functional equation $\Phi(s)=\gamma\Phi(k-s)$.

Notice that with our normalisation the critical line is at $\operatorname{Re} s=k/2$. If the coefficients a_n are completely multiplicative, it follows that $L(f,s)$ also has an Euler product in the usual sense (degree one, as in Part I). This is not surprising for ζ and $L(s,\chi)$ as their analytic continuation, and the functional equation, can be derived by writing them as inverse Mellin transforms of modular forms. Recall our proof of the analytic continuation of ζ in Lemma 2.3. The theta series introduced there,

$$\theta(z)=1+\sum_{n=1}^{\infty}2e^{\pi i n^2 z}$$

is actually a weight $1/2$ modular form for the congruence group $\Gamma_0(2)$ (see (4.3)), [31, pg. 185]. Then

$$\theta(z)^2=\sum_{n=0}^{\infty}r(n)e^{\pi i n z},$$

where $r(n)$ is the number of ways of writing n as a sum of two squares. The series θ^2 is a weight 1 form [54, §14.3]. Similarly we could associate Dirichlet L-functions with modular forms by introducing twists by Dirichlet characters [54, pg. 369, 31, Theorem 1]. It is possible to also give theta series in terms of many other quadratic forms that one is interested in number theory (for example, the Gauß circle problem corresponds exactly to the function $r(n)$). These would yield for example the Dedekind zeta function and its twists.

Hecke went much further than this. For the modular group $\mathrm{SL}_2(\mathbb{Z})$, he fully characterised the modular forms f of weight k corresponding to L-series which also have an Euler product (in addition to the aforementioned properties **H1** and **H2**) of the form

$$L(f,s)=\prod_p L_p(f,s),$$

where $L_p(f,s)$ is called the *Euler factor* of L at p. Actually, the Euler factor is of degree two and can be written as

$$L_p(f,s)=(1-a_p p^{-s}+p^{k-1-2s})^{-1}.$$

The crucial idea was to require that the modular forms are also eigenfunctions of certain

operators T_p with
$$T_p f = a_p f.$$
The operators T_p average f over multiple lattices and guarantee that the coefficients a_p are multiplicative. Nowadays we call T_p the Hecke operators. We will give precise definitions in three dimensions in Section 4.4. All of this generalises to the case when we replace modular forms with automorphic forms. In Chapter 7 one of the crucial ingredients is the behaviour of degree two L-functions associated to the Hecke–Maaß cusp forms. On the other hand, we will also see how the Fourier coefficients of the generalised eigenfunctions of Δ, the Eisenstein series, factorise in terms of the Dedekind zeta function. This theme of matching modular/automorphic forms with L-functions with nice properties was the precursor to the Langlands program and the functoriality conjectures. It is a firm belief that all nice L-functions arise from suitable automorphic forms [6].

4.2 Hyperbolic Plane

4.2.1 Hyperbolic Geometry in Two Dimensions

Most of the basic facts about hyperbolic geometry in this introduction can be found for example in [53, 57] and so will mostly be used without explicit reference. Denote the complex upper half-plane by
$$\mathbb{H}^2 = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}.$$
Equipped with the metric
$$ds^2 = \frac{dx^2 + dy^2}{y^2} \tag{4.1}$$
the upper half-plane becomes a model for the hyperbolic plane, which is a two dimensional Riemannian manifold of constant negative curvature -1. We can compute lengths on $\mathbb{H}^2$ by integrating the line element (4.1). We see that the geodesics on $\mathbb{H}^2$ are Euclidean straight lines and semicircles orthogonal to the Euclidean boundary $\widehat{\mathbb{R}}$. It is then possible to compute that the hyperbolic distance d on $\mathbb{H}^2$ is given by
$$\cosh d(z,w) = 1 + 2u(z,w),$$

where the function $u:\mathbb{H}^2\times\mathbb{H}^2\longrightarrow\mathbb{R}^+$ is called a *point-pair invariant*, and it has the explicit form

$$u(z,w)=\frac{|z-w|^2}{4\operatorname{Im} z\operatorname{Im} w}.$$

Notice that $u(z,w)$ is strictly a function of the distance $d(z,w)$ only.

The orientation preserving isometries of $\mathbb{H}^2$ can be described by Möbius transformations. The group $G=\mathrm{SL}_2(\mathbb{R})$ with $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in G$ acts on $\mathbb{H}^2$ by

$$\gamma z=\frac{az+b}{cz+d},\qquad \text{for any } z\in\mathbb{H}^2. \tag{4.2}$$

This action is transitive in the sense that $Gz=\mathbb{H}^2$ for any $z\in\mathbb{H}^2$. By well-known properties of Möbius transformations it follows that this action maps geodesics to geodesics. We also note that $\frac{d}{dz}gz=(cz+d)^{-2}$, where the factor $|cz+d|^{-2}$ is called the *deformation* of g at z. The matrices g and $-g$ define the same transformations so we can consider $\mathrm{PSL}_2(\mathbb{R})=\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$. Actually, $\mathrm{PSL}_2(\mathbb{R})$ is isomorphic to the group of *orientation preserving* isometries of $\mathbb{H}^2$. Together with conjugation $z\mapsto-\overline{z}$, these generate all the isometries. In particular, the point-pair invariant is invariant under G, that is for any $g\in G$, $u(gz,gw)=u(z,w)$ for any pair of points $z,w\in\mathbb{H}^2$.

The G-invariant metric (4.1) gives rise to the volume measure

$$d\mu(z)=\frac{dx\,dy}{y^2},$$

which is also G-invariant. There is an interesting phenomenon of hyperbolic geometry that follows from the form of $d\mu(z)$. Consider a hyperbolic disc of radius r centered at z. The circumference of this disc is $L=2\pi\sinh r$, while its area is $A=4\pi\sinh^2\frac{r}{2}$, [53, pg. 11]. Hence, $L\sim A$ as $r\to\infty$. This is completely different from what happens in Euclidean geometry. The consequence is that it becomes impossible to use simple geometric arguments to obtain information about the error term in lattice point problems. Spectral analysis of automorphic forms on the other hand has yielded the strongest results known today.

We can classify the Möbius transformations depending on the way they act on $\mathbb{H}^2$. There are three different types of transformations: translations, dilations and rotations. These are called *parabolic*, *hyperbolic* and *elliptic*, respectively. We can also classify them in terms of their fixed points or the trace of the corresponding matrix $g\in G$. Any

$g \neq \pm I$ has at least one and at most two fixed points in $\widehat{\mathbb{H}}^2$.

$$\begin{aligned}
g \text{ is parabolic} &\iff g \text{ has one fixed point in } \widehat{\mathbb{R}} &&\iff |\operatorname{Tr} g| = 2,\\
g \text{ is hyperbolic} &\iff g \text{ has two distinct fixed points in } \widehat{\mathbb{R}} &&\iff |\operatorname{Tr} g| > 2,\\
g \text{ is elliptic} &\iff g \text{ has exactly one fixed point in } \mathbb{H}^2 &&\iff |\operatorname{Tr} g| < 2.
\end{aligned}$$

Finally, we can always conjugate the matrices in $\mathrm{SL}_2(\mathbb{R})$ so that parabolic, hyperbolic and elliptic transformations with given fixed points can be brought to the following forms

$$\begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}, \qquad \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}, \qquad \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

respectively. In the above we assume that $\lambda \neq 1$ and $\lambda > 0$. Also, isometries map geodesics to geodesics and in particular the geodesic joining the two fixed points of a hyperbolic element g is mapped to itself by g.

Remark 4.1. In anticipation of some of the more general discussion in the following chapters, we remark that the hyperbolic plane is also a rank one symmetric space. To see this let $G = \mathrm{SL}_2(\mathbb{R})$ and consider the maximal compact subgroup $K = SO(2)$ of G, which can be identified as the stabiliser of the point i. Hence, with the diffeomorphism $g \mapsto gi$ we have that $G/K \cong \mathbb{H}^2$. Since $\mathbb{H}^2$ has negative curvature, the rank of the symmetric space is one. In fact any hyperbolic n-space is a rank one symmetric space.

The Laplace–Beltrami operator Δ on $\mathbb{H}^2$ is derived from the metric and can be written as

$$\Delta = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right),$$

with the eigenvalue equation

$$\Delta u + \lambda u = 0,$$

so that the eigenvalues are non-negative. From now on we call Δ the (hyperbolic) Laplacian. It is an invariant differential operator in the sense that it commutes with the action of G on functions on $\mathbb{H}^2$:

$$\Delta(f(gz)) = (\Delta f)(gz).$$

It is easy to see that y^s and y^{1-s} are eigenfunctions of Δ with corresponding eigenvalues $\lambda = s(1-s) \geq 0$. We can also find eigenfunctions f that have a constant period in x.

These turn out to be multiples of the Whittaker function,

$$W_s(z) = 2y^{1/2}K_{s-1/2}(2\pi y)e^{2\pi i x}.$$

Here $K_\nu(y)$ is the K-Bessel function [53, B.4].

In order to introduce arithmeticity we look at quotients of $\mathbb{H}^2$ by discrete groups of isometries, that is, $\Gamma \subset G$ which are discrete under the usual topology on $\mathbb{R}^4$. These groups (considered as subgroups of $\mathrm{PSL}_2(\mathbb{R})$ now) act discontinuously on $\mathbb{H}^2$, meaning that none of the orbits Γz accumulate in $\mathbb{H}^2$ for any $z \in \mathbb{H}^2$. We call such Γ a *Fuchsian group*. We classify these groups depending on whether every point on the boundary $\widehat{\mathbb{R}}$ is a limit point of Γz for some $z \in \mathbb{H}^2$, or whether the limit points are nowhere dense on $\widehat{\mathbb{R}}$. Then Γ is of the *first kind* or of the *second kind*, respectively. We will only work with Fuchsian groups of the first kind. A more practical, and equivalent, characterisation is that Γ is of the first kind if and only if it has finite covolume. Therefore such groups are also sometimes called *finite volume groups*. Each Γ has a fundamental domain $F \subset \mathbb{H}^2$ which is defined by the following properties

(i) $F \neq \emptyset$ and F is closed,

(ii) $\bigcup_{\gamma \in \Gamma} \gamma F = \mathbb{H}^2$,

(iii) $F^\circ \bigcap (\gamma F)^\circ = \emptyset$, for any $\gamma \neq I$,

where F° denotes the interior of F. The sides of F are geodesic segments and they form a hyperbolic n-gon where n is always even. We will work with fundamental domains in more detail in Chapter 6.

We further classify the finite volume groups into *cocompact* and *cofinite* groups depending on whether their fundamental domain is compact or not, respectively. The fundamental domain of a cofinite group Γ must have at least one vertex in $\widehat{\mathbb{R}}$, corresponding to a parabolic motion in Γ. Such a vertex is called a *cusp* of Γ. From now on we always assume that Γ is a Fuchsian group of the first kind. Moreover, for simplicity we limit to the case when Γ has at most one cusp (or indeed none), which can be taken to be ∞ after conjugation in $\mathrm{SL}_2(\mathbb{R})$. For any $z \in \widehat{\mathbb{H}}^2$, we define its stabiliser in Γ by

$$\Gamma_z = \{\gamma \in \Gamma : \gamma z = z\}.$$

This is always a cyclic group. Here the action of $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ extends to $\widehat{\mathbb{R}}$ by letting $\gamma(-d/c) = \infty$ and $\gamma\infty = a/c$ if $c \neq 0$, and $\gamma\infty = \infty$ otherwise. Some standard examples of cofinite Fuchsian groups are the modular group $\mathrm{SL}_2(\mathbb{Z})$ and the congruence groups. We let

$$\Gamma(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \bmod N \right\}$$

be the *principal congruence group of level* N. Any subgroup of $\mathrm{SL}_2(\mathbb{Z})$ which contains $\Gamma(N)$ is called a *congruence group*. One example we have already seen is

$$\Gamma_0(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} * & * \\ & * \end{pmatrix} \bmod N \right\}, \tag{4.3}$$

which is known as the *Hecke congruence group of level* N. We also let

$$\Gamma_1(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \bmod N \right\}.$$

The quotient space $M = \Gamma \backslash \mathbb{H}^2$ is a Riemann surface with the metric induced by the natural projection from the universal covering space $\mathbb{H}^2$ with deck transformations corresponding to the group Γ. If M is compact then it has genus at least two.

Elliptic transformations correspond to marked points (branch points) on M and different parabolic subgroups give the inequivalent cusps of M. Moreover, recall that if g is a hyperbolic transformation, then the geodesic joining its fixed points is mapped to itself by g. Hence, closed geodesics of M correspond to hyperbolic conjugacy classes in Γ and in particular the prime geodesics correspond to conjugacy classes generated by primitive hyperbolic elements. Now suppose that $\mathfrak{p}$ is a primitive hyperbolic conjugacy class. Then the *norm* of $\mathfrak{p}$ is defined in terms of the trace as

$$\operatorname{Tr} \mathfrak{p} = N\mathfrak{p}^{1/2} + N\mathfrak{p}^{-1/2}. \tag{4.4}$$

Since trace is invariant under conjugation it follows that the norm does not depend on the choice of the representative in $\mathfrak{p}$. The length of the closed geodesic corresponding to $\mathfrak{p}$ is then given by $\log N\mathfrak{p}$ (counted with multiplicity). Solving the equation (4.4) gives the following expression for the length

$$2\log\left(\frac{\operatorname{Tr}\mathfrak{p} + \sqrt{(\operatorname{Tr}\mathfrak{p})^2 - 4}}{2} \right). \tag{4.5}$$

4.2.2 Automorphic Forms and $L^2(\Gamma\backslash\mathbb{H}^2)$

Let Γ be a Fuchsian group of the first kind and set $M=\Gamma\backslash\mathbb{H}^2$. Our goal is to decompose $f\in L^2(M)$ in terms of the eigenfunctions of Δ. Here we equip the Hilbert space $L^2(M)$ with the standard inner product

$$\langle f,g\rangle=\int_M f(z)\overline{g(z)}\,d\mu(z).$$

We say that a function $f:\mathbb{H}^2\longrightarrow\mathbb{C}$ is *automorphic with respect to* Γ (or Γ-automorphic or Γ-invariant for short) if

$$f(\gamma z)=f(z),\qquad \text{for all } \gamma\in\Gamma,$$

so that f induces a function $f:M\longrightarrow\mathbb{C}$. If, furthermore, f is an eigenfunction of the Laplacian

$$\Delta f+\lambda f=0,\qquad \lambda=s(1-s),$$

then we say that f is an *automorphic form* (or a *Maaß form*). The existence of a cusp for cofinite Γ allows us to do Fourier expansion in a simple manner. Let f be an automorphic form that does not grow too quickly in the cusp, i.e. $f(z)=o(e^{2\pi y})$ as $y\to\infty$. Then, since $f(z+1)=f(z)$, we have the Fourier expansion [53, Theorem 3.1]

$$f(z)=a_0(y)+\sum_{n\neq 0}\widehat{a}_n W_s(nz),$$

where

$$a_0(y)=\frac{A}{2}(y^s+y^{1-s})+\frac{B}{2s-1}(y^s-y^{1-s}),\qquad s\neq 1/2,$$

for some A and B. Moreover, for such automorphic forms we know that the non-zero coefficients decay exponentially

$$f(z)=a_0(y)+O(e^{-2\pi y}),\qquad \text{as } y\to\infty. \tag{4.6}$$

Important examples of automorphic forms are the (non-holomorphic) Eisenstein series, which are defined in terms of the cusps of Γ. In our case there is only the cusp at infinity so we define the Eisenstein series at ∞ for $\operatorname{Re} s>1$ by

$$E(z,s)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma}(\operatorname{Im}\gamma z)^s,$$

where Γ_∞ is generated by $\left(\begin{smallmatrix}1&1\\&1\end{smallmatrix}\right)$. The Eisenstein series $E(z,s)$ is an eigenfunction of Δ

with the eigenvalue $\lambda = s(1-s)$, but it is not square integrable. The Fourier expansion of $E(z,s)$ always has a non-vanishing zero-th coefficient. For $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ this series takes the form (e.g. [80])

$$E(z,s) = y^s + \varphi(s)y^{1-s} + \frac{2y^{1/2}}{\xi(2s)} \sum_{n\neq 0} |n|^{s-1/2} \sigma_{1-2s}(|n|) K_{s-1/2}(2\pi|n|y) e^{2\pi i n x},$$

where ξ is the completed Riemann zeta function, $\sigma_s(n)$ is the sum of divisors function, and $\varphi(s)$ is the *scattering matrix* with the explicit form (for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$)

$$\varphi(s) = \frac{\xi(2-2s)}{\xi(2s)}. \tag{4.7}$$

The Eisenstein series has an analytic continuation as a meromorphic function in s to all of $\mathbb{C}$. The scattering matrix $\varphi(s)$ (for any cofinite Γ) is holomorphic in $\operatorname{Re} s \geq 1/2$, except for a finite number of simple poles $s_j \in (1/2, 1]$. The poles of $E(z,s)$ in $\operatorname{Re} s \geq 1/2$ are among the poles of $\varphi(s)$ and they are simple. The residues are Maaß forms orthogonal to the cusp forms. In particular, $E(z,s)$ has no poles on the critical line and is non-zero if $s \neq 1/2$. The point $s = 1$ (corresponding to the zero eigenvalue) is always a simple pole of $E(z,s)$ with residue [53, Theorem 6.13]

$$\operatorname*{res}_{s=1} E(z,s) = \frac{1}{\mathrm{vol}(M)}.$$

Furthermore, the Eisenstein series satisfies the functional equation [53, Theorem 6.5]

$$E(z,s) = \varphi(s) E(z, 1-s)$$

for any $s \in \mathbb{C}$.

Let ψ be some smooth, compactly supported function on $\mathbb{R}^+$. We define the *incomplete Eisenstein series* (with respect to ψ) by

$$E(z|\psi) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \psi(\operatorname{Im} \gamma z).$$

As opposed to $E(z,s)$, the incomplete Eisenstein series are not eigenfunctions of Δ, but they are of course in $L^2(M)$ as $E(z|\psi)$ is bounded and M has finite volume. The crucial point is that Eisenstein series can be used to represent the incomplete Eisenstein series through Mellin inversion and this can be used to construct the eigenpacket of Δ

in $L^2(M)$. Concretely, we have

$$E(z|\psi) = \frac{1}{2\pi i}\int_{(\sigma)} E(z,s)\mathcal{M}\psi(-s)\,ds,$$

where $\sigma > 1$ and $\mathcal{M}\psi$ denotes the Mellin transform

$$\mathcal{M}\psi(s) = \int_0^\infty \psi(y)y^{s-1}\,dy.$$

Ultimately, $L^2(M)$ can be decomposed as $L^2(M) = \widetilde{\mathcal{E}}(M) \oplus \widetilde{\mathcal{C}}(M)$, where $\mathcal{E}(M)$ denotes the subspace spanned by the incomplete Eisenstein series, $\mathcal{C}(M)$ is its orthogonal complement (inside the space of smooth and bounded automorphic functions), and $\tilde{}$ denotes closure in $L^2(M)$. Automorphic forms contained in the space $\mathcal{C}(M)$ are called *cusp forms*. An equivalent characterisation for a cusp form is to say that it is an automorphic form whose zero-th Fourier coefficient vanishes. In light of (4.6) this means that all cusp forms vanish at the cusp.

An important object to the study of the spectrum of Δ is the *automorphic kernel*. These arise naturally from considering invariant integral operators on M. It is possible to show that an integral operator with kernel k is invariant under Γ (and thus commutes with Δ) if and only if k is a function of a point-pair invariant. Let $k : \mathbb{R}^+ \longrightarrow \mathbb{C}$ be a smooth, compactly supported function. This condition can be relaxed to just $k(u), k'(u) \ll (1+u)^{-2}$, see [53, pg. 68]. The automorphic kernel is then obtained by averaging k as a function of the point-pair invariant over the whole group

$$K(z,w) = \sum_{\gamma\in\Gamma} k(u(z,\gamma w)),$$

where the series converges absolutely. Then K is an automorphic form as a function of z (or, equivalently, w). We often write just simply $k(z,\gamma w)$ and the dependence on $u(\cdot,\cdot)$ is implicit. Observe that as a function on $M \times M$, K is certainly not bounded along the diagonal $z = w$ on M if Γ has a cusp and $y \to \infty$. The power of automorphic kernels comes from their spectral decomposition, the pre-trace formula (4.8).

We can finally describe the spectral resolution of the Laplacian on M. For cocompact Γ there is only a discrete spectrum. So suppose Γ is cofinite. In the space $\mathcal{C}(M)$, Δ has a pure point spectrum, which means that $\mathcal{C}(M)$ has an orthonormal basis consisting of cusp forms. In contrast, the spectrum of Δ in $\mathcal{E}(M)$ is continuous, spanned by $E(z, \frac{1}{2} + it)$ for all t, except for a finite dimensional subspace of the point spectrum, called the *residual spectrum*, coming from the residues of the Eisenstein series in

$1/2 < s \leq 1$. We let $\{u_j\}_{j\geq 0}$ be an orthonormal basis for the discrete spectrum. As a result, we have that any $f \in L^2(M)$ can be decomposed as

$$f(z) = \sum_j \langle f, u_j \rangle u_j(z) + \frac{1}{4\pi} \int_{-\infty}^{\infty} \langle f, E(\cdot, \tfrac{1}{2}+it) \rangle E(z, \tfrac{1}{2}+it)\, dt,$$

which converges in the norm topology. By [53, Theorem 7.3] this expansion converges pointwise absolutely if f and Δf are smooth and bounded. For automorphic kernels this decomposition has a particularly nice form. The coefficients are expressible in terms of an explicit integral transformation, the *Selberg–Harish-Chandra transform*. For a kernel $k(u)$ its Selberg–Harish-Chandra transform $h(t)$ is given by the following formula [53, (1.62)].

$$\begin{aligned} q(v) &= \int_v^{\infty} k(u)(u-v)^{-1/2}\, du, \\ g(r) &= 2q(\sinh^2 \tfrac{r}{2}), \\ h(t) &= \int_{-\infty}^{\infty} e^{irt} g(r)\, dr. \end{aligned}$$

On the other hand, if $h(t)$ is even and is holomorphic in the strip $|\mathrm{Im}\, t| \leq 1/2+\epsilon$ and satisifes $h(t) \ll (|t|+1)^{-2-\epsilon})$, then $h(t)$ has an inverse Selberg–Harish-Chandra transform given by [53, (1.63)]

$$\begin{aligned} g(r) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{irt} h(t)\, dt, \\ q(v) &= \frac{1}{2} g(2 \log \sqrt{v+1} + \sqrt{v}), \\ k(u) &= -\frac{1}{\pi} \int_u^{\infty} (v-u)^{-1/2}\, dq(v). \end{aligned}$$

For the pair k and h satisfying the above conditions, the spectral decomposition of K is then given by [53, Theorem 7.4]

$$K(z,w) = \sum_j h(t_j) u_j(z) \overline{u}_j(w) + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(r) E(z, \tfrac{1}{2}+it) \overline{E}(w, \tfrac{1}{2}+it)\, dr. \tag{4.8}$$

The Selberg–Harish-Chandra transform is a special case of the more general Plancherel theorem for spherical eigenfunctions on a symmetric space [39, 40].

Let us summarise our knowledge of the spectrum of Δ. We write the eigenvalues as $\lambda_j = s_j(1-s_j) = \frac{1}{4} + t_j^2$. There are both the discrete and continuous spectrum if Γ is not cocompact. The continuous spectrum spans $\lambda \in [1/4, \infty)$ (so that $t_j \in \mathbb{R}$, $s_j = \frac{1}{2} + it_j$)

with multiplicity one and the eigenpacket is given by the Eisenstein series at the (only) cusp of Γ at ∞, $E(z,\frac{1}{2}+it)$. The discrete spectrum is split into the residual and cuspidal eigenvalues. There are finitely many residual eigenvalues $s_j \in (\frac{1}{2},1)$, i.e. $\lambda_j \in [0,1/4)$, and they correspond to residues of the Eisenstein series, which are Maaß forms. In particular, the residue of $E(z,s)$ at $s=1$ shows that $\lambda_0 = 0$ is an eigenvalue with the corresponding constant eigenfunction $u_0 = \mathrm{vol}(M)^{-1/2}$. There can also be finitely many cuspidal eigenvalues in $1/2 < s_j < 1$, together with the non-zero residual eigenvalues these are called the *small eigenvalues* of Δ (or sometimes the *exceptional eigenvalues*). The rest of the discrete spectrum embeds into the continuous spectrum $\lambda_j \in [1/4,\infty)$. We call the quantity $\lambda_1 - \lambda_0$ the *spectral gap*.

There are many fundamental aspects of the spectrum that are still not known. For general non-congruence cofinite Γ it is not even known whether the cuspidal spectrum is finite or not. The belief is that typically the discrete spectrum is very small and that the Fourier coefficients of cusp forms should carry arithmetic properties. See Phillips and Sarnak [84] for a discussion where the authors give various characterisations of this problem. For some groups the exceptional spectrum (both the cuspidal and residual) certainly exists. For congruence groups we know that apart from the zero eigenvalue, the residual spectrum does not appear [53, Theorem 11.3]. It is actually believed that these groups do not have any small eigenvalues. This is known as the Selberg eigenvalue conjecture. The current best result is due to Booker and Strömbergsson [7] who prove, with the Selberg Trace Formula, that the conjecture holds for $\Gamma_1(N)$, where N is square-free and $N < 857$.

4.2.3 Selberg Trace Formula

As a final topic in our overview of the spectral theory in $\mathbb{H}^2$ we look at the Selberg Trace Formula (STF), which encompasses information about the spectrum and geometry of M. It is for this reason that the STF has many applications of which we will present a few. Our theorem in Chapter 5 is also a direct application of the STF.

A trace formula is the result of calculating the trace of an operator in two different ways. For example, the Poisson summation formula is a trace formula for $\mathbb{R}/\mathbb{Z}$. Consider an automorphic kernel K given by k, and the associated invariant integral operator. We say that K is of *trace class* if K is absolutely integrable along the diagonal

$z = w$. In this case we define the *trace* of K by

$$\operatorname{Tr} K = \int_M K(z,z)\,d\mu(z).$$

The idea is to evaluate the trace by integrating the definition of K or by using the spectral expansion (4.8). The resulting formula is called the *pre-trace* formula and for compact M it takes the form

$$\sum_{\gamma\in\Gamma} \int_M k(z,\gamma z)\,d\mu(z) = \sum_j h(t_j).$$

For non-compact M the situation is more difficult since a typical automorphic kernel is not of trace class. It is possible to deal with this by subtracting the contribution to the trace coming from the cuspidal zone [53, §10.1]. Real insight is gained when the geometric side (with the series over Γ) is decomposed depending on whether the conjugacy class of γ is parabolic, hyperbolic or elliptic. The result is the following theorem (simplified for one cusp).

Theorem 4.1 (Selberg Trace Formula, [53, Theorem 10.2]). *Suppose $h(t)$ has an inverse Selberg–Harish-Chandra transform $k(u)$ and let g be the Fourier transform of h. Then,*

$$\begin{aligned}
\sum_j h(t_j) + \frac{1}{4\pi}\int_{-\infty}^{\infty} h(t)\frac{-\varphi'}{\varphi}(\tfrac{1}{2}+it)\,dt \\
&= \frac{\operatorname{vol}(\Gamma\backslash\mathbb{H}^2)}{4\pi}\int_{-\infty}^{\infty} h(r)r\tanh\pi r\,dr \\
&\quad + \sum_{\mathfrak{p}}\sum_{l=1}^{\infty}(N\mathfrak{p}^{l/2}-N\mathfrak{p}^{-l/2})^{-1}g(l\log N\mathfrak{p})\log N\mathfrak{p} \\
&\quad + \sum_{\mathscr{R}}\sum_{0<l<m_{\mathscr{R}}}\left(2m_{\mathscr{R}}\sin\frac{\pi l}{m_{\mathscr{R}}}\right)^{-1}\int_{-\infty}^{\infty} h(r)\frac{\cosh\pi(1-2l/m_{\mathscr{R}})r}{\cosh\pi r}\,dr \\
&\quad + \frac{h(0)}{4}(1-\varphi(\tfrac{1}{2})) - g(0)\log 2 \\
&\quad - \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r)\psi(1+ir)\,dr,
\end{aligned}$$

where all the series and integrals converge absolutely.

Let us explain all the quantities appearing in the trace formula. Here $\mathfrak{p}$ ranges over primitive hyperbolic conjugacy classes with norm $N\mathfrak{p}$ given by (4.4), and $\mathscr{R}$ ranges over primitive elliptic classes of order $m_{\mathscr{R}} > 1$. Also, ψ is the usual digamma function and φ is the scattering matrix. By carefully choosing the test function $h(t)$, it is possible to deduce many results from the STF. Apart from the Selberg eigenvalue conjecture that was mentioned before, examples of applications of STF to the Prime Geodesic Theorem, Weyl's law, and density theorems for the eigenvalues, to name a few, are presented in [53].

4.3 Dedekind Zeta Function

Before moving on to automorphic forms in three dimensions, we will give a brief primer on the algebraic number theory of number fields and the associated zeta functions. We mostly refer to [17, 18] in this section.

Let $K=\mathbb{Q}(\sqrt{D})$ be an imaginary quadratic field, so $D<0$. Without loss we may assume that D is square-free. Denote by $\mathcal{O}_K$ the ring of integers of K, which is a free $\mathbb{Z}$-module of rank two. Often we write $\mathcal{O}$ when K is fixed. A $\mathbb{Z}$-basis for $\mathcal{O}$ is given by $\langle 1,\omega\rangle$, where [17, pg. 137]

$$\omega=\frac{d_K+\sqrt{d_K}}{2},$$

and d_K is the *discriminant* given by

$$d_K=\begin{cases}4D, & \text{if } D\equiv 2 \text{ or } 3 \bmod 4,\\ D, & \text{if } D\equiv 1 \bmod 4.\end{cases}$$

Since K is a quadratic field extension, d_K is also a *fundamental discriminant* so that $d_K\equiv 0$ or $1 \bmod 4$. The group of units in $\mathcal{O}$ is denoted by $\mathcal{O}^\times$. The class number of K is denoted by h_K or just h. There is only one pair of complex embeddings of K into $\mathbb{C}$ given by the complex conjugation. We will be working with imaginary quadratic fields of class number one. In this case $\mathcal{O}$ is a principal ideal domain.

The Dedekind zeta function of K is defined for $\operatorname{Re} s>1$ by

$$\zeta_K(s)=\sum_{\mathfrak{a}\subset\mathcal{O}}{}' \frac{1}{N\mathfrak{a}^s}=\prod_{\mathfrak{p}}\frac{1}{1-N\mathfrak{p}^{-s}},$$

where the prime in the summation denotes that it is taken over non-zero ideals $\mathfrak{a}$, and the Euler product is taken over prime ideals $\mathfrak{p}\subset\mathcal{O}$. As with the Riemann zeta function, it follows that ζ_K is non-zero for $\operatorname{Re} s>1$. Define the completed zeta function ξ_K of K by

$$\xi_K(s)=\left(\frac{\sqrt{|d_K|}}{2\pi}\right)^s\Gamma(s)\zeta_K(s). \tag{4.9}$$

Notice that this is not the standard way to define the completed zeta function as usually the factor $|d_K|^{s/2}$ is left out and appears in the functional equation instead. However, our convention will make the life a little bit easier notationally in Chapter 7. Then ξ_K satisfies the functional equation

$$\xi_K(s) = \xi_K(1-s). \tag{4.10}$$

For a statement, see for example [18, Theorem 10.5.1 (2)] which transforms to our form in (4.10) after an application of the duplication formula for $\Gamma(s/2)$, (A.5). Moreover, by the Dirichlet Class Number Formula [18, Theorem 10.5.1 (4)], ζ_K has a simple pole at $s=1$ with residue

$$\operatorname*{res}_{s=1} \zeta_K(s) = \frac{2\pi h R}{|\mathscr{O}^\times|}, \tag{4.11}$$

where h is the class number of K, R is the regulator of K and $|\mathscr{O}^\times|$ is the number of units in $\mathscr{O}$. Notice that for imaginary quadratic fields $R=1$.

It is possible to deduce many properties of ζ_K from those of the Riemann zeta function and $L(s,\chi)$. This is because for quadratic number fields $K=\mathbb{Q}(\sqrt{D})$ with discriminant d_K we can factorise the Dedekind zeta function as (e.g. [18, Proposition 10.5.5])

$$\zeta_K(s) = \zeta(s) L(s, \chi_{d_K}), \tag{4.12}$$

where $\chi_{d_K}(n) = \left(\frac{d_K}{n}\right)$ is the Legendre–Kronecker symbol. In fact, χ_{d_K} is a primitive quadratic Dirichlet character (see [17, Theorem 2.2.15]).

We now proceed to use the factorisation (4.12) to obtain various estimates on ζ_K. It is a simple consequence of the functional equation of ζ and Phragmén–Lindelöf that ζ (and $L(s,\chi)$ by analogue) satisfies the following convexity bound [76, §2.12]:

$$\zeta(\sigma+it) \ll_{a,b,\epsilon} |t|^{\mu_0(\sigma)+\epsilon},$$

where $\sigma \in (a,b)$, $\epsilon > 0$ is fixed and μ_0 is as before in (2.19). Hence,

$$\zeta_K(\sigma+it) \ll_{a,b,\epsilon} |t|^{2\mu_0(\sigma)+\epsilon}. \tag{4.13}$$

For the analysis of QUE in the case $\sigma_\infty = 1$, the above estimate is not sufficient. Instead, we need subconvex bounds. For the Riemann zeta function we know due to Weyl that (see [76, §6.6]):

$$\zeta(\tfrac{1}{2}+it) \ll t^{1/6} \log t, \tag{4.14}$$

as $t \to \infty$. The same is of course true for Dirichlet L-functions. The proof of (4.14) uses the approximate functional equation and methods of exponential sums due to van der Corput. The bound (4.14) is not the best known estimate, but it is sufficient for our purposes. The factorisation (4.12) implies that

$$\zeta_K(\tfrac{1}{2}+it) \ll_K t^{1/3+\epsilon}, \tag{4.15}$$

as $t \to \infty$. In fact, (4.15) is known for any algebraic number field of degree n with the exponent replaced by $n/6+\epsilon$, [42]. As with the Riemann zeta function, (4.15) is far away from the conjectured bound

$$\zeta_K(\tfrac{1}{2}+it) \ll_K t^{\epsilon},$$

for any $\epsilon > 0$. This is also known as the Lindelöf Hypothesis for K. An even stronger bound is expected according to Heath-Brown [42, pg. 324]. We will also need a non-trivial estimate for $1/\zeta_K(s)$ in the region of absolute convergence. We use again the factorisation (4.12) and the corresponding result for ζ, [100, 3.5.1 and Theorem 3.11] (and analogously for $L(s,\chi)$ together with [61]) to get

$$\log^{-1}|t| \ll \zeta(\sigma+it) \ll \log|t|,$$

where $1 \leq \sigma \leq 2$. Thus, we deduce that

$$\log^{-2}|t| \ll_K \zeta_K(\sigma+it) \ll_K \log^2|t|. \tag{4.16}$$

Finally, we will also need a strong estimate on the logarithmic derivative of ζ_K. As before, this follows from the Weyl bound for $\sigma \geq 1$, see [100, Theorems 3.11 and 5.17]. Therefore

$$\frac{\zeta'}{\zeta}(\sigma+it) \ll \frac{\log t}{\log\log t},$$

which can also be obtained for $L(s,\chi)$, see [19]. Hence, we conclude that

$$\frac{\zeta_K'}{\zeta_K}(\sigma+it) \ll_K \frac{\log t}{\log\log t}. \tag{4.17}$$

4.4 Hyperbolic 3-space

The geometry and analysis in three dimensions is somewhat similar to $\mathbb{H}^2$. Hence, we limit ourselves to a shorter discussion mainly highlighting the differences (and some of the more interesting similarities). This section is almost entirely based on the book *Groups acting on hyperbolic space* by Elstrodt, Grunewald, and Mennicke [25]. However, we adopt a mixture of notation from the aforementioned book as well as that used by Sarnak [90], which will hopefully be more clear and appropriate in the context of this thesis. Also, please notice that the s-plane for the eigenvalues differs in Elstrodt ($\lambda = 1 - s^2$) compared to what is typically used in number theory ($\lambda = s(2-s)$). We use the latter convention, which agrees with [90].

4.4.1 Hyperbolic Geometry in Three Dimensions

Let $\mathbb{H}^3$ be the three dimensional upper half-space given by

$$\mathbb{H}^3 = \{(x_1, x_2, y) : x_1, x_2 \in \mathbb{R}, y \in \mathbb{R}^+\}.$$

Denote points $p \in \mathbb{H}^3$ by $p = z + yj = (z, y) = (x_1, x_2, y)$, where $z = x_1 + ix_2 \in \mathbb{C}$ and $y > 0$. It is possible to think of $\mathbb{H}^3$ as a subset of the quaternions with the last coordinate set to zero. The Euclidean norm of a point $p \in \mathbb{H}^3$ is given by

$$||p||^2 = |z|^2 + y^2 = x_1^2 + x_2^2 + y^2.$$

The hyperbolic metric

$$ds^2 = \frac{dx_1^2 + dx_2^2 + dy^2}{y^2} \tag{4.18}$$

gives rise to the volume element

$$d\mu(p) = \frac{dx_1\, dx_2\, dy}{y^3}. \tag{4.19}$$

Equipped with (4.18), $\mathbb{H}^3$ is a model for the three dimensional hyperbolic space. The geodesics on $\mathbb{H}^3$ are given by lines and Euclidean semicircles perpendicular to the Euclidean boundary of $\mathbb{H}^3$, which can be identified with $\mathbb{P}^1\mathbb{C}$. The group $\mathrm{SL}_2(\mathbb{C})$ acts transitively on $\mathbb{H}^3$ with the action of $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C})$ given by

$$\gamma p = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot (z + yj) = \left(\frac{(az+b)\overline{(cz+d)} + a\bar{c}y^2}{||cp+d||^2}, \frac{y}{||cp+d||^2} \right), \tag{4.20}$$

where

$$||cp+d||^2 = |cz+d|^2 + |c|^2 y^2.$$

This action is an orientation preserving isometry. We can also state the action in terms of the inverse in the skew field of quaternions as

$$\gamma p = (ap + b)(cp + d)^{-1},$$

which is in a form similar to the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}^2$. The action of $\gamma \in \Gamma$ on the boundary $\mathbb{P}^1\mathbb{C}$ is given by

$$\gamma(x : y) = (ax + by : cx + dy).$$

The group of all orientation preserving isometries on $\mathbb{H}^3$ is isomorphic to $\mathrm{PSL}_2(\mathbb{C})$. Together with complex conjugation, $z + yj \mapsto \overline{z} + yj$, these generate all the isometries of $\mathbb{H}^3$ under the hyperbolic metric. The classification of isometries is as before, except we also have the case when $\mathrm{Tr}\,\gamma \notin \mathbb{R}$, in which case γ is called *loxodromic*. We say that a subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ is *discrete* if the inverse image of Γ in $\mathrm{SL}_2(\mathbb{C})$ is discrete under the standard topology in $\mathbb{C}^4$. These groups are also called *Kleinian groups*. In this case, Γ acts discontinously on $\mathbb{H}^3$ so that any orbit Γp, for any $p \in \mathbb{H}^3$, has no accumulation points in $\mathbb{H}^3$. We denote the fundamental domain of Γ acting on the space $\mathbb{H}^3$, by $\mathscr{F}_{\mathbb{H}^3}(\Gamma)$. The definition of cocompact and cofinite groups carries over from two dimensions. For two points $p = z + jy$ and $p' = z' + jy'$ in $\mathbb{H}^3$, the hyperbolic distance between p and p', $d(p, p')$, is given by

$$\cosh d(p, p') = \delta(p, p'),$$

where $\delta : \mathbb{H}^3 \times \mathbb{H}^3 \longrightarrow \mathbb{R}^+$ is the standard point-pair invariant,

$$\delta(p, p') = \frac{|z - z'|^2 + y^2 + y'^2}{2yy'}, \tag{4.21}$$

and satisfies $\delta(\gamma p, \gamma p') = \delta(p, p')$, for any $\gamma \in \mathrm{PSL}_2(\mathbb{C})$. The Laplacian on $\mathbb{H}^3$ is given by

$$\Delta = y^2\left(\frac{\partial^2}{\partial {x_1}^2} + \frac{\partial^2}{\partial {x_2}^2} + \frac{\partial^2}{\partial y^2}\right) - y\frac{\partial}{\partial y}.$$

We can also identify $\mathbb{H}^3$ as a homogeneous space through the stabiliser of the point $j = (0, 0, 1)$, which is $\mathrm{SU}(2)$. Hence we have that $\mathbb{H}^3 \cong \mathrm{SL}_2(\mathbb{C})/\mathrm{SU}(2)$, [25, §1.1.6].

4.4.2 Automorphic Forms and $L^2(\Gamma\backslash\mathbb{H}^3)$

Let Γ be either cofinite or cocompact. We again restrict to the case when Γ has at most one cusp, which we take to be ∞. Let $M=\Gamma\backslash\mathbb{H}^3$, which is a Riemannian manifold of constant negative curvature -1 with the metric induced from that of $\mathbb{H}^3$. The Γ-automorphic functions on M are given by $f:\mathbb{H}^3\longrightarrow\mathbb{C}$ such that $f(\gamma p)=f(p)$ for any $\gamma\in\Gamma$ and $p\in\mathbb{H}^3$. Further, if f is also an eigenfunction of Δ, that is $\Delta f+\lambda f=0$, and of polynomial growth in the possible cusp of M, then we say that f is an automorphic form with respect to Γ. Of particular interest to us are the square integrable automorphic forms $f\in L^2(\Gamma\backslash\mathbb{H}^3)$. As in two dimensions, these can be used to give a complete description of the spectrum of Δ.

4.4.3 Eisenstein Series

The Eisenstein series for Γ at the cusp at ∞ is given for $\operatorname{Re} s>2$ (notice the normalisation so that the critical line is at $\operatorname{Re} s=1$) by

$$E(p,s)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} y(\gamma p)^s,$$

where

$$\Gamma_\infty=\{\gamma\in\Gamma:\gamma\infty=\infty\}.$$

Notice that this differs from the definition in [25] by a constant multiple, $[\Gamma_\infty:\Gamma'_\infty]$. For $\Gamma=\mathrm{PSL}_2(\mathcal{O}_K)$ we know that this factor is equal to $|\mathcal{O}_K^\times|/2$, [25, pg. 379]. Here Γ'_∞ is the maximal unipotent subgroup of Γ_∞. The Eisenstein series $E(p,s)$ is an eigenfunction of Δ with the eigenvalue $\lambda=s(2-s)$, but it is not square integrable. Let

$$\xi_K(s)=\left(\frac{\sqrt{|d_K|}}{2\pi}\right)^s\Gamma(s)\zeta_K(s),$$

be the completed Dedekind zeta function and define

$$\varphi(s)=\frac{\xi_K(s-1)}{\xi_K(s)}=\frac{2\pi}{\sqrt{|d_K|}}\frac{1}{s-1}\frac{\zeta_K(s-1)}{\zeta_K(s)},\tag{4.22}$$

which is the scattering matrix for $E(p,s)$. Then the Fourier development of $E(p,s)$ for $\Gamma=\mathrm{PSL}_2(\mathcal{O}_K)$ (where K is an imaginary quadratic field of class number one) is given by

$$E(p,s)=y^s+\varphi(s)y^{2-s}+\frac{2y}{\xi_K(s)}\sum_{0\neq n\in\mathcal{O}}|n|^{s-1}\sigma_{1-s}(n)K_{s-1}\left(\frac{4\pi|n|y}{\sqrt{|d_K|}}\right)e^{2\pi i\langle\frac{2\overline{n}}{\sqrt{d_K}},z\rangle},\quad(4.23)$$

where

$$\sigma_s(n)=\sum_{\substack{(d)\subset\mathcal{O}\\ d|n}}|d|^{2s}.$$

This form of the Fourier expansion can be found in [86, 58]. It also appears in a more general form in [1, (13), 25, §6 Theorem 2.11.]. The Eisenstein series $E(p,s)$ can be analytically continued to all of $\mathbb{C}$ as a meromorphic function of s. Any pole of the scattering matrix is also a pole for $E(p,s)$. The only possible poles of $\varphi(s)$ and $E(p,s)$ in the region $\operatorname{Re} s>1$ are the finitely many simple poles with $s\in(1,2]$. The pole at $s=2$ always occurs and it is simple with residue the constant function

$$\operatorname*{res}_{s=2}E(v,s)=\frac{|\mathcal{F}_\infty|}{\operatorname{vol}(M)},$$

where $\mathcal{F}_\infty$ is the fundamental domain of Γ_∞ acting on the boundary $\mathbb{C}$, [25, §6 Theorem 1.11]. All residues of $E(p,s)$ for $s\in(1,2]$ are square integrable Maaß forms. There are no poles for the Eisenstein series or the scattering matrix on the critical line. On this line $\varphi(1+it)$ is unitary. Moreover, since

$$\varphi(s)\varphi(2-s)=1,$$

the Eisenstein series has a functional equation [25, §6 Theorem 1.2]

$$E(p,s)=\varphi(s)E(p,2-s).\quad(4.24)$$

The incomplete Eisenstein series are defined for smooth $\psi(x)$ with compact support on $\mathbb{R}^+$ by

$$E(p|\psi)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\psi(y(\gamma p)).$$

As before, it is possible to decompose $L^2(M)$ into the orthogonal spaces spanned by the closures of incomplete Eisenstein series on one hand, and the Maaß cusp forms on the other hand (this follows from [25, §6 (1.26)]).

4.4.4 Spectral Resolution of Δ and the Pre-trace Formula

The spectrum of Δ is split in the same way as before. We write $\lambda_j = s_j(2-s_j) = 1 + t_j^2$ for the eigenvalues. For cocompact Γ there is only the discrete spectrum which has an orthonormal basis of automorphic eigenforms. In the cofinite case, Δ has both the discrete and continuous spectrum. There are possibly finitely many residual eigenvalues $1 < s_j < 2$ along with the constant eigenfunction $u_0 = |\mathrm{vol}(M)|^{-1/2}$. The rest of the discrete spectrum comprises of Maaß cusp forms with possibly finitely many eigenvalues in $1 < s_j < 2$ and the rest embedded in the continuous spectrum, $\lambda_j \in [1, \infty)$. The eigenpacket of the continuous spectrum is given by the Eisenstein series on the critical line, $E(p, 1+it)$.

As was the case in two dimensions, it is not known whether the cuspidal spectrum is infinite for a generic cofinite Γ. The situation is in some sense more complicated, or rather less accessible to the methods available in two dimensions, due to the groups having no deformations [25, pp. 308–309] (which is true in fact for any dimension greater than two). The Fourier expansion of a cusp form $u_j(p)$ with eigenvalue λ_j is given by

$$u_j(p) = y \sum_{0 \neq n \in \mathcal{O}_K^*} \rho_j(n) K_{it_j}(2\pi|n|y) e^{2\pi i \langle n, z \rangle}, \tag{4.25}$$

where $\mathcal{O}_K^*$ is the dual lattice,

$$\mathcal{O}_K^* = \{m : \langle m, n \rangle \in \mathbb{Z} \text{ for all } n \in \mathcal{O}_K\}.$$

In order to describe the pre-trace formula we need the Selberg–Harish-Chandra transform (or just the Selberg transform for short) for three dimensions. Notice that in three dimensions zero distance corresponds to the point-pair invariant being equal to one. Let $k : [1, \infty) \longrightarrow \mathbb{C}$ be a smooth function with rapid decay in all derivatives. We consider k implicitly as a function of the point-pair invariant δ. The Selberg transform $h(\lambda)$ of k is then given by [25, §3 Theorem 5.3]

$$h(\lambda) = \frac{4\pi}{s-1} \int_0^\infty k(\cosh u) \sinh((s-1)u) \sinh u \, du, \tag{4.26}$$

where for the case $s = 1$ one must take the limit as $s \to 1^+$. It is also possible to invert this transform, but we will not describe it here (see the proof of Lemma 6.20). Hence, the spectral expansion for the automorphic kernel

$$K(p, q) = \sum_{\gamma \in \Gamma} k(p, \gamma q),$$

is given by

$$K(p,q)=\sum_j h(\lambda_j)u_j(p)\overline{u}_j(q)+\frac{1}{4\pi}\int_{-\infty}^{\infty}h(1+t^2)E(p,1+it)\overline{E}(q,1+it)\,dt,$$

where the convergence is pointwise. The series $K(p,q)$ converges absolutely and uniformly on compact subsets of $\mathbb{H}^3\times\mathbb{H}^3$. Also, for a fixed $q\in\mathbb{H}^3$, the function $K(\cdot,q)$ is square integrable on M, [25, §6 Theorem 4.1].

4.4.5 Hecke Operators and Maaß Forms

We now introduce Hecke operators for $\Gamma=\mathrm{PSL}_2(\mathcal{O}_K)$ with K an imaginary quadratic field of class number one. These are defined by averaging over non-unimodular lattices, so we need an action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{H}^3$. For $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{GL}_2(\mathbb{C})$ and $p\in\mathbb{H}^3$, this is given by

$$\gamma p=(\sqrt{\det\gamma})^{-1}(ap+b)(cp+d)^{-1}\sqrt{\det\gamma},$$

where the inverse is taken as a subset of quaternions [86, pg. 112]. Now, for $n\in\mathcal{O}_K$, $n\neq 0$, we define $\mathcal{M}_n$ to be the set of two by two matrices over $\mathcal{O}_K$ of determinant n. The *Hecke operator* T_n is then defined by

$$T_nf(p)=\sum_{\tau\in\mathrm{SL}_2(\mathcal{O}_K)\backslash\mathcal{M}_n}f(\tau p),$$

for $f:M\longrightarrow\mathbb{C}$. The power of the Hecke operators comes from the fact that they are self-adjoint and they commute with themselves as well as with the Laplacian for any $f\in C^2(M)$. In particular, we can find an orthonormal system of simultaneous eigenfunctions of Δ and the Hecke operators. Now, let ϕ_j be a Hecke–Maaß cusp form with the Fourier development given by (4.25). Let the Hecke eigenvalues of ϕ_j be $T_n\phi_j=\lambda_j(n)\phi_j$. The fundamental theorem of Hecke theory is then the following.

Theorem 4.2 ([44, Satz 16.8, pg. 119]). *The Hecke eigenvalues satisfy*

$$\rho_j(n)=\rho_j(1)\lambda_j(n).$$

The Hecke eigenvalues are multiplicative and, in particular,

$$L(\phi_j,s)=\rho_j(1)\sum_{n\in\mathcal{O}_K}\frac{\lambda_j(n)}{N(n)^s}=\rho_j(1)\prod_{\mathfrak{p}}(1-\lambda_j(\mathfrak{p})N\mathfrak{p}^{-s}+N\mathfrak{p}^{1-2s})^{-1}.$$

Chapter 5

Landau-type Formulae for Exponential Sums

For the discussion in this chapter it is useful to define an analogue of the von Mangoldt function for the lengths of closed geodesics. Let

$$\Lambda_\Gamma(x) = \begin{cases} \log N\mathfrak{p}, & \text{if } x = N\mathfrak{p}^\ell,\ \ell \in \mathbb{N}, \\ 0, & \text{otherwise,} \end{cases} \tag{5.1}$$

where $\mathfrak{p}$ is a primitive hyperbolic conjugacy class of Γ. Let $\lambda_j = \frac{1}{4} + t_j^2$ be the eigenvalues of Δ in $\Gamma\backslash\mathbb{H}^2$, where $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. For $X > 1$, we define the following sum

$$S(T, X) = \sum_{|t_j| \leq T} X^{it_j},$$

which is symmetrised by including both t_j and $-t_j$. By Weyl's law we have trivially that

$$S(T, X) \ll T^2. \tag{5.2}$$

This exponential sum is directly related to the error term in the Prime Geodesic Theorem. Iwaniec [52] demonstrates an analogue of the classical explicit formula of ζ:

$$\sum_{N\mathfrak{a} \leq X} \Lambda_\Gamma(N\mathfrak{a}) = X + \sum_{|t_j| \leq T} \frac{X^{s_j}}{s_j} + O\left(\frac{X}{T} \log^2 X\right), \tag{5.3}$$

where $T \leq \sqrt{X}(\log X)^{-2}$ and the sum on the left-hand side is over hyperbolic conjugacy classes of Γ with norm up to X. Since there are no small eigenvalues, $s_j = \frac{1}{2} + it_j$, a summation by parts shows that the right-hand side of (5.3) becomes

$$X + O(X^{1/2}T^{-1}S(T,X) + XT^{-1}\log^2 X).$$

A balancing of error terms shows that $O(X^{3/4+\epsilon})$ is the best possible error term unless we can improve on (5.2). In [52] Iwaniec also proved that

$$S(T,X) \ll X^{11/48+\epsilon}T. \tag{5.4}$$

With the help of this estimate he then obtained the error term

$$\pi_\Gamma(x) = \mathrm{li}(x) + O(x^{35/48+\epsilon})$$

in the Prime Geodesic Theorem for $\mathrm{PSL}_2(\mathbb{Z})$. A different estimate for the exponential sum was obtained by Luo and Sarnak [66], who proved that

$$S(T,X) \ll X^{1/8}T^{5/4}\log^2 T.$$

Notice that while this is stronger in the X-aspect than the estimate (5.4), it is worse in terms of T. It seems difficult to get rid of this unfortunate feature and to obtain simultaneous improvements in both X and T.

The rest of this chapter arose as an investigation related to the work of Petridis and Risager [77]. In particular, they present the following conjecture that up to a factor of the order of X^ϵ, the exponential sum has square root cancellation in T.

Conjecture 5.1 ([77, Conjecture 2.2])**.** *For every $\epsilon > 0$ and $X > 2$, we have*

$$S(T,X) \ll_\epsilon T^{1+\epsilon}X^\epsilon.$$

This would also lead to the strongest possible error term in the Prime Geodesic Theorem, $\pi_\Gamma(x) - \mathrm{li}(x) = O(x^{1/2+\epsilon})$. Notice that the exponent is the same as the conjecture for the classical Prime Number Theorem. They arrive at this conjecture by studying the local averaging of the hyperbolic lattice point problem. In particular, their result for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ is

$$\int_{\Gamma\backslash\mathbb{H}^2} f(z)N(z,z,X)\,d\mu(z) = \frac{\pi X}{\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)}\int_{\Gamma\backslash\mathbb{H}^2} f(z)\,d\mu(z) + O_{f,\epsilon}(X^{7/12+\epsilon}),$$

where $N(z,w,X)$ counts the number of lattice points Γz within a circle of radius $\operatorname{arccosh}(X/2)$ centered at $w \in \mathbb{H}^2$, and f is a smooth, compactly supported function on $\Gamma\backslash\mathbb{H}^2$. Lattice point problems will be discussed in more detail in Chapter 6. The proof in [77] is based on analysing an integrated version of the pre-trace formula for a suitable test function f. The exponential sum $S(T,X)$ appears when estimating the contribution of the cuspidal spectrum. Petridis and Risager disentangled the contribution of the eigenfuctions in the pre-trace formula by using the average rate of Quantum Unique Ergodicity of Maaß forms. After a summation by parts they are naturally led to estimate $S(T,X)$.

We now report on the numerical investigation of the function $S(T,X)$ and prove a theorem about its behaviour for a fixed $X>1$, as $T\to\infty$. Let $\Lambda(X)$ be the von Mangoldt function extended to $\mathbb{R}$. The main result is the following theorem.

Theorem 5.2. *For a fixed $X>1$, we have*

$$S(T,X)=\frac{|\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2|}{\pi}\frac{\sin(T\log X)}{\log X}T+\frac{T}{\pi}(X^{1/2}-X^{-1/2})^{-1}\Lambda_\Gamma(X)$$
$$+\frac{2T}{\pi}X^{-1/2}\Lambda(X^{1/2})+O(T/\log T),$$

as $T\to\infty$.

We will prove this theorem by applying the Selberg Trace Formula (Theorem 4.1).

5.1 Numerical Discussion

We developed a Python program to investigate the behaviour of exponential sums. The programs, more plots and animations are available on the website [59]. Instead of $S(T,X)$ we study the slightly more general function

$$\widetilde{S}(T,X)=2\sum_{0<t_j\leq T}X^{it_j}.$$

We then have $\operatorname{Re}\widetilde{S}(T,X)=S(T,X)$ and the imaginary part $\operatorname{Im}\widetilde{S}(T,X)$ corresponds to the sine kernel. We used 53 000 eigenvalues from the data of Then [99] with 13 decimal digit precision. We have also used the data of Strömbergsson and Booker [98], which has a much higher precision of 53 decimal digits for 2 280 eigenvalues. We verify that the computations are robust, that is, the number of eigenvalues or their precision has no significant impact on our calculations. These verifications are also available on the website [59]. Our investigation resulted in the following observations:

Experimental Observation 1. *The growth of $S(T,X)$ is consistent with the conjecture.*

Experimental Observation 2. *For a fixed $X>1$, $S(T,X)$ has a peak of order T whenever X is equal to a power of a norm of a primitive hyperbolic class of Γ or an even power of a prime number $p \in \mathbb{N}$.*

Experimental Observation 2 is also in agreement with the results of Chazarain [15] that for the wave kernel the singularities occur at the lengths of closed geodesics (or in our case when $\log X$ is a multiple of a length of a prime geodesic). The peaks at even powers of rational primes are due to the poles of the scattering matrix φ. Experimental Observation 2 lead us to prove asymptotics for $S(T,X)$ for a fixed $X>1$, which is presented in the next section. We now summarise our progress towards these observations and also present some problems that we were not able to explain.

Taking into account the conjecture, we plot the normalised sum

$$\Sigma(T,X) = \widetilde{S}(T,X)T^{-1}.$$

In all of the plots below, blue corresponds to the real part and orange to the imaginary part of $\Sigma(T,X)$, respectively. In Figures 5.1 to 5.4 we have fixed $T=800$ with $X\to\infty$. After that we move on to plots where we have fixed X and vary the T. Recall that we expect a peak of order T at all even prime powers as well as powers of the norms of the primitive hyperbolic classes. The first few norms (up to 8 decimals) are given by formula (4.5) as

$$\begin{aligned} g_1 &= 6.85410196 & g_5 &= 46.97871376 \\ g_2 &= 13.92820323 & g_6 &= 61.98386677 \\ g_3 &= 22.95643924 & g_7 &= 78.98733975 \\ g_4 &= 33.97056274 & g_8 &= 97.98979486. \end{aligned}$$

We start by considering $\Sigma(T,X)$ in terms of X.

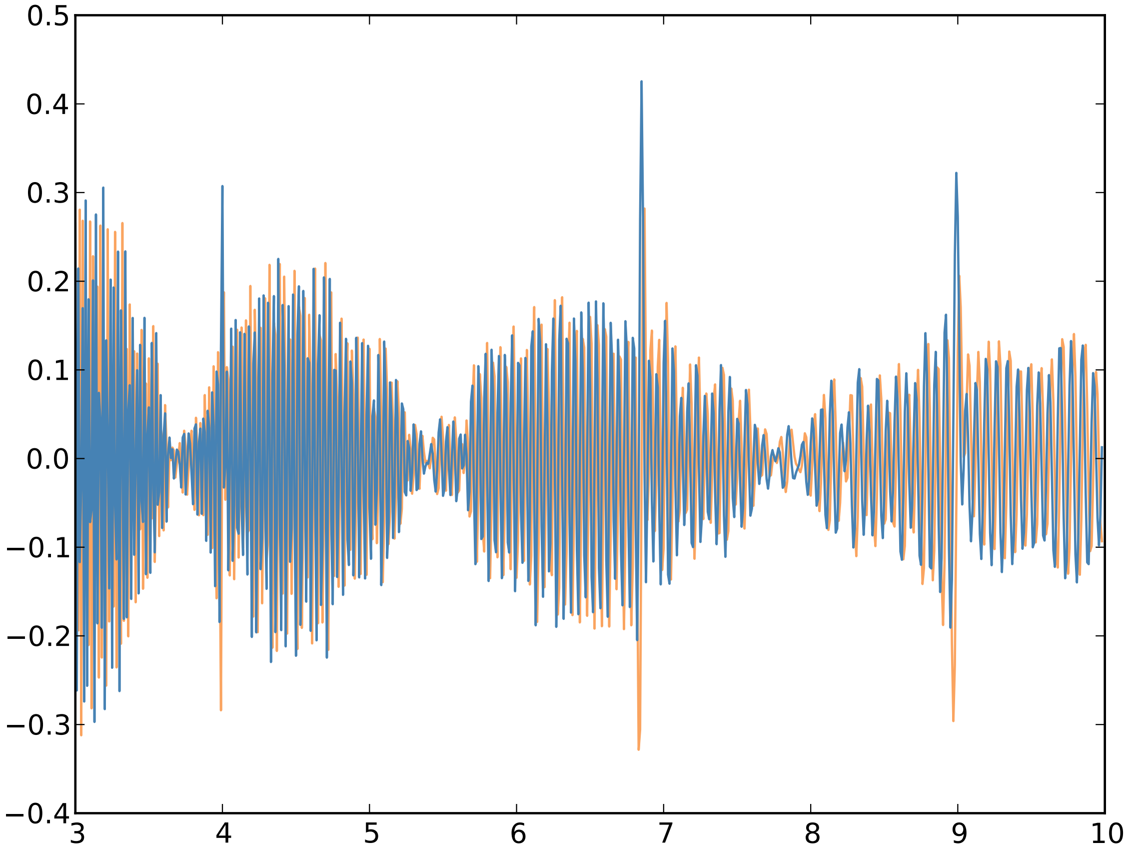


Figure 5.1: $\Sigma(T,X)$ in terms of X for $X \in [3,10]$.

Figure 5.1 clearly shows peak points at $X = 4 = 2^2$, $X = g_1$ and $X = 9 = 3^2$. Notice that the sine kernel has peaks at the exact same locations as $S(T,X)$, but they are of different nature. This will be more clear from the next figures. Also, both the real and imaginary part have noticeable intervals where they vanish (e.g. around $X = 3.6$, 5.4 and 7.7 in Figure 5.1). We are not certain how significant this is. In the following plot, Figure 5.2, we can see the peak points $X = g_2$ and $X = 16 = 2^4$, as well as strong vanishing around $X = 17.2$.

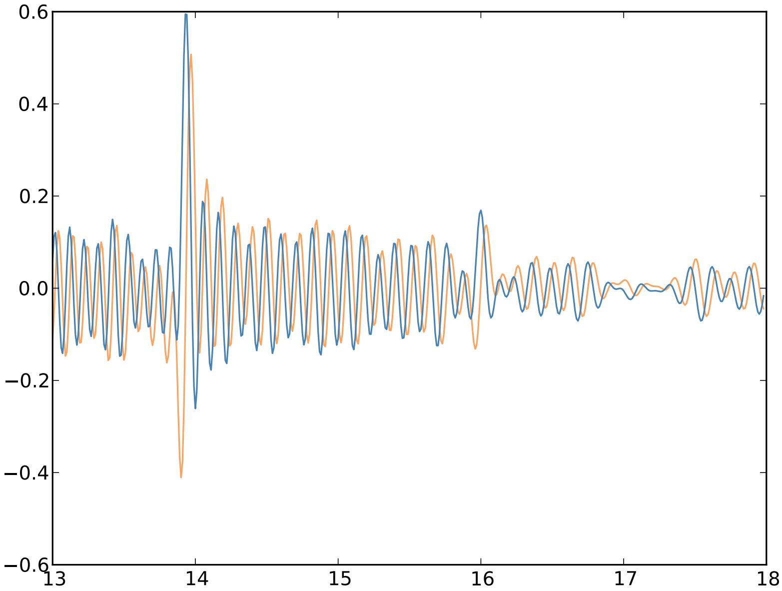


Figure 5.2: $\Sigma(T, X)$ in terms of X for $X \in [13, 18]$.

The next plot, Figure 5.3, highlights the differences in the oscillatory behaviour of the real and imaginary part.

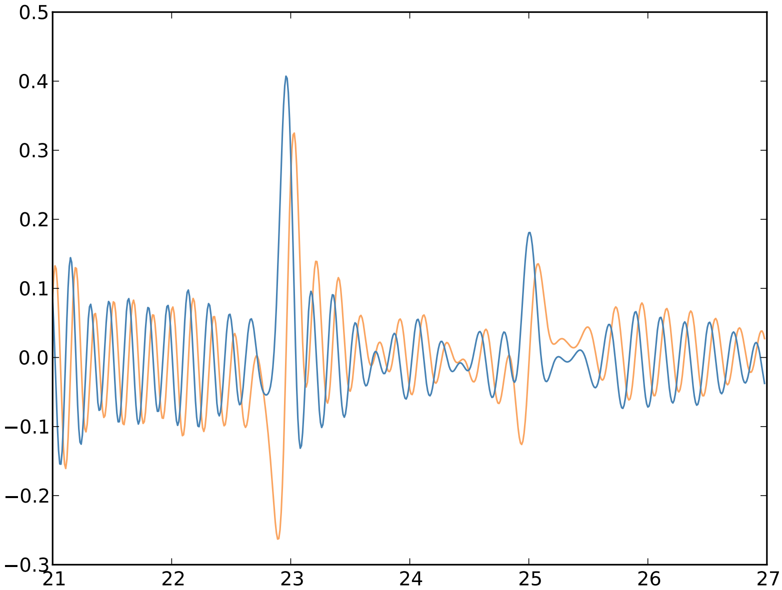


Figure 5.3: $\Sigma(T, X)$ in terms of X for $X \in [21, 27]$.

First of all, overall the oscillations for the cosine and sine kernels are very similar, but slightly out of sync as one might expect. However, at the peak points the sine kernel actually seems to vanish with negative and positive growth around that point. This is perhaps easier to see from the T-plots that follow. We have not been able to quantify this behaviour of the sine kernel. Figures 5.1–5.3 verify Theorem 5.2 numerically in accordance with Experimental Observation 2. In Figure 5.4 we look at $\Sigma(T,X)$ for X in a much larger interval. The graph agrees with Experimental Observation 1. On the other hand we cannot dispose of X^{ϵ} in the conjecture. The frequencies t_j are conjecturally linearly independent over $\mathbb{Q}$, which makes $S(T,X)$ the partial sums of an almost periodic function. Therefore, for a choice of arbitrarily large X, compared to T, $S(T,X)$ will be of size T^2.

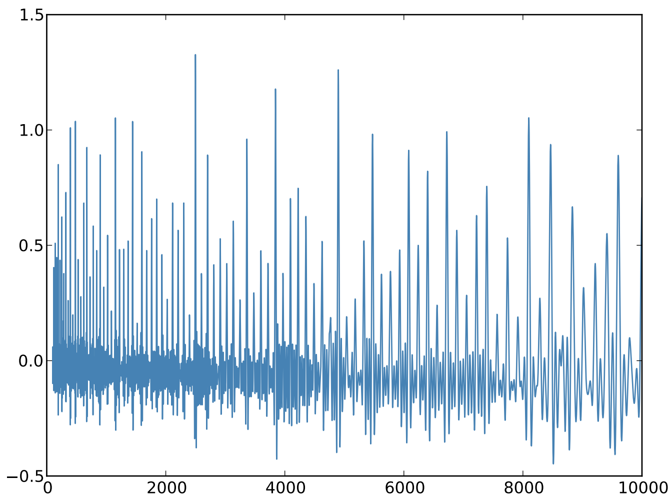
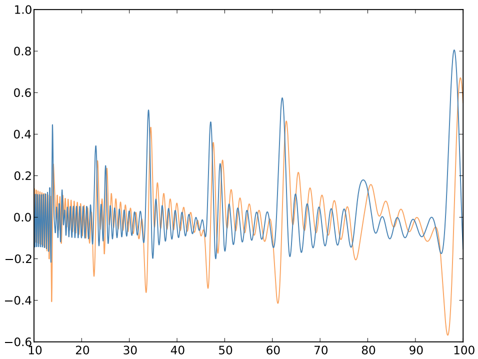
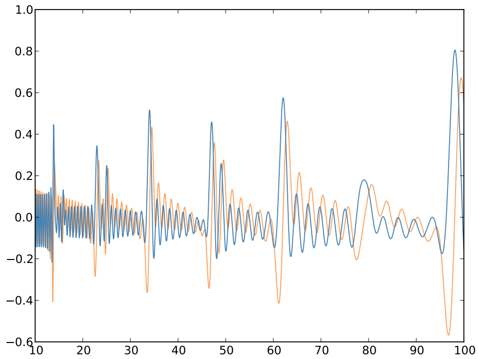


Figure 5.4: $\operatorname{Re}\Sigma(T,X)$ in terms of X for $X \in [100, 10\,000]$.

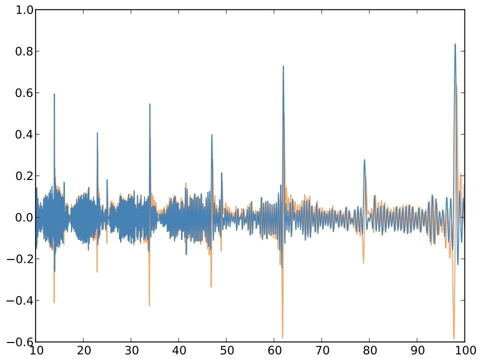
To verify the rigidity of our computations, we can look at plots in the same range of X for various values of T and for different sets of eigenvalue data.



(a) $T = 178$, data of Booker and Strömbergsson,



(b) $T = 178$, data of Then,



(c) $T = 800$, data of Then,

Figure 5.5: $\Sigma(T, X)$ for $X \in [10, 100]$.

It is clear from Figure 5.5 that the graph is unaffected by the higher precision of eigenvalues and that the peak poins have a very durable nature in terms of the number of eigenvalues. The last statement will be more evident from the T-plots that follow.

We now focus on plots in terms of T for a fixed X. In the next plot, Figure 5.6, we see again the vanishing of the imaginary part at the peak point $X = g_6$.

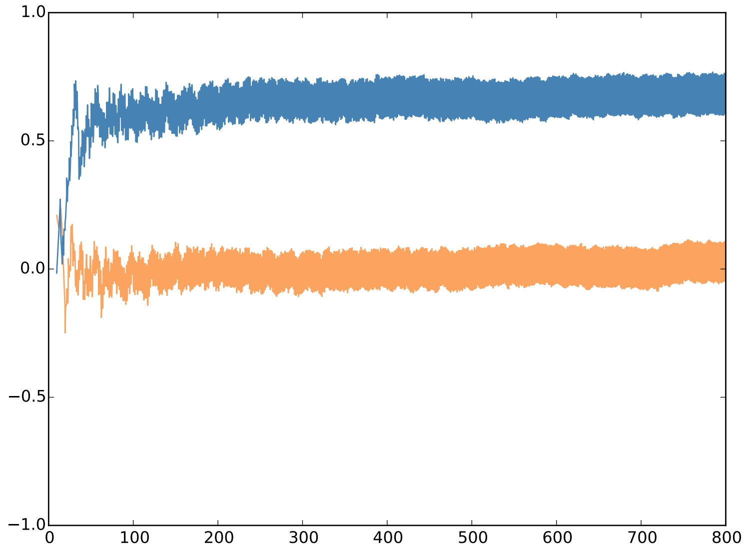


Figure 5.6: $\Sigma(T,X)$ at $X = 61.98$.

The robustness in terms of the number of eigenvalues can be noticed again, since the sum reaches its maximal value fairly quickly in terms of T. It is possible to predict fairly accurately how the main term of $S(T,X)$ should behave. We have already observed the peaks at (exponents of) the lengths of closed geodesics and at even powers of rational primes. From the next plot, Figure 5.7, it becomes more clear that the terms corresponding to the geodesics should dominate and the peaks at prime powers become smaller. Compare Figures 5.6 and 5.7.

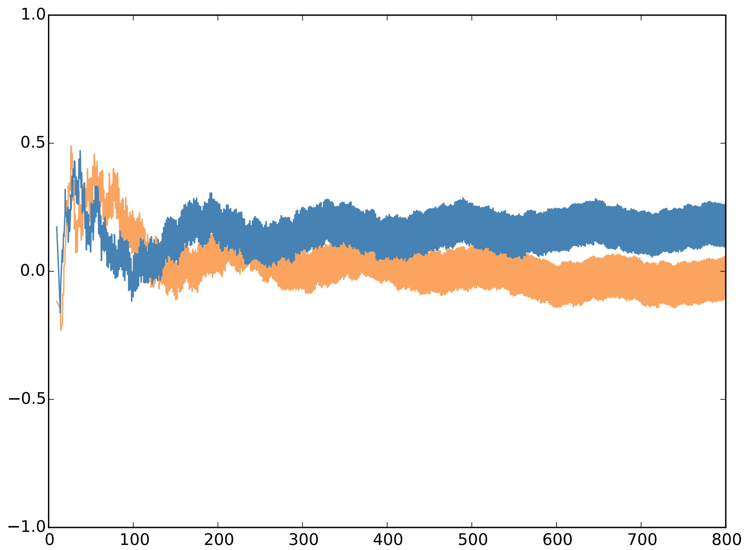
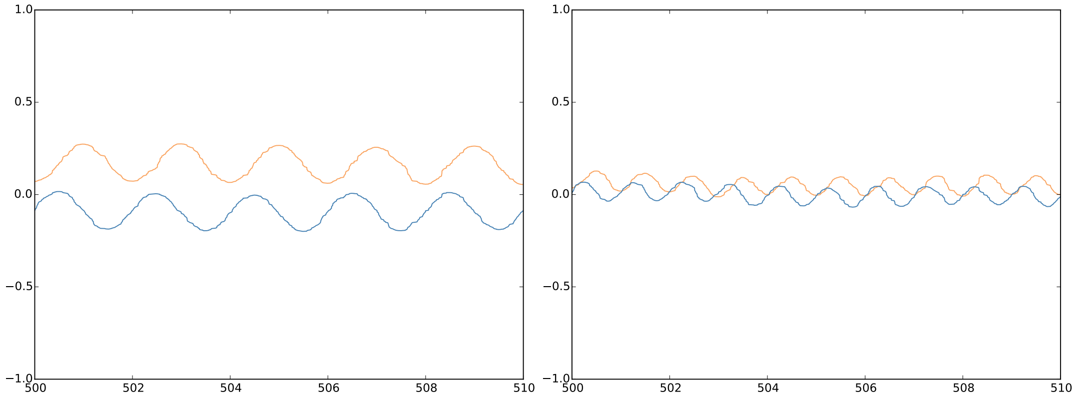


Figure 5.7: $\Sigma(T,X)$ at $X = 49$.

We can also predict the form of the oscillatory main term. Below we are at $X = e^{\pi}$

and $X = e^{2\pi}$.



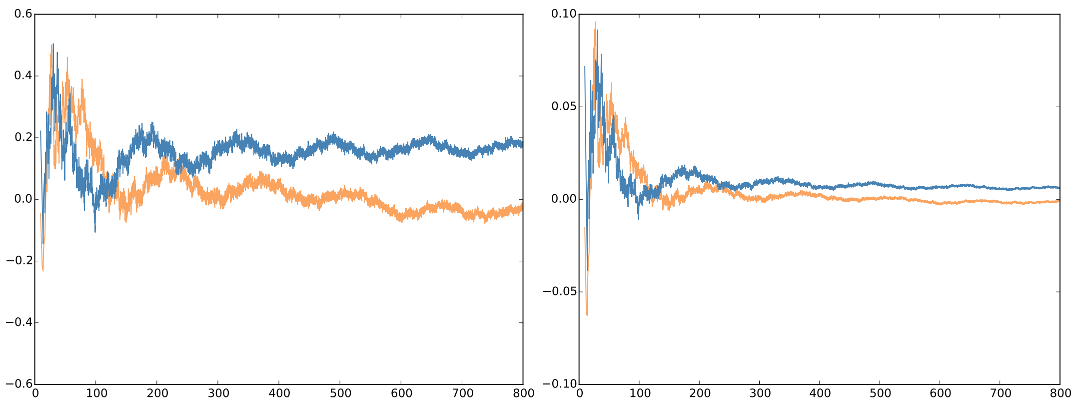
(a) $X = e^{\pi}$, (b) $X = e^{2\pi}$,

Figure 5.8: $\Sigma(T,X)$ at different values of X.

Notice that in intervals of T of equal length, there are twice as many periods of oscillations in Figure 5.8b than in Figure 5.8a. Similar investigation in the X-plots allows us to predict a component of the main term of the form $cT\sin(T\log X)$ for some constant c. In Theorem 5.2 we of course identify this oscillation precisely. We subtract it from $S(T,X)$ and define

$$\widetilde{\Sigma}(T,X) = T\Sigma(T,X) - \frac{|\Gamma\backslash\mathbb{H}^2|}{\pi}\frac{\sin(T\log X)}{\log X}T.$$

Next we focus on our choice of normalisation by T^{-1}. In Figure 5.9 we plot $\widetilde{\Sigma}(T,X)$ in terms of T at $X = 49$, which is one of the peak points.



(a) $\widetilde{\Sigma}(T,X)T^{-1}$, (b) $\widetilde{\Sigma}(T,X)T^{-3/2}$,

Figure 5.9: Different normalisations of $\widetilde{\Sigma}(T,X)$ at $X = 49$.

Clearly the normalisation T^{-1} seems to be closer to the correct one (notice the different ranges for the magnitude), which is evidence towards our Experimental Observation 1. On the website [59] we provide animations of the T-plots, which present the above observations in a more accessible way.

It is of interest to compare the behaviour of $S(T,X)$ with a similar sum over the Riemann zeros, even just to verify the correctness of our programs. Recall that Landau's formula says for a fixed $x>1$ that

$$\sum_{0<\gamma\le T} x^{\rho} = -\frac{T}{2\pi}\Lambda(x) + O(\log T). \tag{5.5}$$

We denote the left-hand side of (5.5) by $Z(T,x)$. We used our program with 10000 zeros of $\zeta(s)$ to 9 decimal places, provided by Odlyzko [73]. With our program we obtain the following plot for the normalised sum $T^{-1}Z(T,x)$.

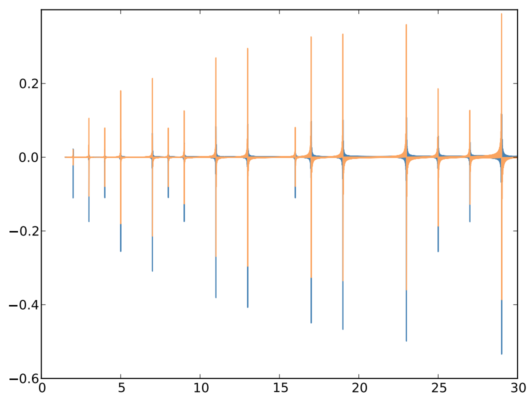


Figure 5.10: $T^{-1}Z(T,x)$ for $x\in[1.5,30]$.

Notice how the imaginary part has both a negative and positive peak at the prime powers as was the case with $\Sigma(T,X)$.

5.2 Application of the Selberg Trace Formula

We now prove Theorem 5.2. Let ψ be a positive even test function supported on $[-1,1]$ with $\int\psi=1$. Then define $\psi_\epsilon(x)=\epsilon^{-1}\psi(x/\epsilon)$. So ψ_ϵ is supported on $[-\epsilon,\epsilon]$ and $\int\psi_\epsilon=1$. Also, let G be the convolution $G(r)=(\mathbb{1}_{[-T,T]}*\psi_\epsilon)(r)$ for some $\epsilon>0$ to be chosen later. Define a function h, depending on T, X and ϵ, given by

$h(r)=G(r)(X^{ir}+X^{-ir})$. Let g be the Fourier transform of h as in [53, (1.64)]. Then the Selberg Trace Formula says that

$$\mathscr{S}(T,X)+\frac{1}{4\pi}\int_{-\infty}^{\infty}h(r)\frac{-\varphi'}{\varphi}\left(\frac{1}{2}+ir\right)dr=I(T,X)+H(T,X)+E(T,X)+L(T,X),$$

where

$$\begin{aligned}
\mathscr{S}(T,X)&=\sum_{j>0}h(t_j),\\
I(T,X)&=\frac{|F|}{4\pi}\int_{-\infty}^{\infty}h(r)r\tanh(\pi r)\,dr,\\
H(T,X)&=\sum_{\mathfrak{p}}\sum_{\ell=1}^{\infty}\left(N\mathfrak{p}^{\ell/2}-N\mathfrak{p}^{-\ell/2}\right)^{-1}g(\ell\log N\mathfrak{p})\log N\mathfrak{p},\\
E(T,X)&=\sum_{\mathscr{R}}\sum_{0<\ell<m}\left(2m\sin\frac{\pi\ell}{m}\right)^{-1}\int_{-\infty}^{\infty}h(r)\frac{\cosh\pi\left(1-\frac{2\ell}{m}\right)r}{\cosh\pi r}\,dr,\\
L(T,X)&=\frac{h(0)}{4}(1-\varphi(\tfrac{1}{2}))-g(0)\log 2-\frac{1}{2\pi}\int_{-\infty}^{\infty}h(r)\frac{\Gamma'}{\Gamma}(1+ir)\,dr,
\end{aligned}$$

where φ is the determinant of the scattering matrix, $|F|$ is the volume of the fundamental domain of $\Gamma\backslash\mathbb{H}^2$, $\mathfrak{p}$ and $\mathscr{R}$ range over the primitive hyperbolic and elliptic classes of $\mathrm{PSL}_2(\mathbb{Z})$, respectively. First observe that $S(T,X)=\mathscr{S}(T,X)+O(T\epsilon)$, so we can work with $\mathscr{S}$. For the identity motion we have

$$\begin{aligned}
I(T,X)&=\frac{|F|}{2\pi}\int_{-\infty}^{\infty}G(r)\cos(r\log X)r\tanh\pi r\,dr\\
&=\frac{|F|}{\pi}\int_{0}^{\infty}G(r)\cos(r\log X)r\left(1-\frac{2}{e^{2\pi r}+1}\right)dr\\
&=\frac{|F|}{\pi}(I_1(T,X)+I_2(T,X)).
\end{aligned}$$

From I_1 we obtain a part of the main term:

$$\begin{aligned}
I_1(T,X)&=\int_{0}^{\infty}G(r)\cos(r\log X)r\,dr\\
&=\left(\int_{0}^{T-\epsilon}+\int_{T-\epsilon}^{T+\epsilon}\right)G(r)\cos(r\log X)r\,dr\\
&=I_{11}+I_{12},
\end{aligned}$$

since G is even and supported on $[-T-\epsilon, T+\epsilon]$. Then,

$$I_{11} = \int_0^{T-\epsilon} \cos(r\log X) r\,dr = \frac{\sin((T-\epsilon)\log X)}{\log X}(T-\epsilon) + O(1),$$
$$I_{12} \ll \int_{T-\epsilon}^{T+\epsilon} r\,dr = O(T\epsilon).$$

Also,

$$I_2(T,X) = -\int_0^\infty G(r)\cos(r\log X) r \frac{2}{e^{2\pi r}+1}\,dr \ll \int_0^\infty r e^{-2\pi r}\,dr = O(1).$$

For $g(r)$ we compute

$$\begin{aligned} g(r) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-irt} h(t)\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} G(t) e^{-irt}(e^{it\log X} + e^{-it\log X})\,dt \\ &= \frac{1}{2\pi}\left(\widehat{G}\left(\frac{r-\log X}{2\pi}\right) + \widehat{G}\left(\frac{r+\log X}{2\pi}\right)\right). \end{aligned}$$

Here $\widehat{G}$ denotes the Fourier transform of G. So in particular $g(\ell \log N\mathfrak{p}) \sim T/\pi$ if $X = N\mathfrak{p}^\ell$ and decays as $O((\ell\log N\mathfrak{p})^{-k-1}\epsilon^{-k})$ otherwise, for any $k \in \mathbb{N}$. For the elliptic terms we need to evaluate

$$\int_{-\infty}^{\infty} h(r)\frac{\cosh \pi(1-\frac{2\ell}{m})r}{\cosh \pi r}\,dr \ll \int_0^\infty \frac{e^{-2\pi r\ell/m} + e^{-2\pi r}}{1+e^{-2\pi r}}\,dr = O(1).$$

Hence $E(T,X)$ is bounded. By the explicit formula (4.7) of φ for $\mathrm{PSL}_2(\mathbb{Z})$, we have

$$\begin{aligned} \int_{-\infty}^{\infty} h(r)\frac{-\varphi'}{\varphi}\left(\frac{1}{2}+ir\right)dr &= \int_{-\infty}^{\infty} h(r)\left(-2\log\pi + \frac{\Gamma'}{\Gamma}\left(\frac{1}{2}\pm ir\right) + 2\frac{\zeta'}{\zeta}(1\pm 2ir)\right)dr. \\ &= C_1 + C_2 + C_3. \end{aligned}$$

The integral C_1 is the Fourier transform of G and is thus bounded. For C_2 we use Stirling asymptotics to get

$$C_2 = \int_{-\infty}^{\infty} h(r)\log(\tfrac{1}{4}+r^2)\,dr + O(1),$$

which is $O(\log T)$. The same computation shows that $L(T,X) = O(\log T)$. The remaining part of the main term comes from C_3. We first expand h and isolate the important terms:

$$C_3 = 2\left(\int_{-T-\epsilon}^{-T+\epsilon} + \int_{-T+\epsilon}^{T-\epsilon} + \int_{T-\epsilon}^{T+\epsilon}\right)(X^{ir}+X^{-ir})G(r)\frac{\zeta'}{\zeta}(1\pm 2ir)\,dr.$$

The first and third integrals are bounded by $O(\epsilon \log T)$. Notice that $G(r)=1$ in the range of the second integral, hence we can write it as

$$2\int_{\frac{1}{2}-(T-\epsilon)i}^{\frac{1}{2}+(T-\epsilon)i}(X^{s-1/2}+X^{1/2-s})\left(\frac{\zeta'}{\zeta}(2s)+\frac{\zeta'}{\zeta}(2-2s)\right)ds.$$

We separate this into two integrals by adding and subtracting the singular part:

$$\begin{aligned}
C_3 &= 2\int_{\frac{1}{2}-(T-\epsilon)i}^{\frac{1}{2}+(T-\epsilon)i}(X^{s-1/2}+X^{1/2-s})\left(\frac{\zeta'}{\zeta}(2s)-\frac{1}{2s-1}\right)ds \\
&\quad + 2\int_{\frac{1}{2}-(T-\epsilon)i}^{\frac{1}{2}+(T-\epsilon)i}(X^{s-1/2}+X^{1/2-s})\left(\frac{\zeta'}{\zeta}(2-2s)-\frac{1}{(2-2s)-1}\right)ds \\
&= 2(C_{31}+C_{32}).
\end{aligned}$$

For the first integral we move the contour to $\operatorname{Re} s=1$ and for the second one to $\operatorname{Re} s=0$. It is easy to see that the top and bottom parts of the contours yield $O(\log T)$. For the line at $\operatorname{Re} s=1$ we get

$$\begin{aligned}
C_{31} &= \int_{1-(T-\epsilon)i}^{1+(T-\epsilon)i}(X^{s-1/2}+X^{1/2-s})\left(\frac{\zeta'}{\zeta}(2s)-\frac{1}{2s-1}\right)ds \\
&\qquad = \int_{-T+\epsilon}^{T-\epsilon}(X^{1/2+ir}+X^{-1/2-ir})\left(\frac{\zeta'}{\zeta}(2+2ir)-\frac{1}{1+2ir}\right)dr,
\end{aligned}$$

For the rest of the proof we will follow an argument due to Landau [60, Hilfssatz 2]. We start by writing out the Dirichlet series:

$$\begin{aligned}
C_{31} &= -\int_{-T+\epsilon}^{T-\epsilon} X^{1/2+ir}\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^{2+2ir}}\,dr + O(\log T) \\
&= -\sum_{n\neq\sqrt{X}}\frac{\sqrt{X}\Lambda(n)}{n^2}\int_{-T+\epsilon}^{T-\epsilon}\left(\frac{X}{n^2}\right)^{ir}dr - X^{-1/2}\Lambda(X^{1/2})\int_{-T+\epsilon}^{T-\epsilon}dr + O(\log T).
\end{aligned}$$

Since $X > 1$, the term in C_{31} with the negative exponent gets absorbed into the error term. Hence,

$$\begin{aligned}\left|\int_{-T+\epsilon}^{T-\epsilon} X^{1/2+ir}\frac{\zeta'}{\zeta}(2+2ir)\,dr + 2X^{-1/2}\Lambda(X^{1/2})(T-\epsilon)\right| \\ \leq \sum_{n\neq\sqrt{X}} \frac{\sqrt{X}\Lambda(n)}{n^2}\left|\frac{(\frac{X}{n^2})^{i(T-\epsilon)} - (\frac{X}{n^2})^{-i(T-\epsilon)}}{\log(X/n^2)}\right| \\ \ll 2\sqrt{X}\left|\frac{\zeta'}{\zeta}(2)\right|.\end{aligned}$$

So we see that $C_{31} = -2X^{-1/2}\Lambda(X^{1/2})(T-\epsilon) + O(\log T)$. A similar argument shows that C_{32} has the same asymptotics. Letting $\epsilon = 1/\log T$ concludes the proof of Theorem 5.2. □

Remark 5.1. It is not as easy to obtain results about the sine kernel through this method. First of all, the Selberg Trace Formula is only valid for even functions so we cannot directly apply it. Formally, we can attempt to differentiate the expression for the cosine kernel and sum by parts to obtain the desired sum, but there are convergence issues to worry about. Even then, it is difficult to see the behaviour that we are interested in. Based on the discussion in the previous section we expect the sine kernel to vanish at the peak points, but to be of large order around them. It is not clear how one could see this through the Selberg Trace Formula.

Chapter 6

Lattice Point Problems

Let Γ be a discrete group acting discontinuously on a hyperbolic space $\mathscr{H}$ and denote the quotient space by M. The standard hyperbolic lattice point problem asks to count the number of points in the orbit Γp within a given distance from some fixed point $q \in \mathscr{H}$. For example, in two dimensions the counting function is

$$N(z,w,X) = \#\{\gamma \in \Gamma : 4u(\gamma z, w) + 2 \leq X\},$$

where u is the standard point-pair invariant on $\mathbb{H}^2$ and $z, w \in \mathbb{H}^2$, and it measures the number of lattice points γz in a hyperbolic disc of radius arccosh$(X/2)$ centered at w. This problem was first considered by e.g. Huber and Selberg. In the 1970s Selberg proved that for fixed $z, w \in \mathbb{H}^2$,

$$N(z,w,X) = \sqrt{\pi} \sum_{s_j \in (1/2,1]} \frac{\Gamma(s_j - 1/2)}{\Gamma(s_j+1)} X^{s_j} u_j(z)\overline{u}_j(w) + E(z,w,X),$$

where $E(z,w,X) = O(X^{2/3})$. The bound on the error term $E(z,w,X)$ has not been improved for any cofinite Γ or any choice of points $z, w \in \mathbb{H}^2$. To find more evidence of the conjectured bound for the error term $E(z,w,X) = O(X^{1/2+\epsilon})$, it is useful to consider various averages. For example, Hill and Parnovski [47] look at the variance of the counting function in terms of the centre over the whole fundamental domain of any cofinite Γ in hyperbolic n-space. For the case $\mathscr{H} = \mathbb{H}^2$ and $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, assuming no non-zero eigenvalues $\lambda \leq 1/4$, their result is

$$\int_{\Gamma\backslash\mathbb{H}^2} \left| N(z,w,X) - \frac{\pi}{\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)} X \right|^2 d\mu(w) = O(X),$$

where $\mu(w)$ is the standard hyperbolic measure on $\mathbb{H}^2$. On the other hand, Petridis and Risager [77] looked at a local average of $N(z,z,X)$ over z. Suppose that f is a smooth, non-negative, compactly supported function on M. For $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ they proved that

$$\int_{\Gamma\backslash\mathbb{H}^2} f(z)N(z,z,X)\,d\mu(z) = \frac{\pi X}{\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)}\int_{\Gamma\backslash\mathbb{H}^2} f(z)\,d\mu(z) + O(X^{7/12+\epsilon}),$$

where the error term depends on ϵ and f only. This improves Selberg's bound halfway to the expected $1/2+\epsilon$ on average. Their method requires knowledge of the average rate of QUE for Maaß cusp forms on M and other arithmetic information only available to groups similar to $\mathrm{PSL}_2(\mathbb{Z})$. In 1996 Chamizo [10] showed that it is possible to apply large sieve methods on M. As an application, he proved that by averaging over a large number of radii, one gets the expected bound on the error term $E(z,w,X)$:

$$\frac{1}{X}\int_X^{2X} |E(z,w,x)|^2\,dx = O(X\log^2 X). \tag{6.1}$$

Furthermore, Chamizo also proves a similar result for the second and fourth moments of discrete averages over sufficiently spaced centres, which leads to

$$\left(\int_{\Gamma\backslash\mathbb{H}^2} |E(z,w,X)|^{2m}\,d\mu(z)\right)^{\frac{1}{2m}} = O(X^{1/2}\log X), \tag{6.2}$$

for $m = 1,2$.

Instead of measuring the distance between two points of $\mathscr{H}$, it is also possible to consider geodesic segments between various subspaces of M. In two dimensions, Huber [50] looked at geodesic segments between a point and a fixed closed geodesic ℓ. The geodesic ℓ corresponds to a hyperbolic conjugacy class $\mathfrak{H}$, given by some power ν of a primitive hyperbolic element $g \in \Gamma$. For cocompact Γ, Huber explained that counting

$$N_z(T) = \#\{\gamma \in \mathfrak{H} : d(z,\gamma z) \le T\} \tag{6.3}$$

is equivalent to counting the geodesic segments from z to ℓ according to length. If Γ has no small eigenvalues, then Huber's main result in [49] says that

$$N_z(T) = \frac{2}{\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)}\frac{\mu}{\nu}X + O(X^{3/4}), \tag{6.4}$$

where μ is the length of the invariant geodesic corresponding to $\mathfrak{H}$, and $X = \frac{\sinh T/2}{\sinh \mu/2}$. Independently, Good [35] proved a stronger error bound of $O(X^{2/3})$. Good's methods also extend to more general counting problems for cofinite groups Γ.

There is another interesting geometric interpretation that Huber gave for the counting problem in conjugacy classes. After conjugation we may assume that the geodesic ℓ lies on the imaginary axis. Then the counting in $N_z(T)$ is equivalent to counting γz in the cosets $\gamma \in \Gamma/\langle g\rangle$, such that γz lies inside the sector formed by the imaginary axis and some angle Θ.

Chatzakos and Petridis [14] showed that it is possible to apply Chamizo's methods to obtain results analogous to (6.1) and (6.2). This was done by extending Huber's method to produce Good's error term $O(X^{2/3})$ for both cocompact and cofinite Γ.

In n dimensions, Herrmann [46] investigated the number of geodesic segments from a point to any Jordan measurable subset Y of a totally geodesic submanifold $\mathscr{Y} \subset \mathscr{H}$. Let $N(r, Y, \Gamma p)$ be the number of orthogonal geodesic segments from γp, for any $\gamma \in \Gamma$, to Y with length at most r. For cocompact Γ, Herrmann proves that

$$N(r, Y, \Gamma p) \sim \frac{2}{n-1} \frac{\pi^{(n-k)/2}}{\Gamma(\frac{n-k}{2})} \frac{\mathrm{vol}(Y)}{\mathrm{vol}(\Gamma\backslash\mathscr{H})} \cosh^{n-1} r. \tag{6.5}$$

His method is geometric, not depending on the action of the group Γ on $\mathscr{H}$. He introduces an associated Dirichlet series and studies its analytic continuation. It is difficult to prove strong error terms with this method. We are interested in the error term of (6.5) for $n = 3$ and $k = n-1$. We study this for $Y = \mathscr{Y}$ by adapting the method of Huber and Chatzakos–Petridis to $\mathscr{H} = \mathbb{H}^3$ for cocompact $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$.

For the rest of this chapter we fix $\mathscr{H} = \mathbb{H}^3$. We also focus solely on cocompact Γ, see Remark 6.6 for a discussion on cofinite groups. Let $\{u_j\}_{j\geq 0}$ be a complete orthonormal system of eigenfunctions of the Laplacian Δ on M with eigenvalues $\lambda_j = s_j(2-s_j) \geq 0$. Let $\mathscr{P}$ be a totally geodesic hyperplane in $\mathscr{H}$ and define $v(p) = \arctan(x_2(p)/y(p))$. We prove the following theorem.

Theorem 6.1. *Let Γ be a cocompact discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. Set $H = \Gamma \cap \mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{C})}(\mathscr{P})$ and let $\hat{u}_j$ be the period integral of u_j over the fundamental domain of H restricted to $\mathscr{P}$. Define*

$$N(p, X) = \#\{\gamma \in H\backslash\Gamma : (\cos v(\gamma p))^{-1} \leq X\}.$$

Then we have

$$N(p, X) = M(p, X) + E(p, X),$$

where

$$M(p, X) = \frac{\mathrm{vol}(H\backslash\mathscr{P})}{\mathrm{vol}(\Gamma\backslash\mathscr{H})} X^2 + \sum_{1<s_j<2} \frac{2^{s_j-1}}{s_j} \hat{u}_j u_j(p) X^{s_j}, \tag{6.6}$$

and

$$E(p,X) = O(X^{3/2}).$$

Here we understand $\mathrm{vol}(H\backslash\mathscr{P})$ as the hyperbolic area in two dimensions. We also apply Chamizo's large sieve results in this case. For the radial average, Chamizo only provides a large sieve inequality in two dimensions. We generalise it to three dimensions and prove an improvement of $1/6$ on the pointwise error term on average. It is clear for structural reasons that the radial averages with this method get worse for higher dimensions. See Section 6.3.4 for a more indepth discussion. In the same vein it is only possible to obtain the conjectured bound for the second moment for spatial averages. A similar feature can already be observed in two dimensions. Chamizo proves large sieve inequalities for all moments of the mean square in his thesis [12]. However, it turns out that for the spatial average only the second and fourth moments yield improvements over the pointwise bound and likewise for the second moment in the radial case.

We summarise our mean square results in the following theorems. The corresponding discrete averages and more precise statements are given in Theorems 6.13 and 6.14.

Theorem 6.2. *Let Γ be a cocompact discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. Then, for $X > 1$,*

$$\frac{1}{X}\int_X^{2X} |E(p,x)|^2\,dx \ll X^{2+2/3}\log X.$$

Remark 6.1. It is also possible to obtain the radial mean square (6.1) in the standard two dimensional lattice point problem by direct integration in the spectral expansion of the error term. This is done for a smoothed error term by Phillips and Rudnick [85]. It is possible to deduce the result of Chamizo from their computations [16]. It would be interesting to see if this can be done in our problem and whether it improves on the above estimate coming from the large sieve.

Theorem 6.3. *Let Γ be a cocompact discrete subgroup of $\mathrm{PSL}_2(\mathbb{C})$. Then, for $X > 1$,*

$$\int_{\Gamma\backslash\mathscr{H}} |E(p,X)|^2\,d\mu(p) \ll X^2\log^2 X.$$

Remark 6.2. The other possible case in three dimensions, $k = 1$, is substantially harder with our method. Currently, the spectral expansion of the automorphic function corresponding to $N(p,X)$ can be written in terms of $\hat{u}_j$ and an explicit solution to an ordinary differential equation. For $k = 1$ we can no longer solve the corresponding eigenvalue equation as it remains a partial differential equation.

Remark 6.3. The majority of computations in this paper are more elementary (explicit), though not necessarily shorter, than in two dimensions (cf. [14]). While the spectral analysis is analogous, the three dimensional geometry introduces new obstacles that were not present in the lower dimension. For example, we have to pay more attention to the fixed submanifold and the period integrals defined over it. We expect that there will always be a distinction between even and odd dimensional hyperbolic spaces. This is due to the fact that the special functions in the spectral expansion of the counting function (see Lemma 6.7 and (6.27), (6.28), (6.29)), the spherical eigenfunctions of Δ and the Selberg transform can all be expressed in an elementary form. The spectral analysis should work similarly in all dimensions in the case $k = n-1$. For the most part, we prefer to work without relying on the explicit expressions too much. Where this is not convenient, we give remarks as to how we expect the computations to generalise to higher dimensions.

Remark 6.4. Dynamical systems and ergodic methods have also been applied to study lattice point counting. Their advantage is that the results generally apply to a larger set of manifolds. On the other hand, these methods fail to produce finer results, such as strong error terms. For example, Parkkonen and Paulin [75] extend the counting in conjugacy classes problem of Huber to higher dimensions for loxodromic, parabolic and elliptic conjugacy classes for any discrete group of isometries Γ. See also [74] for a survey on a wider variety of counting problems analogous to Herrmann [46]. Moreover, Eskin and McMullen [26] obtain main terms for a variety of counting problems on affine symmetric spaces defined by Lie groups. In particular, they give an alternate proof for the main term on homogenous affine varieties, which was also proved by Duke, Rudnick, and Sarnak [23] through spectral methods.

6.1 The Geometric Setup of the Problem

Recall Huber's counting function $N_z(T)$ as defined in (6.3). With the notation from the beginning of the chapter, define

$$A(f)(z) = \sum_{\gamma \in \mathfrak{H}} f\left(\frac{\cosh d(z, \gamma z) - 1}{\cosh \mu - 1}\right),$$

where $\mathfrak{H}$ is the conjugacy class, μ the length of the invariant geodesic and f is some compactly supported function with finitely many discontinuities. If f is the indicator function on $[0, T]$, then $A(f)(z) = N_z(T)$. The key idea in Huber's proof of (6.4) is the identification of a coordinate system in which the coefficients $a(f, t_j)$ of the spectral expansion of $A(f)$ decompose into a product $a(f,t_j)=2\hat{u}_j d(f,t_j)$. Here $\hat{u}_j$ is a period integral of the Maaß form u_j over a segment of the invariant geodesic and $d(f,t_j)$ is an integral transform depending on a solution of the integrated eigenvalue equation.

The $d(f,t)$ transform plays the role of the Selberg–Harish-Chandra transform from Selberg theory. Chatzakos and Petridis identified the special function in the integral transform explicitly, which was crucial for the application of the large sieve to the mean square of the error term. In three dimensions there are more possibilities for generalisation, depending on the dimension of the totally geodesic submanifold. The case of a geodesic and a point is difficult as the eigenvalue equation remains a partial differential equation, which we cannot solve. Instead, we focus on the problem with a totally geodesic hyperplane and a point. Geometrically, the former case corresponds to counting in a cone, while the latter is still counting in a sector.

The totally geodesic hypersurfaces in $\mathcal{H}$ are the Euclidean hyperplanes and semi-spheres orthogonal to the complex plane. Motivated by Huber, we define a new set of coordinates. Let

$$x=x_1, \qquad u=\log\sqrt{x_2^2+y^2}, \qquad v=\arctan\frac{x_2}{y},$$

and transform to $p=(x(p),u(p),v(p))$. We often write (x,u,v) for the same point as a shorthand if the point in question is clear. The effect of this change of coordinates on the metric and the Laplacian is summarised in the next lemma.

Lemma 6.4. *With the (x,u,v) coordinates defined as above, we have*

$$\begin{aligned} ds^2 &= \frac{dx^2}{e^{2u}\cos^2 v}+\frac{du^2+dv^2}{\cos^2 v},\\ d\mu(p) &= \frac{dx\,du\,dv}{e^u\cos^3 v}, \end{aligned}$$

and

$$\Delta = e^{2u}\cos^2 v\frac{\partial^2}{\partial x^2}+\cos^2 v\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right)-\cos^2 v\frac{\partial}{\partial u}+\sin v\cos v\frac{\partial}{\partial v}.$$

Proof. Recall that given a Riemannian metric

$$ds^2=\sum_{i,j} g_{ij}\,dx^i dx^j,$$

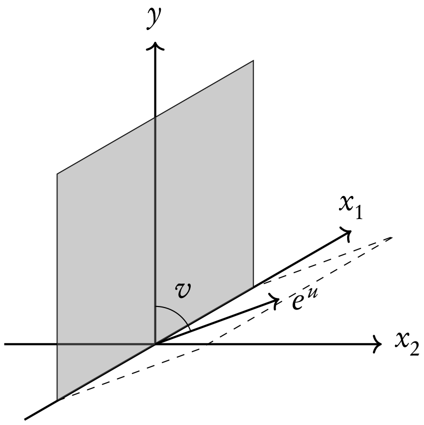


Figure 6.1: The (x,u,v) coordinates in $\mathscr{H}$.

then the associated Laplace–Beltrami operator is given by

$$\Delta = \frac{1}{\sqrt{g}} \sum_{i,j} \frac{\partial}{\partial x^i} \sqrt{g}\, g^{ij} \frac{\partial}{\partial x^j}, \tag{6.7}$$

where $(g^{ij}) = (g_{ij})^{-1}$ is the inverse metric tensor and $g = \det(g_{ij})$. Now, the Jacobian of the transformation $(x_1, x_2, y) \mapsto (x, u, v)$ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin v\, e^u & \cos v\, e^u \\ 0 & \cos v\, e^u & -\sin v\, e^u \end{pmatrix},$$

so that the hyperbolic metric tensor in these coordinates is

$$(g_{ij}) = \begin{pmatrix} \frac{1}{e^{2u}\cos^2 v} & & \\ & \frac{1}{\cos^2 v} & \\ & & \frac{1}{\cos^2 v} \end{pmatrix}. \tag{6.8}$$

We can then read off ds^2 from (6.8) and compute $d\mu(p)$ as the square root of the determinant of (g_{ij}). For the Laplacian we use formula (6.7) to get

$$\Delta = e^u \cos^3 v \left(\frac{\partial}{\partial x} \left(\frac{e^u}{\cos v} \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial u} \left(\frac{1}{e^u \cos v} \frac{\partial}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{1}{e^u \cos v} \frac{\partial}{\partial v} \right) \right),$$

which simplifies to the required form. □

Now, let $\mathscr{P}$ be a totally geodesic hyperplane in $\mathscr{H}$. After conjugation by an element of $\mathrm{PSL}_2(\mathbb{C})$, we may assume that $\mathscr{P}$ is given by the set $\{p \in \mathscr{H} : v = 0\}$ (i.e. $x_2 = 0$, see Figure 6.1). Let $p \in \mathscr{H}$. We denote the orthogonal projection (along geodesics) of p onto $\mathscr{P}$ by $p_0 = (x(p), u(p), 0)$. Next we identify all the elements of $\mathrm{PSL}_2(\mathbb{C})$

that stabilise the plane $\mathscr{P}$. Since we are no longer working with a single geodesic, the stabiliser will be larger than in the two dimensional setting.

Lemma 6.5. *The stabiliser of* $\mathscr{P}$ *in* $\mathrm{PSL}_2(\mathbb{C})$ *is*

$$\mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{C})}(\mathscr{P}) = \mathrm{PSL}_2(\mathbb{R}) \bigcup \begin{pmatrix} i & \\ & -i \end{pmatrix} \mathrm{PSL}_2(\mathbb{R}).$$

Proof. Denote the stabiliser by A. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in A$. By definition we have that $\gamma p \in \mathscr{P}$ for any $p \in \mathscr{P}$, that is $x_2(p) = 0$ implies that $x_2(\gamma p) = 0$. Hence,

$$x_2(\gamma p) = \frac{\mathrm{Im}(a\bar{c}x_1^2 + (a\bar{d} + b\bar{c})x_1 + b\bar{d} + a\bar{c}y^2)}{||cp + d||^2} = 0.$$

This needs to be true for any x_1 and y. Comparing coefficients we get

$$\mathrm{Im}(a\bar{c}) = 0, \tag{6.9}$$
$$\mathrm{Im}(a\bar{d} + b\bar{c}) = 0, \tag{6.10}$$
$$\mathrm{Im}(b\bar{d}) = 0. \tag{6.11}$$

Now, write a, b, c, d in polar form as

$$a = r_a e^{\theta_a i}, \quad b = r_b e^{\theta_b i},$$
$$c = r_c e^{\theta_c i}, \quad d = r_d e^{\theta_d i}.$$

Then equation (6.9) implies that $\theta_a = \theta_c + n\pi$ for some $n \in \mathbb{Z}$. Similarly, from (6.11) we see that $\theta_b = \theta_d + m\pi$ for some $m \in \mathbb{Z}$. Equation (6.10) then gives that

$$\mathrm{Im}(r_a r_d e^{i(\theta_a - \theta_b + m\pi)} + r_b r_c e^{i(\theta_b - \theta_a + n\pi)}) = 0.$$
$$\iff (-1)^m r_a r_d \sin(\theta_a - \theta_b) - (-1)^n r_c r_d \sin(\theta_a - \theta_b) = 0. \tag{6.12}$$

We also have from $ad - bc = 1$ that

$$\mathrm{Im}(r_a r_d e^{i(\theta_a + \theta_b - m\pi)} - r_b r_c e^{i(\theta_b + \theta_a - n\pi)}) = 0.$$
$$\iff (-1)^m r_a r_d \sin(\theta_a + \theta_b) - (-1)^n r_c r_d \sin(\theta_a + \theta_b) = 0. \tag{6.13}$$

Also from the real part of $ad - bc = 1$ we get

$$(-1)^m r_a r_d \cos(\theta_a + \theta_b) - (-1)^n r_c r_d \cos(\theta_a + \theta_b) = 1, \tag{6.14}$$

which means that

$$(-1)^m r_a r_d - (-1)^n r_c r_d \neq 0. \tag{6.15}$$

Combining (6.12), (6.13) and (6.15) yields

$$\sin(\theta_a - \theta_b) = 0, \quad \sin(\theta_a + \theta_b) = 0,$$

and similarly for θ_c and θ_d. Thus either $\theta_a, \theta_b, \theta_c, \theta_d \in \{0, \pi\}$ or $\theta_a, \theta_b, \theta_c, \theta_d \in \{\frac{-\pi}{2}, \frac{\pi}{2}\}$. In the first case the matrix is real and hence gives $\mathrm{PSL}_2(\mathbb{R})$. In the second case we have, after considering the determinant,

$$\gamma = \begin{pmatrix} i & \\ & -i \end{pmatrix}\begin{pmatrix} r_a & r_b \\ r_c & r_d \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} r_a & r_b \\ r_c & r_d \end{pmatrix}\begin{pmatrix} i & \\ & -i \end{pmatrix}.$$

□

Now define $H = \mathrm{Stab}_{\mathrm{PSL}_2(\mathbb{C})}(\mathscr{P}) \cap \Gamma$. We can then write the counting function as

$$\widetilde{N}(p, \Theta) = \#\{\gamma \in H\backslash\Gamma : |v(\gamma p)| \leq \Theta\},$$

where $\Theta \in (0, \pi/2)$. If we set $(\cos\Theta)^{-1} = X$, then $\widetilde{N}$ takes on the following form (cf. Huber [49])

$$\widetilde{N}(p, \Theta) = N(p, X) = \#\left\{\gamma \in H\backslash\Gamma : \frac{1}{\cos v(\gamma p)} \leq X\right\}.$$

Remark 6.5. Counting in the sector is equivalent to counting orthogonal geodesic segments from γp to $\mathscr{P}$ according to length. Hence, we can easily relate the main term (6.6) to that of Herrmann's (6.5). Given a point $p \in \mathscr{H}$, the projection p_0 in the (x_1, x_2, y)-coordinates is given by $p_0 = (x_1, 0, \sqrt{x_2^2 + y^2})$. Then, by the explicit formula for the point-pair invariant (4.21), we get

$$\delta(p, p_0) = \frac{\sqrt{x_2^2 + y^2}}{y},$$

which simplifes to

$$\delta(p, p_0) = \sec v,$$

in the (x, u, v)-coordinates. Now, for $N(p, X)$ we have $X = \sec\Theta$ so that the maximal distance we are counting is $\operatorname{arccosh}\sec\Theta$. Substituting this into (6.5) with $n = 3$ and $k = 2$ shows that the main terms agree.

As we are working with invariance under H, we have to compute its fundamental domain, $\mathscr{F}_{\mathscr{H}}(H)$. It has a particularly convenient description in the (x, u, v)-coordinates.

First, consider H restricted to the plane $\mathscr{P}$ and denote the fundamental domain of H in this space by $S = \mathscr{F}_{\mathbb{H}^2}(H)$.

Lemma 6.6. *We claim that $\mathscr{F}_{\mathscr{H}}(H) = F$, where F is given by the union of rotated copies of S:*

$$F = \bigcup_{\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})} S_\theta,$$

and S_θ are defined by

$$S_\theta = \{p \in \mathscr{H} \, : \, v(p) = \theta, p_0 \in S\}.$$

Proof. This follows immediately from computing the action of H on $\mathscr{H}$, which is seen to be independent of $v(p)$ in the x and u-coordinates. First, let $\iota = \left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$, which acts on $p \in \mathscr{H}$ as a rotation by π about the imaginary axis:

$$\iota p = \iota(z + yj) = -z + yj,$$

so that

$$x(\iota p) = -x(p), \qquad u(\iota p) = u(p), \qquad v(\iota p) = -v(p).$$

On the other hand, for $\tau = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{R})$, we find that

$$\begin{aligned} x(\tau p) &= \frac{ac|z|^2 + (ad + bc)x + bd + acy^2}{c^2|z|^2 + 2cdx + d^2 + c^2y^2} \\ &= \frac{ac||p||^2 + (ad + bc)x + bd}{c^2||p||^2 + 2cdx + d^2}. \end{aligned}$$

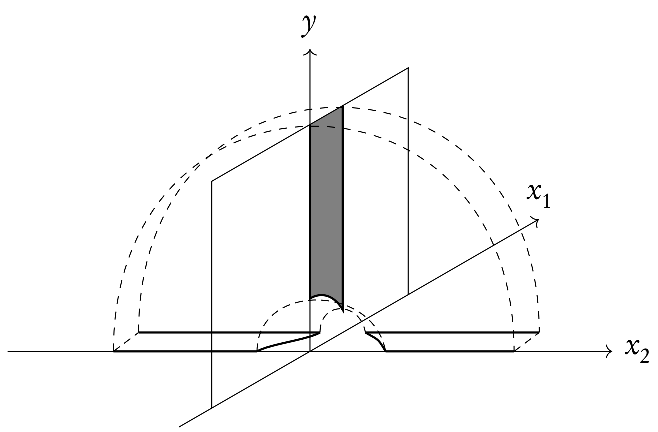


Figure 6.2: Fundamental domain of H in $\mathscr{H}$ with the part for $H\backslash\mathscr{P}$ highlighted (for $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$).

And since $||p||^2 = x^2 + e^{2u}$ it follows that $x(\tau p)$ does not depend on $v(p)$. Similarly,

$$x_2(\tau p) = \frac{x_2}{||cp+d||^2},$$
$$y(\tau p) = \frac{y}{||cp+d||^2},$$

so that

$$u(\tau p) = \log \frac{\sqrt{x_2^2+y^2}}{||cp+d||^2} = u(p) - \log||cp+d||^2,$$
$$v(\tau p) = v(p).$$

It is easy to see that $\iota\tau\iota \in \mathrm{SL}_2(\mathbb{R})$. Hence any $\gamma \in H$ can be written as $\iota\tau$, $\tau\iota$ or τ. The first consequence of the above calculations is that for $p \in \mathscr{H}$ and $\gamma \in H$ the action of the group and the orthogonal projection to $\mathscr{P}$ commute, i.e. $(\gamma p)_0 = \gamma p_0$. Moreover, if $v(p) = \theta$ then $v(\gamma p) = \pm\theta$ for any $\gamma \in H$. Thus, suppose that $p, \gamma p \in S_{\pm\theta}$ for some $\theta \in (0, \pi/2)$ and $\gamma \in H$. It follows that $p_0, \gamma p_0 \in S$, which is a contradiction as S is a fundamental domain for H on the plane. This shows that $\mathscr{F}_{\mathscr{H}}(H) \subseteq F$. Suppose, $\mathscr{F}_{\mathscr{H}}(H) \subsetneq F$. Then for some θ there is a point $p \in S_\theta$ and $\gamma \in H$ with $\gamma p \in F \setminus \mathscr{F}_{\mathscr{H}}(H)$. Projecting back to S gives a contradiction. $\square$

Remark 6.6. For cocompact Γ it is easy to see that $N(p, X)$ is uniformly bounded in p. On the other hand, for cofinite Γ it is not true in general, although it is still possible to see that $N(p, X)$ is well-defined (finite for fixed X and p). For example, if $\Gamma = \mathrm{PSL}_2(\mathbb{Z}[i])$ then as $y(p) \to \infty$, $N(p, X)$ becomes unbounded. This introduces complications for the convergence of the corresponding automorphic form and is the main reason for our restriction to cocompact groups. The problem lies in the fact that for non-compact M the totally geodesic surface can pass through the cusp. It should be possible to overcome this difficulty by restricting the group Γ appropriately.

6.2 Spectral Analysis

Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ be cocompact. Define a function $A(f)$ on $\mathscr{H}$ by

$$A(f)(p) = \sum_{\gamma \in H\backslash\Gamma} f\left(\frac{1}{\cos^2 v(\gamma p)}\right),$$

where $f : [1, \infty) \longrightarrow \mathbb{R}$ has a compact support and finitely many discontinuities. Then it is easy to see that $A(f)$ is automorphic since $v(\gamma p)$ is constant on the cosets. We also note that for sufficiently smooth $f \in C^2[1,\infty)$, A converges pointwise to its spectral expansion, see [50, pg. 23]. Let $\{u_j\}_{j\geq 0}$ be a complete orthonormal system of automorphic eigenfunctions of $-\Delta$ with corresponding eigenvalues λ_j. Since our problem differs from the standard lattice point counting problem, in that $A(f)$ does not define an automorphic kernel, we do not have the usual expansion in terms of the Selberg transform of f. The correct substitute for this is the spectral expansion of $A(f)$ in terms of the u_j's. Let $a(f,t_j)$ be the coefficients of the spectral expansion of $A(f)$ on $\Gamma\backslash\mathscr{H}$ given by

$$a(f,t_j) = \langle A(f), u_j\rangle = \int_{\Gamma\backslash\mathscr{H}} A(f)(p)\overline{u}_j(p)\,d\mu(p).$$

Then the spectral expansion of $A(f)$ in terms of the u_j's is

$$A(f)(p) = \sum_j a(f,t_j)u_j(p).$$

We now compute the coefficients $a(f,t_j)$ explicitly in the manner of [49, Lemma 2.3] and [14, Lemma 2.1]. Following this, we identify the special function that appears in the spectral expansion and prove some simple estimates on it. For simplicity, consider u_j instead of $\overline{u}_j$.

Lemma 6.7. *We have*

$$a(f,t_j) = 2\hat{u}_j c(f,t_j),$$

where

$$\hat{u}_j = \int_{H\backslash\mathscr{P}} u_j(x,u,0)\frac{du\,dx}{e^u},$$

is a period-integral of u_j over the fundamental domain of H restricted to the plane $\mathscr{P}$. Also,

$$c(f,t_j) = \int_0^{\frac{\pi}{2}} f\left(\frac{1}{\cos^2 v}\right)\frac{\xi_{\lambda_j}(v)}{\cos^3 v}\,dv,$$

where ξ_λ is the solution of the ordinary differential equation

$$\cos^2 v\,\xi_\lambda''(v) + \sin v\cos v\,\xi_\lambda'(v) + \lambda\xi_\lambda(v) = 0,$$

with the initial conditions

$$\xi_\lambda(0) = 1, \quad \xi_\lambda'(0) = 0.$$

In the following proofs we work with a fixed λ and denote ξ_λ by ξ.

Proof. Unfolding the spectral coefficients,

$$
\begin{aligned}
a(f,t_j) &= \int_{\Gamma\backslash\mathscr{H}} \sum_{\gamma\in H\backslash\Gamma} f\left(\frac{1}{\cos^2 v(\gamma p)}\right) u_j(p)\, d\mu(p) \\
&= \int_{H\backslash\mathscr{H}} f\left(\frac{1}{\cos^2 v}\right) u_j(x,u,v)\frac{dx\,du\,dv}{e^u\cos^3 v}.
\end{aligned}
$$

We can express this in terms of the period integral as

$$
a(f,t_j) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f\left(\frac{1}{\cos^2 v}\right)\varphi_j(v)\frac{dv}{\cos^3 v},
$$

where

$$
\varphi_j(v) = \int_{S_v} u_j(x,u,v)\frac{du\,dx}{e^u}.
$$

It is immediate that φ_j is even. According to Lemma 6.4, in our new coordinates the eigenvalue equation becomes:

$$
e^{2u}\cos^2 v\frac{\partial^2 u_j}{\partial x^2} + \cos^2 v\left(\frac{\partial^2 u_j}{\partial u^2} + \frac{\partial^2 u_j}{\partial v^2}\right) - \cos^2 v\frac{\partial u_j}{\partial u} + \sin v\cos v\frac{\partial u_j}{\partial v} + \lambda u_j = 0.
$$

Now, dividing by e^u and integrating over S_v we get

$$
\begin{aligned}
\cos^2 v\int_{S_v} e^{2u}\frac{\partial^2 u_j}{\partial x^2}\frac{dx\,du}{e^u} + \cos^2 v\int_{S_v}\frac{\partial^2 u_j}{\partial u^2}\frac{du\,dx}{e^u} + \cos^2 v\frac{\partial^2\varphi_j}{\partial v^2} \\
-\cos^2 v\int_{S_v}\frac{\partial}{\partial u}\int u_j\frac{dx\,du}{e^u} + \sin v\cos v\frac{\partial\varphi_j}{\partial v} + \lambda\varphi_j(v) = 0. \qquad (6.16)
\end{aligned}
$$

Next, notice that the Laplacian on S_v in the induced metric is exactly the restriction of Δ to S_v, that is, $\Delta\restriction_{S_v} = \cos^2 v\Delta_{S_v}$, where

$$
\Delta_{S_v} = e^{2u}\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial u^2} - \frac{\partial}{\partial u}.
$$

Hence, for a fixed v, (6.16) becomes

$$
\cos^2 v\int_{S_v}\Delta_{S_v} u_j\frac{dx\,du}{e^u} + \cos^2 v\frac{\partial^2\varphi_j}{\partial v^2} + \sin v\cos v\frac{\partial\varphi_j}{\partial v} + \lambda\varphi_j(v) = 0. \qquad (6.17)
$$

Denote the integral in (6.17) by I_v. Then, by Stokes' theorem we have that

$$
I_v = \int_{\partial S_v}\nabla u_j\cdot\mathbf{n}\,d\ell,
$$

where $d\ell$ is the line element on S_v and $\mathbf{n}$ is the unit normal vector (on the plane) to S_v.

We wish to show that $I_v = 0$. To do this, recall some basic terminology for fundamental domains (in $\mathbb{H}^2$) from Beardon [3]. Let $\mathscr{F}$ be a fundamental domain of a cofinite or cocompact Fuchsian group G. Then $\mathscr{F}$ is a convex hyperbolic polygon with finitely many sides. A *side* of $\mathscr{F}$ is a geodesic segment of the form $\overline{\mathscr{F}} \cap g\overline{\mathscr{F}}$ for any $g \in G$ with $g \neq I$. A *vertex* of $\mathscr{F}$ is a point of the form $\overline{\mathscr{F}} \cap g\overline{\mathscr{F}} \cap h\overline{\mathscr{F}}$ for any $g \neq h \in G$ such that $g, h \neq I$. If $\partial\mathscr{F}$ contains an elliptic fixed point of $g \in G$ of order 2, then we consider the fixed point as a vertex of $\mathscr{F}$ and moreover g identifies the adjacents sides with opposite orientation. In general, we can always find a side-pairing for $\mathscr{F}$, that is, for $i = 1, \ldots, k$, there exist triples (Λ_i, Ψ_i, g_i) such that $g_i\Lambda_i = \Psi_i$, and g_i is the unique element in G that does this, and that Λ_i or Ψ_i are not paired with any other sides of $\mathscr{F}$. Finally, we can always choose $\mathscr{F}$ so that if we consider $\partial\mathscr{F}$ as a contour in $\mathscr{H}$, then the congruent sides occur with opposite orientation as segments of the contour [45, pp. 2–4]. So, let $\{(\Lambda_i, \Psi_i, g_i) : i = 1, \ldots, k\}$ be a side-pairing of S. Then it immediately follows that for any S_v we get a corresponding side-pairing. Denote these by $\{(\Lambda_i^v, \Psi_i^v, g_i) : i = 1, \ldots, k\}$, where $(\Lambda_i^v)_0 = \Lambda_i$ and $(\Psi_i^v)_0 = \Psi_i$. It follows that $I_v = 0$ as the integral over Λ_i^v is cancelled by the one over Ψ_i^v since $\nabla u_j \cdot \mathbf{n}$ is invariant under H.

We are left with

$$\cos^2 v\, \varphi''(v) + \sin v \cos v\, \varphi'(v) + \lambda\varphi(v) = 0, \tag{6.18}$$

where $\varphi'(0) = 0$, as φ is even. Define,

$$\omega(v) = \varphi(v) + \varphi(-v),$$

for $v \in (-\pi/2, \pi/2)$. We then have the relations

$$\omega'(v) = \varphi'(v) - \varphi'(-v), \quad \omega''(v) = \varphi''(v) + \varphi''(v).$$

Hence, adding (6.18) evaluated at $-v$ to itself yields

$$\cos^2 v\, \omega''(v) + \sin v \cos v\, \omega'(v) + \lambda\omega(v) = 0, \tag{6.19}$$

with

$$\omega(0) = 2\hat{u}_j, \quad \omega'(0) = 0.$$

Now, suppose that $\xi(v)$ is a solution to the second order homogenous linear ODE

$$\cos^2 v\, \xi''(v) + \sin v \cos v\, \xi'(v) + \lambda \xi(v) = 0, \tag{6.20}$$

with initial conditions

$$\xi(0) = 1, \quad \xi'(0) = 0.$$

Then we can write the full solution ω of (6.19) as

$$\omega(v) = 2\hat{u}_j \xi(v).$$

Therefore, the $a(f, t_j)$'s can be written as

$$\begin{aligned} a(f, t_j) &= \left(\int_0^{\pi/2} + \int_{-\pi/2}^{0} \right) f\left(\frac{1}{\cos^2 v} \right) \varphi(v) \frac{dv}{\cos^3 v} \\ &= \int_0^{\pi/2} f\left(\frac{1}{\cos^2 v} \right) \omega(v) \frac{dv}{\cos^3 v} \\ &= 2\hat{u}_j \int_0^{\pi/2} f\left(\frac{1}{\cos^2 v} \right) \xi(v) \frac{dv}{\cos^3 v}. \end{aligned}$$

□

We will also need some estimates on ξ_λ. Notice that the following lemma does not use the explicit form of ξ_λ (which we will compute later). This proof is analogous to [49, §4.2].

Lemma 6.8. *For all $v \in [0, \pi/2)$ we have*

$$|\xi_\lambda(v)| \leq 1, \tag{6.21}$$

$$\xi_\lambda(v) \geq 1 - \frac{2 + \lambda}{2} \tan^2 v. \tag{6.22}$$

Proof. Recall that ξ satisfies

$$\cos^2 v \xi''(v) + \sin v \cos v \xi'(v) + \lambda \xi(v) = 0.$$

Multiplying this by $2\xi'(v)$, we can write

$$\cos^2 v \left(\xi'(v)^2 \right)' + 2 \sin v \cos v (\xi'(v))^2 + \lambda (\xi(v)^2)' = 0.$$

Now, integrate over $[0,x]$ and use $\xi(0)=1$ and $\xi'(0)=0$ to get

$$\cos^2 x\xi'(x)^2+2\int_0^x \sin 2v\xi'(v)^2\,dv+\lambda(\xi(x)^2-1)=0.$$

Since $x\in[0,\pi/2)$, $\sin 2v$ is non-negative and so

$$\lambda(1-\xi(x)^2)=\cos^2 x\xi'(x)^2+2\int_0^x \sin 2v\,\xi'(v)^2\,dv\geq 0.$$

This proves the first part. Now we can apply (6.21) to get

$$\begin{aligned}\xi''(v)&=-\tan v\xi'(v)-\lambda\xi(v)\sec^2 v\\&\geq -\tan v\xi'(v)-\lambda\sec^2 v.\end{aligned}$$

Integrating first over $[0,x]$ we get

$$\xi'(x)\geq -\tan x(1+\xi(x))-\lambda\tan x\geq -(2+\lambda)\tan x,$$

and then integrating over $[0,v]$ yields

$$\xi(v)-1\geq(2+\lambda)\log(\cos v)\geq\frac{-1}{2}(2+\lambda)\tan^2 v.$$

Thus,

$$\xi(v)\geq 1-\frac{2+\lambda}{2}\tan^2 v.$$

□

With this, we have the following Hecke type bound for the mean square of the period integrals.

Lemma 6.9. *Let $\hat{u}_j$ be the period integral over $H\backslash\mathscr{P}$ of the automorphic form $u_j\in L^2(M)$. Then, for $T>1$*

$$\sum_{|t_j|\leq T}|\hat{u}_j|^2\ll T.$$

This is a surprising result in the sense that the order of growth is better than what we expect from the local Weyl law, Lemma 6.12. We suspect that the mean square should be bounded in all dimensions, cf. Tsuzuki [102, Theorem 1]. The proof follows a general strategy of applying Parseval's identity to the Fourier expansion and estimating the special functions from below. See for example Theorem 3.2 and Chapter 8 in [53] as well as [49, §2.6].

Proof. Let $K = \sup_{p\in\mathcal{H}} N(p,X)$. This is well-defined as $N(p,X)$ is uniformly bounded. Then

$$\int_{\Gamma\backslash\mathcal{H}} (A(f)(p))^2\, d\mu(p) \le K \int_{\Gamma\backslash\mathcal{H}} A(f)(p)\, d\mu(p).$$

Also, define

$$\tan\theta = \sqrt{\frac{2}{y+2}}.$$

Then

$$a(f,t_j) = 2\hat{u}_j \int_0^\theta \frac{\xi_{\lambda_j}(v)}{\cos^3 v}\, dv$$

with $\xi_0(v) = 1$. In particular, the coefficient for the zero eigenvalues (so $t_j = i$) gives

$$\begin{aligned}
a(f,t_0) &= 2\hat{u}_0 \int_0^\theta \cos^{-3} v\, dv \\
&= 2\hat{u}_0 \left(\frac{1}{2}\sec\theta\tan\theta + \frac{1}{2}\log|\sec\theta + \tan\theta| \right) \\
&= 2\hat{u}_0 \left(\frac{1}{\sqrt{2}} \frac{\sqrt{4+y}}{y+2} + \frac{1}{2}\log\left| \frac{\sqrt{2} + \sqrt{4+y}}{\sqrt{y+2}} \right| \right) \\
&= 2\hat{u}_0 g(y),
\end{aligned}$$

say. On the other hand,

$$a(f,t_0) = 2u_0 \int_{\Gamma\backslash\mathcal{H}} A(f)(p)\, d\mu(p).$$

Hence,

$$\hat{u}_0 g(y) = u_0 \int_{\Gamma\backslash\mathcal{H}} A(f)(p)\, d\mu(p).$$

It follows that

$$\int_{\Gamma\backslash\mathcal{H}} (A(f)(p))^2\, d\mu(p) \le K \frac{\hat{u}_0 g(y)}{u_0}, \tag{6.23}$$

where u_0 is the constant eigenfunction. By Parseval we have

$$\begin{aligned}
\int_{\Gamma\backslash\mathcal{H}} (A(f)(p))^2\, d\mu(p) &= \sum_{j=0}^\infty |a(f,t_j)|^2 \\
&\ge \sum_{\lambda_j \le y} |a(f,t_j)|^2,
\end{aligned}$$

so that

$$\int_{\Gamma\backslash\mathcal{H}} (A(f)(p))^2\, d\mu(p) \ge \sum_{\lambda_j \le y} |\hat{u}_j|^2 \left(\int_0^\theta \frac{\xi_{\lambda_j}(v)}{\cos^3 v}\, dv \right)^2. \tag{6.24}$$

Since $\lambda_j \le y$, from the bound (6.22) it follows that:

$$\xi_{\lambda_j} \ge 1 - \frac{2+\lambda_j}{2}\tan^2 v \ge 1 - \frac{2+y}{2}\tan^2 v.$$

Hence,

$$\begin{aligned}\int_0^\theta \frac{\xi_{\lambda_j}(v)}{\cos^3 v}\,dv &\ge \int_0^\theta \left(1-\frac{2+y}{2}\tan^2 v\right)\frac{dv}{\cos^3 v}\\ &= \frac{1}{8\tan^2\theta}\left(\tan\theta\sec\theta(2\sec^2\theta-3)+(1+4\tan^2\theta)\log\left(\frac{1+\tan\frac{\theta}{2}}{1-\tan\frac{\theta}{2}}\right)\right)\\ &= h(\theta).\end{aligned}$$

We are interested in the behaviour of h as $y \to \infty$, that is, $\theta \to 0^+$. After a tedious but elementary computation we find that

$$\lim_{\theta\to 0^+}\frac{h(\theta)}{\tan\theta} = \frac{2}{3}.$$

This means that $h(\theta) \ge (\frac{2}{3}-\epsilon)\tan\theta$ for small enough θ and for some $\epsilon > 0$. In other words, we have proved that

$$\int_0^\theta \frac{\xi_{\lambda_j}(v)}{\cos^3 v}\,dv \ge c y^{-1/2}, \tag{6.25}$$

for some constant $c > 0$, as $y \to \infty$. Now, combining (6.23), (6.24) and (6.25) we get

$$\sum_{\lambda_j \le y} |\hat{u}_j|^2 \le \frac{K\hat{u}_0}{c^2 u_0} g(y) y \ll y^{1/2}.$$

The result follows from observing that $\lambda_j = 1 + t_j^2$. □

As pointed out earlier, we can actually express ξ in an elementary form. We suspect that this is always possible in odd dimensional hyperbolic space. In even dimensions the special functions are more complicated Legendre or hypergeometric functions. We let $r = \tan v$. Then

$$a(f,t_j) = 2\hat{u}_j \int_0^\infty f(1+r^2)\xi(\arctan r)\sqrt{1+r^2}\,dr. \tag{6.26}$$

Apply the transformation $\tan v = \sinh w$ in (6.20). It becomes

$$\xi''(w) + 2\tanh w\,\xi'(w) + \lambda\xi(w) = 0.$$

It is easy to check that the general solution is

$$\xi(w) = \frac{1}{\cosh w}\left(Ae^{-w\sqrt{1-\lambda}} + Be^{w\sqrt{1-\lambda}}\right),$$

for some constants A and B. With our initial conditions we get

$$A = B = \frac{1}{2}.$$

Since $\lambda = s_j(2-s_j)$, we have $1-\lambda = (s_j-1)^2$. So,

$$\begin{aligned}\xi(w) &= \frac{1}{2\cosh w}\left(e^{-w(s_j-1)} + e^{w(s_j-1)}\right)\\ &= \frac{\cosh w(s_j-1)}{\cosh w},\end{aligned} \tag{6.27}$$

or in terms of r:

$$\xi(r) = \frac{\cosh((s_j-1)\operatorname{arcsinh} r)}{\sqrt{1+r^2}}. \tag{6.28}$$

Thus we can finally write the explicit form for ξ in terms of v as

$$\xi(v) = \frac{\cosh((s_j-1)\operatorname{arcsinh}\tan v)}{\sec v}. \tag{6.29}$$

We will now show how to estimate the spectral coefficients $a(f, t_j)$. With the explicit form (6.28) for ξ, we can write

$$a(f, t_j) = 2\hat{u}_j \int_0^\infty f(1+r^2)\cosh((s_j-1)\operatorname{arcsinh} r)\,dr. \tag{6.30}$$

We are thus led to consider the integral transform

$$c(f, t) = \int_0^\infty f(1+r^2)\cosh((s-1)\operatorname{arcsinh} r)\,dr,$$

where $s = 1+it$. Now, define

$$f\left(\frac{1}{\cos^2 v(p)}\right) = \begin{cases} 1, & \text{if } 0 \le |v| \le \Theta,\\ 0, & \text{if } \Theta < |v| < \frac{\pi}{2},\end{cases}$$

or equivalently

$$f\left(\frac{1}{\cos^2 v(p)}\right) = \begin{cases} 1, & \text{if } 1 \le \sec v \le X,\\ 0, & \text{if } X < \sec v.\end{cases}$$

If we let $r = \tan v$, then we get that

$$f\left(1+r^2\right) = \begin{cases} 1, & \text{if } 0 \le r \le U, \\ 0, & \text{if } U < r, \end{cases}$$

where $U = \tan\Theta = \sqrt{X^2 - 1}$. Notice that

$$U = X\sqrt{1 - X^{-2}} = X\left(1 + O(X^{-2})\right) = X + O(X^{-1}).$$

In particular,

$$A(f)(p) = \widetilde{N}(p, \Theta) = N(p, X).$$

Now, letting $r = \sinh u$, we can rewrite $c(f,t)$ as

$$c(f,t) = \int_0^\infty f(\cosh^2 u)\cosh((s-1)u)\cosh u\, du.$$

Notice that $2\cosh((s-1)u)\cosh u = \cosh su + \cosh(2-s)u$, so that

$$c(f,t) = \frac{1}{2}\int_0^\infty f(\cosh^2 u)\cosh su\, du + \frac{1}{2}\int_0^\infty f(\cosh^2 u)\cosh(2-s)u\, du.$$

Since the integrands are even we arrive at

$$c(f,t) = \frac{1}{4}\int_{\mathbb{R}} f(\cosh^2 u)\cosh su\, du + \frac{1}{4}\int_{\mathbb{R}} f(\cosh^2 u)\cosh(2-s)u\, du,$$

where

$$f(\cosh^2 u) = \begin{cases} 1, & \text{if } |u| \le \operatorname{arcsinh} U, \\ 0, & \text{otherwise.} \end{cases}$$

Since both of the integrals in $c(f,t)$ are of the same type, we define the integral transform $d(f,s)$ given by

$$d(f,s) = \int_{\mathbb{R}} f(\cosh^2 u)\cosh su\, du. \tag{6.31}$$

We list some simple properties of the $d(f,s)$-transform without proof.

Lemma 6.10. *Suppose f and g are compactly supported even functions with finitely many discontinuities, let $\alpha \in \mathbb{R}$, then*

$$\begin{aligned} 4c(f,t) &= d(f,s) + d(f,2-s), \\ d(\alpha f,s) &= \alpha d(f,s), \\ d(f*g,s) &= d(f,s)d(g,s), \end{aligned}$$

where $*$ is the usual convolution. Also

$$d(\mathbb{1}_{[-T,T]},s) = \frac{2\sinh sT}{s},$$

where $\mathbb{1}_{[-T,T]}(\cosh^2 u)$ is the indicator function on $[-T,T]$, and

$$d(f,0) = \int_{\mathbb{R}} f(\cosh^2 u)\,du.$$

Let $1 > \delta > 0$, and define $\chi(\cosh^2 u) = (2\delta)^{-1}\mathbb{1}_{[-\delta,\delta]}(\cosh^2 u)$ to be a characteristic function with unit mass with respect to the $d(f,s)$-transform. Now define

$$\widetilde{f}^{+}(\cosh^2 u) = \begin{cases} 1, & \text{if } |u| \leq \operatorname{arcsinh} U + 2\delta, \\ 0, & \text{otherwise.} \end{cases}$$

$$\widetilde{f}^{-}(\cosh^2 u) = \begin{cases} 1, & \text{if } |u| \leq \operatorname{arcsinh} U - 2\delta, \\ 0, & \text{otherwise.} \end{cases}$$

Let $f^+ = \widetilde{f}^+ * \chi * \chi$ and $f^- = \widetilde{f}^- * \chi * \chi$. Then $f^+(x) = 1$ for $|x| \leq \operatorname{arcsinh} U$ and vanishes for $|x| \geq \operatorname{arcsinh} U + 4\delta$, and similarly $f^-(x)$ vanishes for $|x| \geq \operatorname{arcsinh} U$. It follows that

$$A(f^-)(p) \leq N(p,X) \leq A(f^+)(p).$$

Hence, in order to estimate $N(p,X)$ we need bounds for $A(f^+)$ and $A(f^-)$, which in turn leads us to investigate $c(f^{\pm},t)$. The case for f^- is analogous, so we restrict the treatment below to f^+.

Remark 6.7. Without any smoothing, the spectral expansion for $A(f)$ would of course not converge. In two dimensions it suffices to use a single convolution (linear decay). In our case we need at least two convolutions to ensure convergence. On the other hand, any more smoothing in this manner does not yield improvements for the pointwise bound nor for the application of the large sieve.

Proposition 6.11. *The integral transform $c(f,t)$ satisfies the following properties:*

(i) For $s=1+it$ we can write

$$c(f^+,t)=a(t,\delta)X^{1+it}+b(t,\delta)X^{1-it}, \tag{6.32}$$

where a and b satisfy

$$a(t,\delta),\ b(t,\delta)\ll \min(|t|^{-1},|t|^{-3}\delta^{-2}).$$

(ii) For $s\in[1,2]$ we have

$$c(f^+,t)=\frac{2^{s-2}}{s}X^s+\frac{2^{-s}}{2-s}X^{2-s}+O(\delta X^s),$$

where the case of $s=2$ is understood as

$$c(f^+,i)=\frac{X^2}{2}+O(\delta X^2).$$

Proof. We have

$$\begin{aligned} d(f^+,s)&=\frac{8\sinh s(\operatorname{arcsinh} U+2\delta)\sinh^2 s\delta}{(2\delta)^2 s^3}\\ &=\frac{8\sinh^2 s\delta}{(2\delta)^2 s^3}\sinh(s\log(U+\sqrt{U^2+1})+2s\delta). \end{aligned}$$

By Taylor expansion $U+\sqrt{U^2+1}=2U+O(U^{-1})$, so that

$$(U+\sqrt{U^2+1})^s=(2U)^s+O(sU^{s-2})=(2X)^s+O(sX^{s-2}).$$

Now suppose $s\in[1,2)$, then we may assume that $|s|\delta<1$. So,

$$\begin{aligned} \sinh(s\operatorname{arcsinh} U)&=\frac{1}{2}\big((2X)^s+O(sX^{s-2})\big),\\ \cosh(2s\delta)&=1+O(\delta). \end{aligned}$$

Therefore,

$$\begin{aligned}\sinh s(\operatorname{arcsinh} U+2\delta) &= \sinh(s\operatorname{arcsinh} U)\cosh 2s\delta+\cosh(s\operatorname{arcsinh} U)\sinh 2s\delta\\ &= \frac{1}{2}\big((2X)^s+O(X^{s-2})\big)(1+O(\delta))+O(\delta X^s+\delta X^{s-2})\\ &= \frac{1}{2}(2X)^s+O(\delta X^s).\end{aligned}$$

It follows that

$$d(f^+,s)=\frac{4\sinh^2 s\delta}{(2\delta)^2 s^3}\left((2X)^s+O(\delta X^s)\right).$$

Since $|s|\delta<1$, we also have that $(\sinh s\delta)/s\delta=1+O(\delta)$, and

$$d(f^+,s)=\frac{1}{s}(1+O(\delta))((2X)^s+O(\delta X^s))=\frac{2^s}{s}X^s+O(\delta X^s).$$

So

$$\begin{aligned}c(f^+,t)&=\frac{1}{4}\left(d(f^+,s)+d(f^+,2-s)\right)\\ &=\frac{2^{s-2}}{s}X^s+\frac{2^{-s}}{2-s}X^{2-s}+O(\delta X^s).\end{aligned}$$

Now, for the smallest eigenvalue, $s=2$, we get

$$c(f^+,i)=\frac{1}{4}(d(f^+,2)+d(f^+,0)),$$

where

$$d(f^+,2)=2X^2+O(\delta X^2),$$

and

$$d(f^+,0)=O(\log X),$$

as $d(\chi,0)=1$. This proves (ii) in the proposition. We now consider the case when s is complex, that is, $s=1+it$. Assume $t>0$ and $X>1$, to get

$$\begin{aligned}\sinh((1+it)(\operatorname{arcsinh} U+2\delta)) &= \sinh((1+it)\operatorname{arcsinh} U)\cosh((1+it)2\delta)\\ &\quad+\cosh((1+it)\operatorname{arcsinh} U)\sinh((1+it)2\delta)\\ &=\frac{1}{2}\Big((2X+O(X^{-1}))^{1+it}e^{2\delta(1+it)}\\ &\quad-(2X+O(X^{-1}))^{-1-it}e^{-2\delta(1+it)}\Big).\end{aligned}$$

Thus we can write

$$\sinh((1+it)(\operatorname{arcsinh} U+2\delta))=X^{1+it}\upsilon(t,\delta), \tag{6.33}$$

where $\upsilon(t,\delta)$ is bounded for $0<\delta<1$ and any t. Hence,

$$d(f^+,1+it)=X^{1+it}\upsilon(t,\delta)\frac{2}{s}\left(\frac{\sinh s\delta}{s\delta}\right)^2.$$

Now, suppose $|s|\delta<1$, then

$$\frac{\sinh s\delta}{s\delta}\ll 1.$$

So in this case

$$d(f^+,1+it)=X^{1+it}|t|^{-1}.$$

On the other hand, if $|s|\delta\geq 1$, then $\sinh s\delta=O(1)$ so that

$$d(f^+,1+it)=X^{1+it}|t|^{-3}\delta^{-2}.$$

Working similarly with $d(f^+,1-it)$ proves (i). □

Before we can prove the theorem, we need to know the local Weyl law in our setting. Elstrodt, Grunewald, and Mennicke [25] prove this for Eisenstein series in Chapter 6 Theorem 4.10. It is clear that their proof can be extended to include the cuspidal part. This yields the following lemma.

Lemma 6.12. *For $T>1$, we have for all $p\in\mathcal{H}$ that*

$$\sum_{t_j\leq T}|u_j(p)|^2\ll y(p)^2T+T^3.$$

We now have all the ingredients to prove our main theorem.

Proof of Theorem 6.1. First, write the spectral expansion of $A(f^+)(p)$:

$$\begin{aligned}A(f^+)(p)&=\sum_j 2c(f^+,t_j)\hat{u}_j u_j(p)\\&=X^2\hat{u}_0u_0+\sum_{s_j\in[1,2)}2\hat{u}_ju_j(p)\left(\frac{2^{s_j-2}X^{s_j}}{s_j}+\frac{2^{-s_j}X^{2-s_j}}{2-s_j}+O(\delta X^{s_j})\right)\\&\quad+\sum_{t_j\in\mathbb{R}}2c(f^+,t_j)\hat{u}_ju_j(p).\end{aligned}$$

Now the summation over $s_j \in [1,2)$ is finite, so

$$\sum_{s_j \in [1,2)} 2\hat{u}_j u_j(p) O(\delta X^s) = O(\delta X^2).$$

Also, by the discreteness of the spectrum, for a fixed Γ there is some $\epsilon_0 > 0$ such that there are no s_j in $[2-\epsilon_0, 2)$, and so

$$\sum_{s_j \in [1,2)} 2\hat{u}_j u_j(p) \frac{2^{-s_j} X^{2-s_j}}{2-s_j} = O(X).$$

Hence the spectral expansion becomes

$$A(f^+)(p) = \sum_{s_j \in (1,2]} \frac{2^{s_j-1} X^{s_j}}{s_j} \hat{u}_j u_j(p) + G(f^+, p) + O(X + \delta X^2),$$

where

$$G(f^+, p) = \sum_{0 \neq t_j} 2c(f^+, t_j) \hat{u}_j u_j(p).$$

Again, by the discreteness of the spectrum we can estimate the contribution of small t_j's

$$G(f^+, p) = \sum_{|t_j| \geq 1} 2c(f^+, t_j) \hat{u}_j u_j(p) + O(X).$$

Now, since $c(f^+, t)$ is even in t, we get by a dyadic decomposition

$$\begin{aligned}
\sum_{|t_j| \geq 1} 2c(f^+, t_j) \hat{u}_j u_j(p) &\ll \sum_{t_j \geq 1} c(f^+, t_j) \hat{u}_j u_j(p) \\
&= \sum_{n=0}^{\infty} \left(\sum_{2^n \leq t_j < 2^{n+1}} c(f^+, t_j) \hat{u}_j u_j(p) \right) \\
&\ll \sum_{n=0}^{\infty} \sup_{2^n \leq t_j < 2^{n+1}} c(f^+, t_j) \left(\sum_{2^n \leq t_j < 2^{n+1}} \hat{u}_j u_j(p) \right).
\end{aligned}$$

By the Cauchy–Schwarz inequality and Lemmas 6.9 and 6.12 we have

$$\begin{aligned}
G(f^+, p) &\ll \sum_{n=0}^{\infty} \sup_{2^n \leq t_j < 2^{n+1}} c(f^+, t_j) \left(\sum_{t_j < 2^{n+1}} |\hat{u}_j|^2 \right)^{1/2} \left(\sum_{t_j < 2^{n+1}} |u_j(p)|^2 \right)^{1/2} + X \\
&\ll \sum_{n=0}^{\infty} \sup_{2^n \leq t_j < 2^{n+1}} c(f^+, t_j) 2^{2n+2} + X.
\end{aligned}$$

We separate the sum over n depending on whether $t_j\delta \leq 1$ or $t_j\delta \geq 1$,

$$G(f^+,p) \ll \sum_{n<\log_2\delta^{-1}} 2^{2n+2} \sup_{2^n\leq t_j<2^{n+1}} c(f^+,t_j) + \sum_{n>\log_2\delta^{-1}} 2^{2n+2} \sup_{2^n\leq t_j<2^{n+1}} c(f^+,t_j) + X.$$

Hence, by Proposition 6.11,

$$G(f^+,p) \ll \sum_{n<\log_2\delta^{-1}} 2^{2n+2} X 2^{-n} + \sum_{n>\log_2\delta^{-1}} 2^{2n+2} X\delta^{-2}2^{-3n} + X \ll X\delta^{-1} + X.$$

Putting all this together we find that

$$A(f^+)(p) = \sum_{s_j\in(1,2]} \frac{2^{s_j-1}X^{s_j}}{s_j}\hat{u}_j u_j(p) + O(X + \delta X^2 + \delta^{-1}X). \tag{6.34}$$

The optimal choice for δ comes from equating $\delta X^2 = \delta^{-1}X$, which gives $\delta = X^{-1/2}$. The result follows from noting that $u_0 = \mathrm{vol}(\Gamma\backslash\mathscr{H})^{-1/2}$ and $\hat{u}_0 = \mathrm{vol}(S)\,\mathrm{vol}(\Gamma\backslash\mathscr{H})^{-1/2}$.

□

6.3 Applications of the Large Sieve

We will now apply Chamizo's large sieve inequalities to show that the mean square of the error term $E(p,X)$ satifies the conjectured bound $O(X^{1+\epsilon})$ over a spatial average. In the radial aspect Chamizo proves large sieve inequalities with exponential weights for all moments in two dimensions. We extend his result to three dimensions for the second moment. For structural reasons that are explained later, this inequality has a limited application to our problem. We can only prove a mean square estimate of $O(X^{2+2/3})$ in the radial average. This translates to an improvement of $1/6$ compared to the pointwise bound we obtained in Section 6.2. More specifically, our aim is to prove the following two theorems.

Theorem 6.13. *Let $X>1$ and $X_1,\dots,X_R\in[X,2X]$ such that $|X_k-X_l|>\epsilon>0$ for all $k\neq l$. Suppose $R\epsilon\gg X$ and $R>X^{2/3}$, then*

$$\frac{1}{R}\sum_{k=1}^{R}|E(p,X_k)|^2\ll X^{2+2/3}\log X. \tag{6.35}$$

Taking the limit $R\to\infty$ gives

$$\frac{1}{X}\int_X^{2X}|E(p,x)|^2\,dx\ll X^{2+2/3}\log X. \tag{6.36}$$

For $p,q\in\Gamma\backslash\mathscr{H}$, let

$$\tilde{d}(p,q)=\inf_{\gamma\in\Gamma}d(p,\gamma q)$$

be the induced distance on $\Gamma\backslash\mathscr{H}$.

Theorem 6.14. *Let $X>1$ and $p_1,\dots,p_R\in\Gamma\backslash\mathscr{H}$ with $\tilde{d}(p_k,p_l)>\epsilon>0$ for all $k\neq l$. Suppose $R\epsilon^3\gg 1$ and $R>X$, then*

$$\frac{1}{R}\sum_{k=1}^{R}|E(p_k,X)|^2\ll X^2\log^2 X. \tag{6.37}$$

Taking the limit as $R\to\infty$ gives

$$\int_{\Gamma\backslash\mathscr{H}}|E(p,X)|^2\,d\mu(p)\ll X^2\log^2 X. \tag{6.38}$$

We split the rest of this section into four parts: one for each of the averages; we then write down the proof for a generalisation of the radial large sieve inequality that is used for Theorem 6.13; finally, we explain why the expected results from the application of the large sieve get worse for higher dimensions.

Remark 6.8. The methods in Sections 6.3.1 and 6.3.2 generalise without effort for cofinite groups Γ. However, their application to the error term would depend on being able to identify the oscillations for the terms $t_j\in\mathbb{R}$ in the spectral expansion of $A(f)(p)$ (see equation (6.32)).

6.3.1 Radial Average

We will prove the following proposition.

Proposition 6.15. *Let $X>1$ and $X_1,\dots,X_R\in[X,2X]$ such that $|X_k-X_l|\geq\epsilon>0$ for all $k\neq l$. Then, we have*

$$\sum_{k=1}^{R}|E(p,X_k)|^2\ll X^{3+2/3}\epsilon^{-1}\log X+X^{2+2/3}R+X^{3+1/3}\log X. \tag{6.39}$$

Theorem 6.13 follows immediately from the above proposition.

Proof of Theorem 6.13. We take $\epsilon\asymp R^{-1}X$. Hence the bound (6.39) becomes

$$\sum_{k=1}^{R}|E(p,X_k)|^2\ll X^{2+2/3}R\log X+X^{2+2/3}R+X^{3+1/3}\log X.$$

So if we choose $R>X^{2/3}$ then

$$\frac{1}{R}\sum_{k=1}^{R}|E(p,X_k)|^2\ll X^{2+2/3}\log X.$$

This proves (6.35). For the integral limit (6.36) it suffices to consider a limiting partition of $[X,2X]$ with equally spaced points. □

The large sieve inequality for radial averaging is given by the following theorem.

Theorem 6.16. *Given $p\in\Gamma\backslash\mathcal{H}$, suppose that $X>1$ and $T>1$. Let $x_1,\dots,x_R\in[X,2X]$. If $|x_k-x_l|>\epsilon>0$ for all $k\neq l$, then*

$$\sum_{k=1}^{R}\left|\sum_{|t_j|\leq T}a_jx_k^{it_j}u_j(p)\right|^2\ll(T^3+XT^2\epsilon^{-1})||a||_*^2, \tag{6.40}$$

where

$$||a||_*^2=\sum_{|t_j|\leq T}|a_j|^2. \tag{6.41}$$

In two dimensions, Chamizo [11, Theorem 2.2] proves a corresponding result to the above theorem. The proof in three dimensions is similar, but we write it down in Section 6.3.3. Let f be a compactly supported function with finitely many discontinuities on $[1,\infty)$, and denote

$$E_f(p,X)=A(f)(p)-\sum_{1\leq s_j\leq 2}c(f,t_j)\hat{u}_ju_j(p).$$

Then, recall we have shown that (see (6.34))

$$
\begin{aligned}
E_{f^+}(p,X) &= O(X\delta^{-1}+X),\\
E_{f^-}(p,X) &= O(X\delta^{-1}+X),
\end{aligned}
$$

and

$$
E_{f^-}(p,X) < E(p,X) + O(X^2\delta + X) < E_{f^+}(p,X). \tag{6.42}
$$

We can now prove the proposition.

Proof of Proposition 6.15. For simplicity, we combine the error terms in (6.42). Suppose that $1 > \delta \gg X^{-1}$, then

$$
E_{f^-}(p,X) < E(p,X) + O(X^2\delta) < E_{f^+}(p,X).
$$

Hence,

$$
\sum_{k=1}^{R} |E(p,X_k)|^2 \ll \sum_{k=1}^{R} |E_f(p,X_k)|^2 + RX^4\delta^2,
$$

where f is appropriately chosen as f^+ or f^- depending on k. The main strategy is again to apply dyadic decomposition in the spectral expansion. We use the following notation for the truncated spectral expansion:

$$
S_T(p,X) = \sum_{T<|t_j|\le 2T} 2c(f,t_j)\hat{u}_j u_j(p). \tag{6.43}
$$

We consider three different ranges, which we choose so that the tails of the spectral expansion get absorbed into the error term. The correct ranges are given by

$$
\begin{aligned}
A_1 &= \{t_j : 0 < |t_j| \le 1\},\\
A_2 &= \{t_j : 1 < |t_j| \le \delta^{-3}\},\\
A_3 &= \{t_j : |t_j| > \delta^{-3}\}.
\end{aligned}
$$

Also, define

$$
S_i = \sum_{t_j \in A_i} 2c(f,t_j)\hat{u}_j u_j(p).
$$

We can now write

$$
E_f(p,X) = S_1 + S_2 + S_3.
$$

For the tail we have

$$\begin{aligned}\sum_{t_j\in A_3} 2c(f,t_j)\hat{u}_j u_j(p) &\ll \sum_{|t_j|\geq \delta^{-3}} c(f,t_j)\hat{u}_j u_j(p)\\ &\ll \sum_{t_j>\delta^{-3}} \min(|t|^{-1},|t|^{-3}\delta^{-2})X\hat{u}_j u_j(p).\end{aligned}$$

With a dyadic decomposition we get

$$\begin{aligned}\sum_{t_j\in A_3} 2c(f,t_j)\hat{u}_j u_j(p) &\ll X\delta^{-2}\sum_{n=0}^{\infty}\left(\sum_{2^n\delta^{-3}<t_j\leq 2^{n+1}\delta^{-3}} t^{-3}\hat{u}_j u_j(p)\right)\\ &\ll X\delta^{-2}\sum_{n=0}^{\infty}\delta^9 2^{-3n}\left(\sum_{2^n\delta^{-3}<t_j\leq 2^{n+1}\delta^{-3}} \hat{u}_j u_j(p)\right),\end{aligned}$$

and then from the Cauchy–Schwarz inequality it follows that

$$\begin{aligned}&\ll X\delta^7\sum_{n=0}^{\infty}2^{-3n}\left(\sum_{t_j\leq\delta^{-3}2^{n+1}}|\hat{u}_j|^2\right)^{1/2}\left(\sum_{t_j\leq\delta^{-3}2^{n+1}}|u_j(p)|^2\right)^{1/2}\\ &\ll X\delta^7\sum_{n=0}^{\infty}2^{-3n}(2^{n/2}\delta^{-3/2})(2^{3n/2}\delta^{-9/2})\\ &\ll X\delta,\end{aligned}$$

as required. Next, for the first interval we have

$$S_1=\sum_{t_j\in A_1}2c(f,t)\hat{u}_j u_j(p)\ll X\sum_{|t_j|<1}|t|^{-1}\hat{u}_j u_j(p)\ll X.$$

Hence

$$S_1+S_3=O(X^2\delta).$$

Finally, we split the summation in S_2 into dyadic intervals by letting $T=2^n$ for $n=0,1,\ldots,[\log_2\delta^{-3}]$. We then have

$$E_f(p,X)\ll\sum_{1\leq 2^n\leq\delta^{-3}}S_{2^n}(p,X)+O(X^2\delta).$$

Squaring and summing up over the radii gives

$$\sum_{k=1}^{R}|E_f(p,X_k)|^2\ll\sum_{k=1}^{R}\left|\sum_{1\leq 2^n\leq\delta^{-3}}S_{2^n}(p,X_k)\right|^2+O(RX^4\delta^2). \tag{6.44}$$

Applying the Cauchy–Schwarz inequality to the sum over the dyadic intervals yields

$$\left|\sum_{1\le 2^n<\delta^{-3}} S_{2^n}(p,X_k)\right|^2 \ll \log X \sum_{1\le 2^n<\delta^{-3}} |S_{2^n}(p,X_k)|^2.$$

Substituting this back into (6.44) we get

$$\sum_{k=1}^{R}|E_f(p,X_k)|^2 \ll \log X \sum_{1\le 2^n\le\delta^{-3}}\sum_{k=1}^{R}|S_{2^n}(p,X_k)|^2 + O(RX^4\delta^2). \tag{6.45}$$

Recall from Proposition 6.11 that we can write

$$c(f,t) = X(a(t,\delta)X^{it} + b(t,\delta)X^{-it}),$$

where a and b satisfy

$$a(t,\delta),\ b(t,\delta) = \min(|t|^{-1}, |t|^{-3}\delta^{-2}).$$

Keeping in mind our notation with $T = 2^n$ and (6.43), we apply Theorem 6.16 to get

$$\sum_{k=1}^{R}|S_T(p,X_k)|^2 \ll (T^3 + XT^2\epsilon^{-1})||a||_*^2, \tag{6.46}$$

where

$$\begin{aligned}
||a||_*^2 &\ll \sum_{T<|t_j|\le 2T} |\min(|t_j|^{-1}, |t_j|^{-3}\delta^{-2})X\hat{u}_j|^2 \\
&\ll X^2\min(T^{-2}, T^{-6}\delta^{-4})\left(\sum_{T\le|t_j|<2T}|\hat{u}_j|^2\right) \\
&\ll X^2\min(T^{-1}, T^{-5}\delta^{-4}).
\end{aligned}$$

This simplifies (6.46) to

$$\sum_{k=1}^{R}|S_T(p,X_k)|^2 \ll (T^3 + XT^2\epsilon^{-1})X^2\min(T^{-1}, T^{-5}\delta^{-4}).$$

Therefore (6.45) becomes

$$\begin{aligned}
\sum_{k=1}^{R}|E(p,X_k)|^2 &\ll \log X \sum_{1\le T<\delta^{-3}}\sum_{k=1}^{R}|S_T(p,X_k)|^2 + RX^4\delta^2 \\
&\ll \log X \sum_{1\le T<\delta^{-3}} (T^3 + XT^2\epsilon^{-1})X^2\min(T^{-1}, T^{-5}\delta^{-4}) + RX^4\delta^2.
\end{aligned}$$

We split the summation depending on whether $T<\delta^{-1}$ and get

$$\sum_{k=1}^{R}|E(p,X_i)|^2 \ll X^2\log X\left(\sum_{1\le T\le\delta^{-1}}T^2\right)+X^3\epsilon^{-1}\log X\left(\sum_{1\le T\le\delta^{-1}}T\right)$$
$$+X^2\delta^{-4}\log X\left(\sum_{\delta^{-1}\le T\le\delta^{-3}}T^{-2}\right)$$
$$+X^3\delta^{-4}\epsilon^{-1}\log X\left(\sum_{\delta^{-1}\le T\le\delta^{-3}}T^{-3}\right)+RX^4\delta^2.$$

With trivial estimates we have

$$\sum_{k=1}^{R}|E(p,X_k)|^2 \ll X^2\delta^{-2}\log X+X^3\epsilon^{-1}\delta^{-1}\log X+RX^4\delta^2.$$

The optimal choice for δ comes from $X^4\delta^2=X^2\delta^{-1}$, that is, $\delta=X^{-2/3}$, since $\epsilon R\asymp X$. This gives

$$\sum_{k=1}^{R}|E(p,X_k)|^2 \ll X^{3+1/3}\log X+X^{3+2/3}\epsilon^{-1}\log X+RX^{2+2/3}.$$

□

6.3.2 Spatial Average

We now consider the spatial average. In this case the corresponding large sieve inequality was already proved by Chamizo [11, Theorem 3.2] for n dimensions for any cocompact group. It would not be difficult to extend it to cofinite groups in three dimensions. We state it as Theorem 6.18 simplified to our setting. With similar strategy as in Section 6.3.1, we prove the following proposition which readily yields Theorem 6.14.

Proposition 6.17. *Suppose $X>1$ and let $p_1,p_2,\dots,p_R\in\Gamma\backslash\mathcal{H}$ with $\tilde{d}(p_k,p_l)>\epsilon$ for some $\epsilon>0$. Then we have*

$$\sum_{k=1}^{R}|E(p_k,X)|^2 \ll X^4R^{-1}\log^2 X+X^2\epsilon^{-3}.$$

As before, Theorem 6.14 follows easily.

Proof of Theorem 6.14. We pick $\epsilon^{-3} \ll R$ and $R > X$. Then

$$\frac{1}{R}\sum_{k=1}^{R}|E(p_k,X)|^2 \ll X^4R^{-2}\log^2 X + X^2 \ll X^2\log^2 X.$$

For the integral limit we take hyperbolic balls of radius $\epsilon/2$ uniformly spaced in M. For small radii the volume of such a ball is $(4/3)\pi(\epsilon/2)^3$, [25, pg. 10]. This is compatible with our assumption that $R\epsilon^3$ is bounded from below since M is of finite volume. □

For the proof of Proposition 6.17 we use the large sieve inequality in the following form.

Theorem 6.18 (Theorem 3.2 in [11]). *Given $T > 1$, $p_1,\ldots,p_R \in \Gamma\backslash\mathscr{H}$, if $\tilde{d}(p_k,p_l) > \epsilon > 0$ for all $k \neq l$, then*

$$\sum_{k=1}^{R}\Big|\sum_{|t_j|\leq T} a_j u_j(p_k)\Big|^2 \ll (T^3+\epsilon^{-3})\|a\|_*^2,$$

where $\|a\|_*$ is as in (6.41).

We can then prove Proposition 6.17.

Proof of Proposition 6.17. By a direct application of the Cauchy–Schwarz inequality we have

$$\left(\sum_{k=1}^{n} a_k\right)^2 \ll \sum_{k=1}^{n} k^2 a_k^2.$$

Thus,

$$\Big|\sum_{1\leq T\leq\delta^{-3}} S_T(p_k,X)\Big|^2 \ll \sum_{1\leq T\leq\delta^{-3}} |\log T|^2 |S_T(p_k,X)|^2. \tag{6.47}$$

Repeating this with the identity

$$\left(\sum_{k=1}^{n} a_k\right)^2 \ll \sum_{k=1}^{n}(n+1-k)^2 a_k^2,$$

and combining with (6.47) yields

$$\Big|\sum_{1\leq T\leq\delta^{-3}} S_T(p_k,X)\Big|^2 \ll \sum_{1\leq T\leq\delta^{-3}} |c_T|^2 |S_T(p_k,X)|^2,$$

where $c_T = \min\big(|\log T\delta^3 + 1|, \log T\big)$. Repeating the analysis from the proof of Proposition 6.15 and applying Theorem 6.18 gives

$$\sum_{k=1}^{R} |E(p_k,X)|^2 \ll \sum_{1\le T\le \delta^{-3}} |c_T|^2 (T^3+\epsilon^{-3}) X^2 T^{-1} \min(1, T^{-4}\delta^{-4}) + RX^4\delta^2$$
$$\ll X^2\delta^{-2}\log^2 X + X^2\epsilon^{-3} + RX^4\delta^2,$$

as before. We optimise by setting $X^2\delta^{-2} = RX^4\delta^2$, which gives $\delta = R^{-1/4}X^{-1/2}$. This yields

$$\sum_{k=1}^{R} |E(p_k,X)|^2 \ll X^3 R^{1/2}\log^2 X + X^2\epsilon^{-3}.$$

□

6.3.3 A Large Sieve Inequality

We need two technical lemmas to prove Theorem 6.16. The first one is Lemma 3.2 in [11].

Lemma 6.19. *Let $b=(b_1,\dots,b_R)\in\mathbb{C}^R$ be a unit vector and let $A=(a_{ij})$ be an $R\times R$ matrix over $\mathbb{C}$ with $|a_{ij}|=|a_{ji}|$. Then*

$$|b\cdot Ab| = |\sum_{i,j=1}^{R} b_i \overline{b}_j a_{ij}| \le \max_i \sum_{j=1}^{R} |a_{ij}|.$$

We also need to compute the inverse Selberg transform for the Gaussian at different frequencies.

Lemma 6.20. *Let $h(1+t^2) = e^{-t^2/(2T)^2}\cos(rt)$. The inverse Selberg transform k of h satisfies for all $x>0$*

$$k(\cosh x) \ll T^3 \frac{(x+r)e^{-T^2(x+r)^2} + (x-r)e^{-T^2(x-r)^2}}{\sinh x},$$

and

$$k(1) \ll \min(T^3, r^{-3}).$$

Proof of Lemma 6.20. According to [25, §3 Lemma 5.5], the inverse Selberg transform k of h for $x\ge 1$ is given by

$$-2\pi k(x) = \frac{d}{dx}\frac{1}{2\pi}\int_{-\infty}^{\infty} h(1+t^2)e^{-it\operatorname{arccosh} x}\,dt.$$

Hence, by a direct computation

$$-2\pi k(\cosh x)=\frac{1}{2\pi i}\frac{1}{\sinh x}\int_{-\infty}^{\infty}e^{-t^2/(2T)^2}\cos(rt)te^{-itx}\,dt,$$

which is just the Fourier transform of a Gaussian times $t\cos(rt)$. Denote the non-normalised Fourier transform of an integrable function f by $\mathscr{F}\{f\}$, that is,

$$\mathscr{F}\{f\}(x)=\int_{-\infty}^{\infty}f(\xi)e^{-ix\xi}\,d\xi.$$

Then, by standard results [36, 17.22 (2)]

$$\begin{aligned}\mathscr{F}\{t\cos(rt)e^{-t^2/(2T)^2}\}(x)&=i\frac{d}{dx}\mathscr{F}\{\cos(rt)e^{-t^2/(2T)^2}\}(x)\\&=\frac{i}{2}\frac{d}{dx}\Big(\mathscr{F}\{e^{-t^2/(2T)^2}\}(x-r)+\mathscr{F}\{e^{-t^2/(2T)^2}\}(x+r)\Big),\end{aligned}$$

and since [36, 17.23 (13)]

$$\mathscr{F}\{e^{-t^2/(2T)^2}\}(x)=2\sqrt{\pi}Te^{-x^2T^2},$$

we have

$$\mathscr{F}\{t\cos(rt)e^{-t^2/(2T)^2}\}(x)=-2i\sqrt{\pi}T^3\Big((x+r)e^{-T^2(x+r)^2}+(x-r)e^{-T^2(x-r)^2}\Big).$$

It follows that

$$k(\cosh x)=\frac{2\sqrt{\pi}}{4\pi^2}T^3g(x),$$

where

$$g(x)=\frac{(x+r)e^{-T^2(x+r)^2}+(x-r)e^{-T^2(x-r)^2}}{\sinh x}.$$

Furthermore, taking the limit as $x\to 0$, we get

$$\begin{aligned}\lim_{x\to 0}g(x)&=\lim_{x\to 0}\frac{e^{-T^2(x+r)^2}+e^{-T^2(x-r)^2}-2T^2((x+r)^2e^{-T^2(x+r)^2}+(x-r)^2e^{-T^2(x-r)^2}}{\cosh x}\\&=2e^{-T^2r^2}(1-2T^2r^2).\end{aligned}$$

Let $u(x)=2e^{-x^2}(1-2x^2)$. Hence,

$$k(1)=\frac{1}{2\pi^{3/2}}T^3u(rT).$$

Since $u(x)$ is bounded, we get trivially that $k(1) \ll T^3$. On the other hand,

$$T^3 u(rT) = \frac{1}{r^3}(Tr)^3 u(rT) \ll r^{-3}.$$

It follows that

$$k(1) \ll \min(T^3, r^{-3}).$$

□

Remark 6.9. It is possible to reformulate the above proof in terms of the spherical eigenfunctions of Δ following [25, §3 Lemma 5.2] and [39, 40]. The problem with this more general approach is that we cannot see the Fourier transform that appears naturally in the explicit formulas.

Proof of Theorem 6.16. Let S be the left-hand side of (6.40). Since $\mathbb{C}^R$ is self-dual, it follows from Riesz representation theorem that there exists a unit vector $\mathbf{b} = (b_1, \ldots, b_R)$ in $\mathbb{C}^R$ such that

$$S = \left(\sum_{k=1}^{R} b_k \left(\sum_{|t_j| \le T} a_j x_k^{it_j} u_j(p) \right) \right)^2.$$

Then, by the Cauchy–Schwarz inequality

$$S \le ||a||_*^2 \widetilde{S},$$

where

$$\widetilde{S} = \sum_{|t_j| \le T} \left| \sum_{k=1}^{R} b_k x_k^{it_j} u_j(p) \right|^2.$$

In order to understand the sum $\widetilde{S}$, we smooth it out by a Gaussian centered around zero. This allows us to apply Lemma 6.20, which shows that the Selberg transform for a Gaussian is easy to compute. Thus,

$$\widetilde{S} \ll \sum_{j} e^{-t_j^2/(4T^2)} \left| \sum_{k=1}^{R} b_k x_k^{it_j} u_j(p) \right|^2.$$

After we open up the squares and interchange the order of summation, we apply Lemma 6.19 to get

$$S \ll ||a||_*^2 \max_k \sum_{l=1}^{R} |S_{kl}|,$$

where

$$S_{kl} = \sum_{j} e^{-t_j^2/(4T^2)} \cos(r_{kl} t_j) |u_j(p)|^2,$$

and

$$r_{kl} = \left| \log \frac{x_k}{x_l} \right|.$$

We can identify S_{kl} as the diagonal contribution in the spectral expansion of an automorphic kernel with $h(1+t^2) = e^{-t^2/(4T^2)} \cos(r_{kl} t)$. It follows from Lemma 6.20 that

$$S_{kl} \ll \min(T^3, r_{kl}^{-3}) + \sum_{\gamma \neq \mathrm{id}} T^3 e^{-T^2(d(\gamma p, p) - r_{kl})^2}. \tag{6.48}$$

The standard hyperbolic lattice point problem (e.g. [25, §2 Lemma 6.1]) gives

$$\#\{\gamma \in \Gamma : \delta(p, \gamma q) \leq x\} \ll x^2,$$

where the implied constant depends on Γ and p. We can rewrite this as

$$\log(1 + \#\{\gamma \in \Gamma : r < d(p, \gamma q) \leq r + 1\}) \ll r^2 + 1.$$

This shows that the series in (6.48) converges as $T \to \infty$, so that

$$S_{kl} \ll \min(T^3, r_{kl}^{-3}).$$

Hence, by the mean value theorem

$$\sum_{l=1}^{R} |S_{kl}| \ll \sum_{l=1}^{R} \min(T^3, X^3 |x_k - x_l|^{-3}).$$

The case $l = k$ yields T^3. So suppose $l \neq k$, then separate the x_l for which $T \leq X|x_k - x_l|^{-1}$. By the spacing condition, there are at most $2XT^{-1}\epsilon^{-1}$ such points. Hence,

$$\begin{aligned} \sum_{l=1}^{R} |S_{kl}| &\ll T^3 X T^{-1} \epsilon^{-1} + \int_1^\infty \frac{X^3}{|XT^{-1} + \epsilon u|^3} \, du + T^3 \\ &\ll T^2 X \epsilon^{-1} + T^3. \end{aligned} \tag{6.49}$$

□

6.3.4 Dependence of the Large Sieve on the Dimension

We conclude with a discussion on the dimensional limitations in applying the large sieve to hyperbolic lattice point problems. For simplicity, we restrict the discussion to cocompact Γ, albeit the conclusions apply to cofinite Γ as well. Moreover, the same arguments apply for counting in conjugacy classes as in [14]. Let us first consider the spatial average. In his thesis Chamizo [12, §2] proves, for $z_1, \dots, z_R \in \Gamma\backslash\mathbb{H}^2$ with $\tilde{d}(z_j, z_k) > \varepsilon > 0$ for $j \neq k$, that

$$\sum_{k=1}^{R} \left| \sum_{|t_j| \leq T} a_j u_j(z_k) \right|^{2l} \ll T^{2l}(1 + \varepsilon^{-2} T^{-2}) ||a||_*^{2l}, \tag{6.50}$$

where $l \in \mathbb{N}$ and the implied constant depends only on Γ and l, and $||a||_*$ is defined analogously to (6.41). It is important to note that in (6.50) the T^{2l} essentially corresponds to local Weyl law on $\Gamma\backslash\mathbb{H}^2$, while the $1 + \varepsilon^{-2}T^{-2}$ is the decay provided by the sieving. This leads to a mean square estimate

$$\sum_{k=1}^{R} |E(z_k, w, X)|^{2l} \ll X^{(l-2)/3} R^{(l-2)/3l} X^{l+\epsilon} \varepsilon^{-2} + X^{4l/3+\epsilon} R^{1/3},$$

where $E(z_k, w, X)$ is the error term of the standard lattice point problem on $\Gamma\backslash\mathbb{H}^2$. It is straightforward to see that the inequality is only effective if $l = 1, 2$, as $\varepsilon^{-2} \asymp R$. On the other hand, in n dimensions the upper bound in the large sieve inequality becomes (generalising Theorem 2.3 in [12])

$$T^{nl}(1 + \varepsilon^{-n} T^{-n}) ||a||_*^{2l},$$

so that the growth coming from Weyl's law is counterbalanced by the decay in the sieving term. However, $||a||_*^{2l}$ is always bounded by at least $X^{nl}T^{-1}$, which means that the large sieve does not provide enough savings apart from when $l = 1$. For example, in our case (Proposition 6.17) in three dimensions, the corresponding fourth moment for the error term is

$$\sum_{k=1}^{R} |E(p_k, X)|^4 \ll X^8 R^{-1} \log^3 X + X^5 \epsilon^{-3} R^{-1/4} \log^3 X,$$

which does not improve on the pointwise bound $O(X^{3/2})$. Hence, we expect that in n dimensions it is possible to obtain the appropriate conjecture $O(X^{(n-1)/2+\epsilon}$ for the error term on average only in the second moment, but not for higher.

For radial averages the picture is more grim. The higher moment large sieve inequality in [12] is

$$\sum_{k=1}^{R} \left| \sum_{|t_j| \leq T} a_j x_k^{it_j} u_j(z) \right|^{2l} \ll T^{2l}(1 + X\varepsilon^{-1} T^{-1}) ||a||_*^{2l},$$

where $x_1, \ldots, x_R \in [X, 2X]$ with $|x_j - x_k| > \varepsilon > 0$ for $j \neq k$ and $z \in \Gamma\backslash\mathbb{H}^2$. The corresponding mean square for the standard counting is

$$\sum_{k=1}^{R} |E(z, w, x_k)|^{2l} \ll R^{(l-1)/3l} X^{(4l+2)/3+\epsilon} \varepsilon^{-1} + R^{1/3} X^{4l/3+\epsilon},$$

which only yields an improvement for $l = 1$. Now, recall the form of the radial large sieve in three dimensions (6.40). Heuristically, in n dimensions the second moment would become

$$\sum_{k=1}^{R} \Bigg| \sum_{|t_j| \leq T} a_j x_k^{it_j} u_j(p) \Bigg|^2 \ll T^n (1 + X\varepsilon^{-1} T^{-1}) \|a\|_*^2.$$

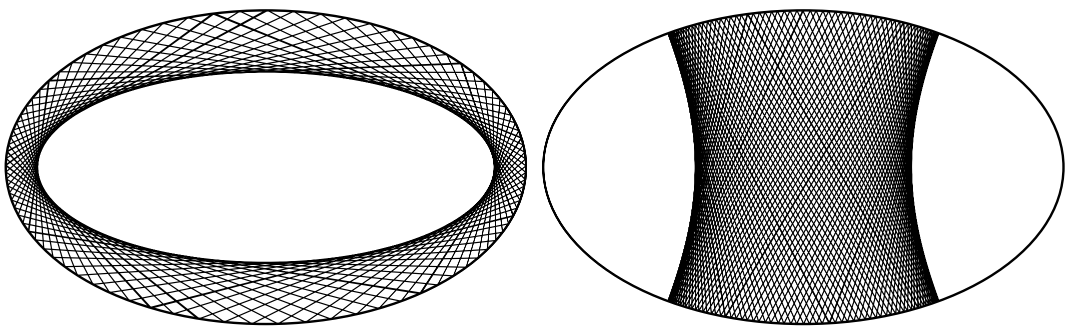
Here the sieving coefficient $1 + X\varepsilon^{-1}T^{-1}$ is independent of the dimension and hence any savings get worse compared to T^n as n increases. Therefore, the difference between the pointwise bound and the large sieve for the radial mean square gets smaller as $n \to \infty$, which is already evident in Theorem 6.13.

Chapter 7

Quantum Unique Ergodicity and Eisenstein Series

7.1 A Brief Introduction to Quantum Chaos

In classical mechanics chaos means extreme (exponential) sensitivity to initial conditions. Consider a billiard, that is, the flow of a particle with no external forces in a domain $\Omega \subset \mathbb{R}^2$ with boundary. In this case the angle of incidence equals the angle of reflection. Depending on the shape of Ω this flow can be integrable or extremely chaotic. The elliptic billiard is integrable. To see this, consider a trajectory that is

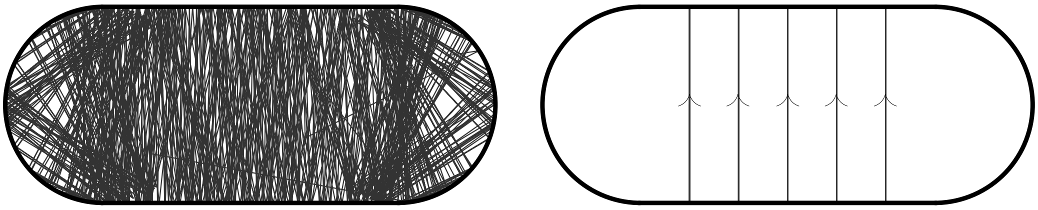


(a) "Whispering gallery" effect, (b) Trajectories bounded by a hyperbola,

Figure 7.1: Trajectories in an elliptic billiard.

tangent to a confocal ellipse C. Such a trajectory will always be restricted as shown in Figure 7.1a and remains tangent to C, [91, pg. 3]. Another integrable component is given by the trajectories bounded by a confocal hyperbola as can be seen in Figure 7.1b.

On the other hand, the stadium S in Figure 7.2a is an example of an ergodic system,



(a) Ergodic flow,

(b) Periodic orbits,

Figure 7.2: The Bunimovich Stadium.

which means that almost all of the trajectories become equidistributed with respect to the Liouville measure on S, [9]. Notice that it is possible for ergodic systems to have families of periodic orbits, called *bouncing ball trajectories*, as show in Figure 7.2b.

In quantum mechanics chaos is understood in the context of the correspondence principle, that is, studying quantum mechanical systems that are chaotic in the semi-classical limit $\hbar \to 0$. The possible energy states of a quantum system $\mathcal{Q}$ in a compact space X are governed by its Hamiltonian $\hat{H}$ and in particular by the spectrum of $\hat{H}$. In the case of a free particle of mass m (e.g. a billiard), the Hamiltonian $\hat{H}$ is given by

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta,$$

where $\hbar$ is the normalised Planck's constant. The time evolution of $\mathcal{Q}$ is determined by the Schrödinger equation

$$i\hbar\frac{\partial \Psi}{\partial t} = \hat{H}\Psi, \tag{7.1}$$

where $\Psi(\mathbf{x}, t)$ is the wave function with Dirichlet boundary conditions. Recall that $|\Psi(\mathbf{x}, t)|^2$ measures the probability of finding the particle at a location $\mathbf{x}$ at time t. Since for physical systems the Hamiltonian is a Hermitian operator [37, (3.18)], the equation (7.1) is a linear partial differential equation. It turns out that if the classical system is chaotic then the dynamics on $\mathcal{Q}$ are ergodic. Finally, since wave functions live in the domain of $\hat{H}$, that is, $L^2(X)$, it is possible to write down the eigenfunction expansion of Ψ as

$$\Psi(\mathbf{x}, t) = \sum_n \langle \Psi, \phi_n \rangle \phi_n(\mathbf{x}, t),$$

where $\langle f, g \rangle = \int_X f\overline{g}$ is the standard inner product on X. The eigenfunction ϕ_n with the eigenvalue E_n satisfies the eigenvalue equation

$$-\frac{\hbar^2}{2m}\Delta\phi_n = E_n\phi_n.$$

The eigenfunctions are stationary states of the system, and we can write [41, (2.2)]

$$\phi_n(\mathbf{x}, t) = e^{itE_n/\hbar}\phi_n(x, 0). \tag{7.2}$$

If we normalise the eigenvalues (energy levels) as $\lambda_n = E_n 2m/\hbar^2$, then the semi-classical limit $\hbar \to 0$ corresponds to the large eigenvalue (high energy) limit $\lambda_n \to \infty$. This identification will be useful to us later on.

7.1.1 Quantum Unique Ergodicity

We have already seen what ergodicity means for the billiards. Formally, we can express this as follows. Let (X, μ) be a compact Hausdorff space with a finite Borel measure μ. Given a μ-invariant measurable homeomorphism $T : X \longrightarrow X$, we say that T is ergodic if for almost all $x \in X$,

$$\lim_{N\to\infty} \frac{\#\{j \in \mathbb{N} : 0 \leq j \leq N,\, T^j(x) \in A\}}{N} = \frac{\mu(A)}{\mu(X)},$$

for every measurable subset $A \subset X$, [41, §1]. This is equivalent to the condition that any T-invariant measurable subset $A \subset X$ satisfies $\mu(A) = 0$ or $\mu(A) = \mu(X)$.

For example, in the case of the billiard, X is the unit tangent bundle of Ω and μ is the Liouville measure on X. The map $T = \Phi_t(x, \theta)$ then gives the location and direction of the particle at time t. Equivalently, T is ergodic if and only if every real-valued T-invariant funtion on X is constant almost everywhere. Of course, μ is an example of such a function. It is natural to ask whether there can be any other T-invariant functions. In general the answer is yes, but it turns out that for some systems this is not the case. Hence, if μ is the only constant T-invariant function on X, we say that T is *uniquely ergodic*. In particular, unique ergodicity implies that every orbit is dense [41, pg. 162].

We now extend these definitions to the quantum setting according to [41, §2]. We will give a simplified explanation as we wish to avoid the consideration of momentum (phase space) which leads to microlocal analysis and pseudodifferential operators [92, pg. 213]. The definitions we choose should be compatible with the classical ones when we take the limit $\lambda \to \infty$. As before, consider a quantum system $\mathcal{Q}$ of a free particle in the space X under the normalised Laplacian Δ. Informally, the relation of equidistribution of quantum orbits to that of the classical system can be formulated as follows: suppose we can express a wave function in terms of high-energy states (ϕ_n for large n), then over a large period of time this wave function should approach the uniform probability distribution on X. More concretely, suppose that we have a sequence of wave functions Ψ_j that can be written as a linear combination of ϕ_n with $\lambda_j \leq n \leq 2\lambda_j$. Let

$$\tau_{\Psi_j} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\Psi_j(\mathbf{x}, t)|^2 \, dt,$$

be the time average of Ψ_j. Then the measures τ_{Ψ_j} should tend to the uniform measure μ on X in the weak-$*$ limit as $j \to \infty$, that is,

$$\int_X f \, d\tau_{\Psi_j} \to \int_X f \, d\mu, \qquad \text{as } j \to \infty,$$

for any continuous function f on X. If this is true for any sequence λ_j and Ψ_j, then the eigenstates ϕ_j become equidistributed. This follows from the observation that

$$\tau_{\phi_n} = |\phi_n(\mathbf{x}, 0)|^2,$$

see (7.2). The converse is true as well, so that if the ϕ_j become equidistributed, then any τ_{Ψ_j} tends to the uniform distribution on X. Hence, if we define $\mu_j = |\phi_j|^2 \mu$, then the system $\mathscr{Q}$ is *quantum unique ergodic* (QUE) if

$$\int_X f \, d\mu_j \to \int_X f \, d\mu,$$

as $j \to \infty$ for any continuous function f on X. With the above identification it is now simple to formulate quantum ergodicity. Again, informally this should mean that almost all quantum orbits become equidistributed. In terms of the ϕ_j, we say that $\mathscr{Q}$ is *quantum ergodic* (QE) if there is a full density subsequence λ_{j_k} of λ_j, that is,

$$\sum_{\lambda_{j_k} \leq T} 1 \sim \sum_{\lambda_j \leq T} 1,$$

as $T \to \infty$, such that $\mu_{j_k} \to \mu$ in the weak-$*$ sense.

7.1.2 Arithmetic Quantum Chaos

Suppose M is a compact negatively curved Riemannian manifold (without boundary) with the unit tangent bundle $X = SM$, then the geodesic flow on X is a classical dynamical system. This flow is ergodic [92]. Thus, we are led to investigate what happens to the "quantised flow", in terms of the eigenfunctions ϕ_j of Δ on M, in the large eigenvalue limit. The following answer was given by Shnirelman [93], Zelditch [106] and Colin de Verdiére [103].

Theorem 7.1 (Quantum Ergodicity Conjecture)**.** *Let M be a compact negatively curved Riemannian manifold with the Laplace–Beltrami operator Δ and the standard measure μ. Let $\{\phi_j\}$ be an orthonormal basis for $L^2(M)$ of eigenfunctions of Δ with corresponding eigenvalues λ_j. Define $\mu_j = |\phi_j|^2\mu$. Then*

$$\mu_{j_k} \to \mu,$$

as $k \to \infty$, for some full density subsequence j_k, in the weak-$*$sense.

On the other hand, the QUE conjecture for general M is hopelessly out of reach. When $M = \Gamma\backslash\mathbb{H}^2$, and Γ is arithmetic, more tools are available. First of all, the geodesic flow on such M is strongly chaotic (Asonov) [92] and so ergodic. Second, more tools are available for the treatment of the ϕ_j due to the existence of Hecke operators and more explicit Fourier expansions of generalised eigenfunctions of Δ. This lead to the following conjecture by Rudnick and Sarnak [89] in 1994.

Quantum Unique Ergodicity Conjecture. *For a compact negatively curved manifold M, QUE holds. In other words, the measures $\mu_j \to \mu$ as $\lambda_j \to \infty$.*

There has been substantial progress towards the conjecture, but we are still far away from the full result. Let $M = \Gamma\backslash\mathbb{H}^2$, where $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ is discrete. For Γ of particular arithmetic type the distribution of the eigenstates is well-understood. The precise definition and a discussion of arithmetic Fuchsian groups is given in [57, §5]. We choose to omit the lengthy details. In essence, arithmeticity is related to whether the trace $\mathrm{Tr}\,\gamma$ for all $\gamma \in \Gamma$ is an algebraic integer in some number field K. For example, $\mathrm{PSL}_2(\mathbb{Z})$ and its congruence subgroups are arithmetic.

In 1995 Luo and Sarnak [66] proved the conjecture for Eisenstein series for non-compact arithmetic Γ and, in particular, for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. The precise result is that given Jordan measurable subsets A and B of M, then

$$\lim_{t\to\infty} \frac{\int_A |E(z,\frac{1}{2}+it)|^2\, d\mu(z)}{\int_B |E(z,\frac{1}{2}+it)|^2\, d\mu(z)} = \frac{\mu(A)}{\mu(B)}, \tag{7.3}$$

where $\mu(B) \neq 0$. They actually compute the asymptotic[1] explicitly

$$\int_A |E(z, \tfrac{1}{2} + it)|^2 \, d\mu(z) \sim \frac{6}{\pi} \mu(A) \log t,$$

as $t \to \infty$. Jakobson [55] extended (7.3) to the unit tangent bundle. The result of Luo and Sarnak was also generalised to $\mathrm{PSL}_2(\mathcal{O}_K)\backslash\mathbb{H}^3$ by Koyama [58], where $\mathrm{PSL}_2(\mathcal{O}_K)$ is the ring of integers of an imaginary quadratic field of class number one, and to $\mathrm{PSL}_2(\mathcal{O}_K)\backslash(\mathbb{H}^2)^n$ with K a totally real field of degree n and narrow class number one by Truelsen [101]. In particular, the asymptotic in [101] for $d\mu_{m,t} = |E(z, \frac{1}{2} + it, m)|^2 d\mu$ is

$$\mu_{m,t} \sim \frac{(2\pi)^n nR}{2d_K \zeta_K(2)} \log t,$$

where $E(z,s,m)$ are a family of Eisenstein series parametrised by $m \in \mathbb{Z}^{n-1}$, ζ_K is the Dedekind zeta function and R and d_K are the regulator and discriminant of K, respectively. The QUE for ϕ_j a Hecke–Maaß eigenform was proven by Lindenstrauss [65] in the compact case and Soundararajan [95] in the non-compact case, thus completing the full QUE conjecture for all arithmetic surfaces. Holowinsky and Soundararajan [48] study QUE in the holomorphic case. They consider holomorphic, L^2 normalised Hecke cusp forms f_k of weight k for $\mathrm{SL}_2(\mathbb{Z})$. They prove that the measures $|y^{k/2} f_k(z)|^2 d\mu$ converge weakly to $d\mu$ as $k \to \infty$. Another interesting direction for the QUE of Eisenstein series has recently been proposed by Young [104], who proves equidistribution of Eisenstein series for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ when they are restricted to "thin sets", e.g. geodesics connecting 0 and ∞ (as opposed to restricting to compact Jordan measurable subsets of $\Gamma\backslash\mathbb{H}^2$ as in [66]).

For a general cofinite $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ it is not clear whether there are infinitely many cusp forms so that the limit of $|\phi_j|^2 d\mu$ might not be relevant. Petridis, Raulf, and Risager [80] (see also [79]) propose to study the scattering states of Δ instead of the cuspidal spectrum. The scattering states arise as the residues of Eisenstein series on the left half-plane ($\operatorname{Re} s < 1/2$) at the non-physical poles of the scattering matrix. These poles are called resonances. Let ρ_n be a sequence of poles of the scattering matrix for $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2$ (this corresponds to half a non-trivial zero of ζ). Define

$$u_{\rho_n}(z) = (\operatorname*{res}_{s=\rho_n} \varphi(s))^{-1} \operatorname*{res}_{s=\rho_n} E(z,s).$$

The normalisation is chosen so that u_{ρ_n} has simple asymptotics $y^{1-\rho_n}$ for its growth at

[1] The actual (erroneous) constant in [66] is $48/\pi$, which arises from a missing factor of 2 in the Fourier expansion of the Eisenstein series and a mistake in the value of an integral of a special function.

infinity. Their result is that for any compact Jordan measurable subset A of $\Gamma\backslash\mathbb{H}^2$,

$$\int_A |u_{\rho_n}(z)|^2 d\mu(z) \to \int_A E(z, 2-\gamma_\infty) d\mu(z),$$

where γ_∞ is the limit of the real part of the sequence of zeros of the Riemann zeta function. This is obtained by studying the quantum limits of Eisenstein series off the critical line. Since the results in two and three dimensions are analogous, we will not state them for $\mathbb{H}^2$.

Fix a square-free integer $D<0$ and let $K=\mathbb{Q}(\sqrt{D})$ be the corresponding imaginary quadratic number field of discriminant d_K. Let $\mathcal{O}$ be the ring of integers of K and take a $\mathbb{Z}$-basis as described in Section 4.3. Let $\Gamma = \mathrm{PSL}_2(\mathcal{O})$. For simplicity, restrict D so that K has class number one. This means that Γ has exactly one cusp (up to Γ-equivalence) which we may suppose is $\infty \in \mathbb{P}^1\mathbb{C}$. The imaginary quadratic fields of class number one are exactly those with $D = -1, -2, -3, -7, -11, -19, -43, -67, -163$. Let ρ_n be a sequence of poles of the scattering matrix $\varphi(s)$ of $E(p,s)$ and define

$$\upsilon_{\rho_n}(p) = (\operatorname*{res}_{s=\rho_n} \varphi(s))^{-1} \operatorname*{res}_{s=\rho_n} E(p,s).$$

From the explicit form of φ (4.22) we know that ρ_n is equal to a non-trivial zero of ζ_K. Define

$$s(t) = \sigma_t + it,$$

where $\sigma_t > 1$ is a sequence converging to $\sigma_\infty \geq 1$. Also, let γ_n be a sequence of real parts of the non-trivial zeros of ζ_K with $\lim \gamma_n = \gamma_\infty$. We will prove the following theorems.

Theorem 7.2. *Let A be a compact Jordan measurable subset of $\Gamma\backslash\mathbb{H}^3$. Then*

$$\int_A |\upsilon_{\rho_n}(p)|^2 d\mu(p) \to \int_A E(p, 4-2\gamma_\infty) d\mu(p)$$

as $n \to \infty$.

Notice that $4-2\gamma_\infty > 2$ so that we are in the region of absolute convergence. Under the GRH the limit becomes $E(p,3)d\mu(p)$.

Theorem 7.3. *Assume $\sigma_\infty = 1$ and $(\sigma_t - 1)\log t \to 0$. Let A and B be compact Jordan measurable subsets of $\Gamma\backslash\mathbb{H}^3$. Then*

$$\frac{\mu_{s(t)}(A)}{\mu_{s(t)}(B)} \to \frac{\mu(A)}{\mu(B)},$$

as $t \to \infty$. In fact, we have

$$\mu_{s(t)}(A) \sim \mu(A) \frac{2(2\pi)^2}{|\mathcal{O}^\times||d_K|\zeta_K(2)} \log t. \tag{7.4}$$

Let F be the fundamental domain of $\mathcal{O}$ as a lattice in $\mathbb{R}^2$. Since $|F| = \sqrt{|d_K|}/2$ and $\mathrm{vol}(\Gamma\backslash\mathbb{H}^3) = |d_K|^{3/2}\zeta_K(2)/(4\pi^2)$, [90, Proposition 2.1], it is also possible to express the constant in (7.4) in terms of the volumes.

Remark 7.1. The constant for the QUE of Eisenstein series *on* the critical line in Koyama [58] is $2/\zeta_K(2)$. However, there is a small mistake in his computations on page 485, where the residue of the double pole of $\zeta_K^2(s/2)$ goes missing. After fixing this (and taking into account the number of units of $\mathcal{O}$ which is normalised away in [58]) his result agrees with our limit (7.4) for $\sigma_\infty = 1$.

Theorem 7.4. *Assume $\sigma_\infty > 1$. Let A be a compact Jordan measurable subset of $\Gamma\backslash\mathbb{H}^3$. Then*

$$\mu_{s(t)}(A) \to \int_A E(p, 2\sigma_\infty) d\mu(p),$$

as $t \to \infty$.

Theorem 7.2 now follows easily.

Proof of Theorem 7.2. By the functional equation (4.24) of $E(p,s)$, we get

$$\begin{aligned}
|\upsilon_{\rho_n}|^2 d\mu(p) &= |(\operatorname*{res}_{s=\rho_n} \varphi(s))^{-1} \operatorname*{res}_{s=\rho_n} E(p,s)|^2 d\mu(p)\\
&= |(\operatorname*{res}_{s=\rho_n} \varphi(s))^{-1} \operatorname*{res}_{s=\rho_n} \varphi(s)E(p,2-s)|^2 d\mu(p)\\
&= |E(p, 2-\rho_n)|^2 d\mu(p).
\end{aligned}$$

We apply Theorem 7.4 with $\sigma_\infty = 2 - \gamma_\infty$ to conclude the proof. $\square$

Theorem 7.2 says that the measures $|\upsilon_\rho|^2 d\mu$ do *not* become equidistributed. We could of course renormalise the measures and use

$$d\nu_{s(t)}(p) = \left| \frac{E(p, s(t))}{\sqrt{E(p, 2\sigma_\infty)}} \right|^2 d\mu(p).$$

Then we have the following corollary.

Corollary 7.5. *Assume $\sigma_\infty > 1$. Let A be a compact Jordan measurable subset of $\Gamma\backslash\mathbb{H}^3$. Then*

$$\nu_{s(t)}(A) \to \mu(A),$$

as $t \to \infty$.

The measures $d\nu_{\rho_n}$ are not eigenfunctions of Δ so their equidistribution is not directly related to the QUE conjecture.

Remark 7.2. Dyatlov [24] investigated quantum limits of Eisenstein series and scattering states for more general Riemannian manifolds with cuspidal ends. He proves results analogous to Theorems 7.2 and 7.4. However, only the case of surfaces is explicitly written down and the constants are not identified as concretely as in arithmetic cases such as [80] or Theorems 7.2 and 7.4. Dyatlov uses a very different method of decomposing the Eisenstein series into plane waves and studying their microlocal limits, which does not use global properties of the surface, such as hyperbolicity.

7.2 Proof of Theorems 7.3 and 7.4

Let $M = \Gamma\backslash\mathbb{H}^3$. Since any function in $L^2(M)$ can be decomposed in terms of the Hecke–Maaß cusp forms $\{u_j\}$ and the incomplete Eisenstein series $E(p|\psi)$, it is sufficient to consider them separately.

7.2.1 Discrete Part

We will first prove that the discrete spectrum vanishes in the limit.

Lemma 7.6. *Let u_j be a Hecke–Maaß cusp form. Then*

$$\int_M u_j(p)|E(p,s(t))|^2 d\mu(p) \to 0,$$

as $t \to \infty$.

Proof. Denote the integral by

$$J_j(s(t)) = \int_M u_j(p)|E(p,s(t))|^2 d\mu(p).$$

We define
$$I_j(s) = \int_M u_j(p)E(p,s(t))E(p,s)\,d\mu(p).$$
Unfolding the integral gives
$$I_j(s) = \int_0^\infty \int_F u_j(p)E(p,s(t))y^s \frac{dx_1 dx_2 dy}{y^3}. \tag{7.5}$$

Let $\varkappa$ be the isometry corresponding to $z + yj \mapsto \overline{z} + yj$. This can also be identified with a map that takes $\gamma \in \mathrm{SL}_2(\mathbb{C})$ to its complex conjugate $\overline{\gamma}$. In particular $\varkappa$ acts on any automorphic f by $f_\varkappa(p) = f(\varkappa p)$. The action of $\varkappa$ on f commutes with Δ. It follows that we can choose a complete orthonormal basis of eigenfunctions of $L^2(M)$ consisting of simultaneous Maaß–Hecke cusp forms and $\varkappa$ eigenfunctions. For any such eigenform ϕ, we have that $\phi(\varkappa p) = \pm\phi(p)$ as $\varkappa$ is an involution. We can therefore decompose the space of such cusp forms into two orthogonal spaces depending on whether $\phi(\varkappa p) = \phi(p)$ or $\phi(\varkappa p) = -\phi(p)$. These are called *even* and *odd* cusp forms, respectively. Also, it is easy to verify that $E(\varkappa p, s) = E(p,s)$. Hence, by a change of variables, we may suppose that the u_j in $I_j(s)$ is even as the integral over the odd cusp forms vanishes.

Substituting Fourier expansions of the Eisenstein series (4.23) and the cusp forms (4.25) (only the cosine remains for even u_j) into (7.5) gives

$$\begin{aligned} I_j(s) = \int_0^\infty \int_F \Bigg(2y \sum_{0\neq n\in\mathcal{O}^*} \rho_j(n) K_{ir_j}(2\pi|n|y)\cos(2\pi\langle n,z\rangle)\Bigg) \\ \Bigg(y^{s(t)} + \varphi(s(t))y^{2-s(t)} \\ + \frac{2y}{\xi_K(s(t))} \sum_{0\neq m\in\mathcal{O}} |m|^{s(t)-1}\sigma_{1-s(t)}(m)K_{s(t)-1}\left(\frac{4\pi|m|y}{\sqrt{|d_K|}}\right) e^{2\pi i\langle \frac{2m}{\sqrt{d_K}}, z\rangle}\Bigg) y^s \frac{dx_1dx_2dy}{y^3} \end{aligned}$$

By the definition of F and the formula $\cos(a+b) = \cos a\cos b - \sin a \sin b$ it is simple to see that
$$\int_F \cos(2\pi\langle n,z\rangle)\,dz = \begin{cases} 0, & \text{if } 0 \neq n \in \mathcal{O}^*, \\ 1, & \text{if } n = 0. \end{cases}$$

Evaluating the integral over F tells us that only the terms with $n = \pm 2m/\sqrt{|d_K|}$ remain and that the integral over the imaginary part goes to zero. Hence, with the identification $\mathcal{O} \to \mathcal{O}^*$ by $\alpha \mapsto (2/\sqrt{d_K})\overline{\alpha}$, we get

$$I_j(s) = \frac{4}{\xi_K(s(t))} \int_0^\infty \sum_{0\neq n\in\mathcal{O}^*} |n|^{s(t)-1}\sigma_{1-s(t)}(n)\rho_j(n)K_{s(t)-1}(2\pi|n|y)K_{it_j}(2\pi|n|y)y^s\frac{dy}{y}.$$

Change of variables $y \mapsto y/|n|$ yields

$$I_j(s) = \frac{4}{\xi_K(s(t))} \sum_{0 \neq n \in \mathcal{O}^*} \frac{|n|^{s(t)-1}\sigma_{1-s(t)}(n)\rho_j(n)}{|n|^s} \int_0^\infty K_{s(t)-1}(2\pi y) K_{it_j}(2\pi y) y^s \frac{dy}{y}.$$

We can evaluate the integral by [36, 6.576 (4) and 9.100] to get

$$I_j(s) = \frac{4}{\xi_K(s(t))} \frac{2^{-3}\pi^{-s}}{\Gamma(s)} \prod \Gamma\left(\frac{s \pm (s(t)-1) \pm it_j}{2}\right) R(s),$$

where the product is taken over all combinations of $\pm$ and

$$R(s) = \sum_{0 \neq n \in \mathcal{O}^*} \frac{|n|^{s(t)-1}\sigma_{1-s(t)}(n)\rho_j(n)}{|n|^s}.$$

Since u_j is a Hecke eigenform, we can factorise $R(s)$ with Theorem 4.2 as follows.

$$\begin{aligned}
R(s) &= \rho_j(1) \prod_{(p):\text{prime ideal}} \sum_{k=0}^{\infty} \frac{\lambda_j(p^k)|p|^{k(s(t)-1)}\sigma_{1-s(t)}(p^k)}{|p|^{ks}} \\
&= \rho_j(1) \prod_{(p)} \sum_{k=0}^{\infty} \frac{\lambda_j(p^k)|p|^{k(s(t)-1)}}{|p|^{ks}} \sum_{l=0}^{k} |p|^{2(1-s(t))l} \\
&= \rho_j(1) \prod_{(p)} \sum_{k=0}^{\infty} \frac{\lambda_j(p^k)|p|^{k(s(t)-1)}}{|p|^{ks}} \frac{1-|p|^{2(1-s(t))(k+1)}}{1-|p|^{2(1-s(t))}} \\
&= \rho_j(1) \prod_{(p)} \frac{1}{(1-|p|^{2(1-s(t))})} \left(\sum_{k=0}^{\infty} \lambda_j(p^k)|p|^{-k(s-s(t)+1)} \right. \\
&\quad \left. - |p|^{-2(s(t)-1)} \sum_{k=0}^{\infty} \lambda_j(p^k)|p|^{-k(s+s(t)-1)} \right) \\
&= \rho_j(1) \prod_{(p)} \frac{1}{1-|p|^{2(1-s(t))}} \left(\frac{1}{1-\lambda_j(p)|p|^{-(s-s(t)+1)} + |p|^{-2(s-s(t)+1)}} \right. \\
&\quad \left. - \frac{|p|^{-2(s(t)-1)}}{1-\lambda_j(p)|p|^{-(s+s(t)-1)} + |p|^{-2(s+s(t)-1)}} \right) \\
&= \rho_j(1) \prod_{(p)} \frac{1-|p|^{-2s}}{1-\lambda_j(p)|p|^{-(s-s(t)+1)} + |p|^{-2(s-s(t)+1)}} \\
&\quad \times \frac{1}{1-\lambda_j(p)|p|^{-(s+s(t)-1)} + |p|^{-2(s+s(t)-1)}} \\
&= \rho_j(1) \frac{L(u_j, \frac{s-s(t)+1}{2}) L(u_j, \frac{s+s(t)-1}{2})}{\zeta_K(s)}
\end{aligned}$$

Now,

$$J_j(t) = I_j(\overline{s(t)}),$$

so

$$\begin{aligned} J_j(t) &= \frac{2^{-1}}{\xi_K(s(t))} \frac{\pi^{-\overline{s(t)}}}{\Gamma(\overline{s(t)})} \prod \Gamma\left(\frac{\overline{s(t)} \pm (s(t)-1) \pm i t_j}{2}\right) \rho_j(1) \frac{1}{\zeta_K(\overline{s(t)})} L(u_j, \tfrac{1}{2} - it) L(u_j, \sigma_t - \tfrac{1}{2}) \\ &= \frac{2^{s(t)-1} \pi^{2it} \rho_j(1)}{|d_K|^{s(t)/2} |\zeta_K(s(t))|^2} L(u_j, \tfrac{1}{2} - it) L(u_j, \sigma_t - \tfrac{1}{2}) \frac{\prod \Gamma\left(\frac{\overline{s(t)} \pm (s(t)-1) \pm i t_j}{2}\right)}{|\Gamma(s(t))|^2}. \end{aligned}$$

With Stirling's Formula (A.6) we see that the quotient of Gamma factors is $O(|t|^{1-2\sigma_t})$. Recall from (4.16) that

$$\log^{-2}|t| \ll \zeta_K(s(t)) \ll \log^2|t|.$$

It follows that we need a *subconvex* bound on the L-functions to guarantee vanishing. Petridis and Sarnak [82] show that there is a $\delta > 0$ such that

$$L(u_j, \tfrac{1}{2} + it) \ll_j |1+t|^{1-\delta}.$$

In fact, they have $\delta = 7/166$, although this is not crucial for us. Hence,

$$J_j(t) \to 0,$$

as $t \to \infty$. □

7.2.2 Continuous Part

Let $h(y) \in C^\infty(\mathbb{R}^+)$ be a rapidly decreasing function at 0 and ∞ so that $h(y) = O_N(y^N)$ for $0 < y < 1$ and $h(y) = O_N(y^{-N})$ for $y \gg 1$ for all $N \in \mathbb{N}$. Denote the Mellin transform of h by $H = \mathcal{M} h$, i.e.

$$H(s) = \int_0^\infty h(y) y^{-s} \frac{dy}{y}$$

and the Mellin inversion formula [76, A2.1] gives

$$h(y) = \frac{1}{2\pi i} \int_{(\sigma)} H(s) y^s \, ds$$

for any $\sigma \in \mathbb{R}$. We consider the incomplete Eisenstein series denoted by

$$F_h(p) = E(p|h) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} h(y(\gamma p)) = \frac{1}{2\pi i} \int_{(3)} H(s) E(p,s)\, ds,$$

where h is a smooth function on $\mathbb{R}^+$ with compact support. We wish to prove the following lemma.

Lemma 7.7. *Let h be a function satisfying the conditions stated above. Then*

$$\int_M F_h(p) |E(v, s(t))|^2 d\mu(p) \sim \begin{cases} \int_M F_h(p) E(p, 2\sigma_\infty) d\mu(p), & \text{if } \sigma_\infty > 1, \\ \frac{2(2\pi)^2}{|\mathcal{O}^\times| |d_K| \zeta_K(2)} \log t \int_M F_h(p) d\mu(p), & \text{if } (\sigma_t - 1)\log t \to 0, \end{cases}$$

as $t \to \infty$.

Proof. Now, unfolding gives

$$\begin{aligned} \int_M F_h(p) |E(p, s(t))|^2 d\mu(p) &= \int_M \frac{1}{2\pi i} \int_{(3)} H(s) E(p,s)\, ds\, |E(p, s(t))|^2 d\mu(p) \\ &= \int_0^\infty \frac{1}{2\pi i} \int_{(3)} H(s) y^s\, ds \int_F |E(p, s(t))|^2 d\mu(p) \\ &= \int_0^\infty h(y) \operatorname{vol}(F) \left(\sum_{n \in \mathcal{O}} |a_n(y, s(t))|^2 \right) \frac{dy}{y^3}. \end{aligned}$$

We will deal separately with the contribution of the $n = 0$ term and the rest. We factor out the constant $\operatorname{vol}(F)$ in the analysis below.

Contribution of the constant term

We know that

$$|a_0(y, s(t))|^2 = y^{2\sigma_t} + 2\operatorname{Re}(\varphi(s(t)) y^{2-2it}) + |\varphi(s(t))|^2 y^{4-2\sigma_t}.$$

The first term is

$$\int_0^\infty h(y) y^{2\sigma_t - 2} \frac{dy}{y} = H(2 - 2\sigma_t),$$

which converges to $H(2 - 2\sigma_\infty)$. For the second term we first have that

$$\varphi(s(t)) \int_0^\infty h(y) y^{-2it} \frac{dy}{y} = \varphi(s(t)) H(2it).$$

Since $H(s)$ is in Schwartz class in t, the function $H(2it)$ decays rapidly, whereas $\varphi(s(t))$ is bounded as is clear from the factorisation (4.12). By taking complex conjugates we see that the second term will also tend to zero. Finally, for the third expression in the constant term we get

$$|\varphi(s(t))|^2 \int_0^\infty h(y) y^{2-2\sigma_t} \frac{dy}{y} = |\varphi(s(t))|^2 H(2\sigma_t - 2).$$

If $\sigma_\infty \neq 1$ then

$$|\varphi(s(t))| = \left| \frac{2\pi}{s(t)-1} \frac{\zeta_K(s(t)-1)}{\zeta_K(s(t))}. \right|$$

To estimate this we need the convexity bound (4.13) for ζ_K which gives

$$\zeta_K(s(t)-1) = \zeta_K(\sigma_t - 1 + it) = O(|t|^{1-\sigma_t/2+\epsilon}).$$

We also need the bound (4.16) for $1/\zeta_K$. Of course we also have that

$$\frac{1}{s(t)-1} = O(|t|^{-1}).$$

Combining all of this, we get

$$\varphi(s(t)) = O(|t|^{-\sigma_t/2+\epsilon}),$$

and so

$$\varphi(s(t)) \to 0, \tag{7.6}$$

as $t \to \infty$, when $\sigma_\infty \neq 1$. So in summary, the contribution of the constant term converges to $H(2-2\sigma_\infty)$ if $\sigma_\infty \neq 1$ and is $O(1)$ otherwise.

Contribution of the non-constant terms

In this case the contribution equals

$$\begin{aligned} A(t) &= \int_0^\infty \frac{1}{2\pi i} \int_{(3)} H(s) y^s \, ds \frac{4y^2}{|\xi_K(s(t))|^2} \sum_{n \in \mathcal{O}}{}' |n|^{2\sigma_t - 2} |\sigma_{1-s(t)}(n)|^2 \left| K_{s(t)-1}\left(\frac{4\pi |n| y}{\sqrt{|d_K|}} \right) \right|^2 \frac{dy}{y^3} \\ &= \frac{4|\mathcal{O}^\times|}{|\xi_K(s(t))|^2} \frac{1}{2\pi i} \int_{(3)} H(s) \left(\frac{\sqrt{|d_K|}}{4\pi} \right)^s \sum_{n \in \mathcal{O}/\sim}{}' \frac{|\sigma_{1-s(t)}(n)|^2}{|n|^{s+2-2\sigma_t}} \int_0^\infty y^s |K_{s(t)-1}(y)|^2 \frac{dy}{y} \, ds, \end{aligned}$$

where $a \sim b$ if a and b generate the same ideal in $\mathcal{O}$ and prime in the summation denotes that it is taken over $n \neq 0$. We now need to evaluate the series. Keeping in mind that $N(p)=|p|^2$, we get by a standard calculation

$$\begin{aligned}
\sum_{n\in\mathcal{O}/\sim}{}' \frac{\sigma_a(n)\sigma_b(n)}{|n|^s} &= \prod_{(p):\text{prime ideal}} \sum_{k=0}^{\infty} \frac{\sigma_a(p^k)\sigma_b(p^k)}{|p|^{ks}} \\
&= \prod_{(p)} \sum_{k=0}^{\infty} \frac{1}{|p|^{ks}} \left(\frac{1-|p|^{2a(k+1)}}{1-|p|^{2a}}\right)\left(\frac{1-|p|^{2b(k+1)}}{1-|p|^{2b}}\right) \\
&= \prod_{(p)} \frac{1}{1-|p|^{2a}} \frac{1}{1-|p|^{2b}} \sum_{k=0}^{\infty} \Big(|p|^{-ks} \\
&\quad -|p|^{k(2a-s)+2a} - |p|^{k(2b-s)+2b} + |p|^{k(2a+2b-s)+2a+2b} \Big),
\end{aligned}$$

and hence

$$\begin{aligned}
\sum_{n\in\mathcal{O}/\sim}{}' \frac{\sigma_a(n)\sigma_b(n)}{|n|^s} &= \prod_{(p)} \frac{1}{1-|p|^{2a}} \frac{1}{1-|p|^{2b}} \left(\frac{1}{1-|p|^{-ks}} - \frac{|p|^{2a}}{1-|p|^{2a-s}} \right. \\
&\quad \left. - \frac{|p|^{2b}}{1-|p|^{2b-s}} + \frac{|p|^{2a+2b}}{1-|p|^{2a+2b-s}} \right) \\
&= \prod_{(p)} \frac{1-|p|^{2(a+b-s)}}{(1-|p|^{-s})(1-|p|^{2a-s})(1-|p|^{2b-s})(1-|p|^{2a+2b-s})} \\
&= \frac{\zeta_K(\frac{s}{2})\zeta_K(\frac{s}{2}-a)\zeta_K(\frac{s}{2}-b)\zeta_K(\frac{s}{2}-a-b)}{\zeta_K(s-a-b)}.
\end{aligned}$$

For $a=\overline{b}=1-s(t)$ and $s=s-2(\sigma_t-1)$ this becomes

$$\sum_{n\in\mathcal{O}/\sim}{}' \frac{|\sigma_{1-s(t)}(n)|^2}{|n|^{s-2\sigma_t+2}} = \frac{\zeta_K(\frac{s}{2}-\sigma_t+1)\zeta_K(\frac{s}{2}+it)\zeta_K(\frac{s}{2}-it)\zeta_K(\frac{s}{2}+\sigma_t-1)}{\zeta_K(s)}.$$

Again, by [36, 6.576 (4)] we see that

$$\int_0^{\infty} y^s |K_{s(t)-1}(y)|^2 \frac{dy}{y} = \frac{2^{s-3}}{\Gamma(s)} \Gamma(\tfrac{s}{2}-\sigma_t+1)\Gamma(\tfrac{s}{2}+it)\Gamma(\tfrac{s}{2}-it)\Gamma(\tfrac{s}{2}+\sigma_t-1).$$

Hence, $A(t)$ becomes

$$\begin{aligned}
A(t) &= \frac{|\mathcal{O}^\times|}{|\xi_K(s(t))|^2} \frac{1}{4\pi i} \int_{(3)} H(s) \frac{\xi_K(\frac{s}{2}-\sigma_t+1)\xi_K(\frac{s}{2}-it)\xi_K(\frac{s}{2}+it)\xi_K(\frac{s}{2}+\sigma_t-1)}{\xi_K(s)}\, ds. \\
&= \frac{|\mathcal{O}^\times|}{|\xi_K(s(t))|^2} \frac{1}{4\pi i} \int_{(3)} B(s)\, ds,
\end{aligned}$$

say. By the Dirichlet Class Number Formula (4.11) for ζ_K, the completed zeta function ξ_K has a simple pole at $s = 1$ with residue

$$\operatorname*{res}_{s=1} \xi_K(s) = \frac{1}{|\mathcal{O}^\times|}.$$

There is also a simple pole at $s = 0$, [18, Theorem 10.5.1 (3)]. It follows that the poles of $B(s)$ in the region $\operatorname{Re} s \geq 1$ are at $2 \pm 2it$, $2\sigma_t$, $2\sigma_t - 2$, and $4 - 2\sigma_t$. Moving the line of integration to $\operatorname{Re} s = 1$ gives

$$\begin{aligned}
A(t) = \frac{|\mathcal{O}^\times|}{2|\xi_K(s(t))|^2} \Bigg(& \operatorname*{res}_{s=2\pm 2it} B(s) + \operatorname*{res}_{s=2\sigma_t} B(s) + \delta_t \operatorname*{res}_{s=4-2\sigma_t} B(s) \\
& + (1-\delta_t) \operatorname*{res}_{s=2\sigma_t - 2} B(s) + \frac{1}{2\pi i} \int_{(1)} B(s)\, ds \Bigg), \\
= {} & A_1 + A_2 + \ldots + A_5,
\end{aligned}$$

where $\delta_t = 1$ if $\sigma_t < 3/2$ and 0 otherwise. We deal with each of the residues A_i separately.

(i) For the first term we have

$$A_1 = \frac{H(2 \pm 2it)}{|\xi_K(\sigma_t + it)|^2} \frac{\xi_K(2 - \sigma_t \pm it)\xi_K(1 \pm 2it)\xi_K(\sigma_t \pm it)}{\xi_K(2 \pm 2it)}.$$

By Stirling asymptotics and convexity estimates for the Dedekind zeta functions, the quotient of the ξ_K functions is bounded by $|t|^{1-2\sigma_t} \log^{10}|t|$. By virtue of H being of rapid decay in t it follows that $A_1 \to 0$ as $t \to \infty$.

(ii) The second term is

$$A_2 = H(2\sigma_t) \frac{\xi_K(2\sigma_t - 1)}{\xi_K(2\sigma_t)}.$$

If $\sigma_\infty \neq 1$ then

$$A_2 \to H(2\sigma_\infty) \frac{\xi_K(2\sigma_\infty - 1)}{\xi_K(2\sigma_\infty)},$$

but if $\sigma_t \to 1$ then

$$A_2 \sim H(2) \frac{1}{2|\mathcal{O}^\times|\xi_K(2)(\sigma_t - 1)}.$$

(iii) Now, in the third term we use the form (4.22) of φ and the fact that ξ_K satisfies the functional equation (4.10)

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

We can then write

$$A_3 = \delta_t H(4-2\sigma_t)|\varphi(s(t))|^2 \frac{\xi_K(3-2\sigma_t)}{\xi_K(4-2\sigma_t)}.$$

By (7.6) we have that $\varphi(s(t)) \to 0$ as $t \to \infty$ for $\sigma_\infty \neq 1$. Hence, if $\sigma_\infty \neq 1$, then

$$A_3 \to 0.$$

On the other hand, if $\sigma_\infty = 1$, then

$$A_3 \sim \frac{\delta_t}{2|\mathcal{O}^\times|} H(2)|\varphi(s(t))|^2 \frac{-1}{\xi_K(2)(\sigma_t - 1)},$$

which is bounded.

(iv) For the fourth term we have

$$A_4 = (1-\delta_t)\zeta_K(0)H(2\sigma_t - 2)|\varphi(s(t))|^2,$$

which clearly converges to 0 if $\sigma_\infty \neq 1$ and is bounded for $\sigma_\infty = 1$ as in the previous case.

(v) Finally, the fifth term is

$$\begin{aligned} A_5 &= \frac{|\mathcal{O}^\times|}{2|\xi_K(s(t))|^2} \frac{1}{2\pi i} \int_{(1)} B(s)\,ds \\ &= \frac{|\mathcal{O}^\times|}{2|\xi_K(\sigma_t + it)|^2} \mathscr{I}, \end{aligned}$$

where

$$\mathscr{I} = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(1+i\tau) \frac{|\xi_K(\sigma_t - \frac{1}{2} + i\tau)|^2 \xi_K(\frac{1}{2} + i(\tau + t))\xi_K(\frac{1}{2} + i(\tau - t))}{\xi_K(1+2i\tau)}\,d\tau.$$

We now estimate the growth of A_5 in terms of t. The exponential contribution from the gamma functions in the integral is equal to

$$(e^{-\frac{\pi}{2}|t|})^2 e^{-\frac{\pi}{2}|\tau+t|} e^{-\frac{\pi}{2}|\tau-t|} e^{\frac{\pi}{2}|2\tau|} \ll e^{-\pi|t|}.$$

This cancels with the exponential growth of $|\xi_K(s(t))|^2$. Since $H(1+i\tau)$ decays rapidly, we can bound $\zeta_K(\sigma_t - \frac{1}{2} + i\tau)$ polynomially and absorb it into H. Hence

$$A_5 \ll \frac{\log^4|t|}{(|t|^{\sigma_t - 1/2})^2} \int_{-\infty}^{\infty} \widetilde{H}(\tau)|\zeta_K(\tfrac{1}{2} + i(\tau + t))\zeta_K(\tfrac{1}{2} + i(\tau - t))|\, d\tau.$$

where $\widetilde{H}$ is a function of rapid decay. The Dedekind zeta functions can be estimated with the subconvex bound (4.15) of Heath-Brown [42]. We get

$$A_5 \ll |t|^{5/3 - 2\sigma_t + 2\epsilon} \log^4|t| \int_{-\infty}^{\infty} \widetilde{H}(\tau)(t^{-1} + |\tau t^{-1} + 1|)^{1/3+\epsilon}(t^{-1} + |\tau t^{-1} - 1|)^{1/3+\epsilon}\, d\tau,$$

which is $o(1)$ since $\sigma_t \geq 1$.

Hence we have proved that the integral

$$\int_M F_h(p)|E(p, s(t))|^2 d\mu(p)$$

converges to

$$\operatorname{vol}(F)\left(H(2 - 2\sigma_\infty) + H(2\sigma_\infty)\frac{\xi_K(2\sigma_\infty - 1)}{\xi_K(2\sigma_\infty)}\right),$$

if $\sigma_\infty \neq 1$. On the other hand, for $\sigma_\infty = 1$ the contribution is asymptotic to

$$\operatorname{vol}(F)H(2)\frac{1 - |\varphi(s(t))|^2}{2|\mathcal{O}^\times|\xi_K(2)(\sigma_t - 1)} + O(1). \tag{7.7}$$

To finish the proof, we apply Mellin inversion and unfold backwards to see that

$$\begin{aligned}
\operatorname{vol}(F)&\left(H(2 - 2\sigma_\infty) + H(2\sigma_\infty)\frac{\xi_K(2\sigma_\infty - 1)}{\xi_K(2\sigma_\infty)}\right) \\
&= \int_0^\infty h(y)\operatorname{vol}(F)(y^{2\sigma_\infty - 2 + 2} + \varphi(2\sigma_\infty)y^{-2\sigma_\infty + 2})\frac{dy}{y^3} \\
&= \int_0^\infty h(y)\left(\int_F E(z + jy, 2\sigma_\infty)\, dz\right)\frac{dy}{y^3} \\
&= \int_M F_h(p)E(p, 2\sigma_\infty)\, d\mu(p),
\end{aligned}$$

and

$$\operatorname{vol}(F)H(2) = \int_M F_h(p)\, d\mu(p).$$

For the second case, we need to estimate the quotient with the scattering matrix. We will show that

$$\frac{1 - |\varphi(s(t))|^2}{2|\mathcal{O}^\times|\xi_K(2)(\sigma_t - 1)} \sim \frac{2(2\pi)^2}{|\mathcal{O}^\times||d_K|\zeta_K(2)}\log t.$$

To see this let $G(\sigma)=\varphi(\sigma+it)\varphi(\sigma-it)$. Notice that

$$G'(\sigma)=\frac{\varphi'}{\varphi}(\sigma\pm it)G(\sigma),$$

where the $\pm$ denotes the linear combination

$$\frac{\varphi'}{\varphi}(\sigma\pm it)=\frac{\varphi'}{\varphi}(\sigma+it)+\frac{\varphi'}{\varphi}(\sigma-it).$$

We then apply the mean value theorem twice on the intervals $[1,\sigma]$ and $[1,\sigma']$, respectively. We get

$$\begin{aligned}\frac{G(1)-G(\sigma)}{1-\sigma}&=G'(\sigma')\\&=G(\sigma')\frac{\varphi'}{\varphi}(\sigma'\pm it)\\&=\left(G(1)-(1-\sigma')G(\sigma'')\frac{\varphi'}{\varphi}(\sigma''\pm it)\right)\frac{\varphi'}{\varphi}(\sigma'\pm it),\end{aligned}$$

where $1\leq\sigma''\leq\sigma'\leq\sigma$. On noticing that $G(1)=1$, this gives

$$\frac{1-|\varphi(\sigma+it)|^2}{1-\sigma}=\left(1-(1-\sigma')|\varphi(\sigma''+it)|^2\frac{\varphi'}{\varphi}(\sigma''\pm it)\right)\frac{\varphi'}{\varphi}(\sigma'\pm it).$$

Using the asymptotics

$$\frac{\varphi'}{\varphi}(\sigma\pm it)\sim-4\log t \tag{7.8}$$

and the fact that $|\varphi(\sigma+it)|$ is bounded for $\sigma\geq1$ proves the lemma. The estimate (7.8) follows immediately from (A.8) and the Weyl bound (4.17) for ζ_K'/ζ_K. □

Proof of Theorems 7.3 and 7.4. These follow now from Lemmas 7.6 and 7.7 by approximation arguments similar to [66] and [58]. □

Appendix A

Gamma Function

The Gamma function $\Gamma(z)$ is defined for $\operatorname{Re} s > 0$ by the absolutely convergent integral

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

Notice that $\Gamma(1) = 1$. It follows, after taking the limit, from [96, §2 Theorem 5.4] that Γ is holomorphic in this region. Integrating by parts in (A.1) shows that the Gamma function satisfies the functional equation

$$\Gamma(s+1) = s\Gamma(s). \tag{A.2}$$

This gives the relation $\Gamma(n+1) = n!$. The functional equation (A.2) can be used to meromorphically continue Γ to the whole complex plane [96, §6 Theorem 1.3] with simple poles at negative integers of residue $\operatorname{res}_{s=-n} \Gamma(s) = (-1)^n/n!$ for $n \in \mathbb{N} \cup \{0\}$. In particular the functional equation holds for all $s \in \mathbb{C}$ with the understanding that one takes the residues at the poles. Clearly $\overline{\Gamma(s)} = \Gamma(\bar{s})$ holds in the region of absolute convergence, and hence it holds for all $s \in \mathbb{C}$. It is an exercise in contour integration [96, §6 Theorem 1.4] to prove that

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}, \qquad \text{for } s \in \mathbb{C}. \tag{A.3}$$

Similarly [36, 8.334 (2)],

$$\Gamma(\tfrac{1}{2}+s)\Gamma(\tfrac{1}{2}-s) = \frac{\pi}{\cos \pi s}, \qquad \text{for } s \in \mathbb{C}. \tag{A.4}$$

Setting $s=0$ in the last equation gives $\Gamma(\frac{1}{2})=\sqrt{\pi}$. We will also need the *duplication formula* [36, 8.335 (1)]

$$\Gamma(s)\Gamma(s+\tfrac{1}{2})=2^{1-2s}\sqrt{\pi}\Gamma(2s), \qquad \text{for } s\in\mathbb{C}. \tag{A.5}$$

The asymptotic behaviour of $\Gamma(\sigma+it)$ for large t is particularly important. This follows from *Stirling's Formula* [76, A4.8], which states that

$$\log\Gamma(s)=(s-\tfrac{1}{2})\log s-s+\log\sqrt{2\pi}+O(|s|^{-1}),$$

for s with $|\arg s|<\pi-\epsilon$ for any $\epsilon>0$. After a straightforward computation we obtain

$$|\Gamma(\sigma+it)|\sim\sqrt{2\pi}e^{-\pi\frac{|t|}{2}}|t|^{\sigma-1/2}, \qquad \text{as } |t|\to\infty. \tag{A.6}$$

Stirling's Formula can also be used to give the following estimate for the logarithmic derivative of Γ,

$$\frac{\Gamma'}{\Gamma}(s)=\log s-\frac{1}{2s}+O(|s|^{-2}). \tag{A.7}$$

The function Γ'/Γ is sometimes called the *digamma function* and denoted by ψ. Finally, from (A.7) we deduce that for σ bounded and $|t|\to\infty$,

$$\psi(\sigma+it)=\frac{\Gamma'}{\Gamma}(\sigma+it)=\log|t|+O(1). \tag{A.8}$$

Appendix B

Radial Average of the Standard Lattice Point Problem in $\mathbb{H}^3$

In this appendix we will show that the radial large sieve, Theorem 6.16, yields the same improvement of 1/6 over the known error bound $O(X^{3/2})$ in the standard hyperbolic lattice point problem in three dimensions. We prove the following theorem.

Theorem B.1. *Let Γ be a cocompact subgroup of $\mathrm{PSL}_2(\mathbb{C})$ and fix $p,q \in \mathbb{H}^3$. Define, for $X > 2$,*

$$N(p,q,X) = \#\{\gamma \in \Gamma : d(\gamma p, q) \leq \operatorname{arccosh} X\},$$

$$M(p,q,X) = \frac{2\pi}{\operatorname{vol}(\Gamma\backslash\mathbb{H}^3)} X^2 + \pi \sum_{1<s_j<2} 2^{s_j} \frac{\Gamma(s_j-1)}{\Gamma(s_j+1)} u_j(p)\overline{u_j}(q) X^{s_j}, \tag{B.1}$$

and let

$$E(p,q,X) = N(p,q,X) - M(p,q,X).$$

Suppose $X_1, \ldots, X_R \in [X, 2X]$ such that $|X_k - X_l| \geq \epsilon > 0$ for all $k \neq l$, and that $R\epsilon \gg X$ and $R > X^{2/3}$. Then

$$\frac{1}{R}\sum_{k=1}^{R} |E(p,q,X_k)|^2 \ll X^{2+2/3} \log X,$$

and

$$\frac{1}{X}\int_X^{2X} |E(p,q,x)|^2 \, dx \ll X^{2+2/3} \log X. \tag{B.2}$$

Notice that the main term $M(p,q,X)$ agrees with Lax and Phillips [63] and [25, §5 Theorem 1.1].

Proof. Suppose $0 < H < 1$. Define the hyperbolic convolution $*$ by

$$(k_1 * k_2)(p,q) = \int_{\mathbb{H}^3} k_1(p,r)k_2(r,q)\,d\mu(r).$$

It follows from [25, §3 (5.36)] and [25, §3 Theorem 5.3] that the Selberg transform is multiplicative with respect to $*$. Recall the form of the Selberg transform h of k (4.26). The case $s = 1$ can be explicitly written as

$$h(1) = 4\pi \int_0^\infty k(\cosh u)u \sinh u\, du. \tag{B.3}$$

We can split h as we did in Section 6.2 and write it in terms of the d-transform (6.31) as

$$\begin{aligned} h(\lambda) &= \frac{\pi}{(s-1)}\left(\int_{\mathbb{R}} k(\cosh u)\cosh su\, du - \int_{\mathbb{R}} k(\cosh u)\cosh((2-s)u)\, du\right) \\ &= \frac{\pi}{(s-1)}(d(k,s) - d(k,2-s)), \end{aligned}$$

with the modification that the argument of k in $d(k,t)$ is now $\cosh u$ instead of $\cosh^2 u$.

Now, let k_1 and k_2 be characteristic functions of $[1, \cosh R]$ with $R = H + \operatorname{arccosh} X$ and $R = H$, respectively. Also let h_1 and h_2 be their respective Selberg transforms. Define

$$K^+(p,q) = \frac{(k_1 * k_2)(p,q)}{\operatorname{vol}(B(0,H))},$$

and analogously for K^-, so that we get

$$\sum_{\gamma\in\Gamma} K^-(p,\gamma q) < N(p,q,X) < \sum_{\gamma\in\Gamma} K^+(p,\gamma q).$$

Let h be the Selberg transform of K^+. Then the spectral expansion of K^+ is given by

$$\begin{aligned} \sum_{\gamma\in\Gamma} K^+(p,\gamma q) &= \sum_{t_j} h(t_j)u_j(p)\overline{u_j}(q) \\ &= \sum_{1\le s_j\le 2} h(t_j)u_j(p)\overline{u_j}(q) + \sum_{t_j\in\mathbb{R}\setminus\{0\}} h(t_j)u_j(p)\overline{u_j}(q). \end{aligned}$$

Since the Selberg transform is multiplicative under convolution, we have

$$h(t) = \frac{h_1(t)h_2(t)}{\operatorname{vol}(B(0,H))}.$$

We can now use our estimates and methods from Lemma 6.10 and Proposition 6.11 to evaluate h_1 and h_2. We will also use the fact that [25, §1 (2.7)]

$$\mathrm{vol}(B(0,H)) \sim \frac{4}{3}\pi H^3,$$

for small H. We now compute $h(\lambda)$ for each λ.

(i) For the small eigenvalues $\lambda \in (0,1)$ we get

$$\begin{aligned} h(\lambda) = &\frac{1}{\mathrm{vol}(B(0,H))}\left(\frac{\pi}{(s-1)}\right)^2 \left(\frac{2\sinh s(\operatorname{arccosh} X + H)}{s}\right. \\ &\left.- \frac{2\sinh(2-s)(\operatorname{arccosh} X + H)}{2-s}\right)\left(\frac{2\sinh sH}{s} - \frac{2\sinh(2-s)H}{2-s}\right) \\ = &\frac{3/4}{\pi H^3}\left(\frac{\pi}{(s-1)}\right)^2\left(\frac{(2X)^s}{s} - \frac{(2X)^{2-s}}{2-s} + O(HX^s)\right)\frac{H^3}{3}\left(s^2-(2-s)^2 + O(H^2)\right), \end{aligned}$$

by comparing the Taylor series of $(\sinh sH)/s$ and $(\sinh(2-s)H)/(2-s)$.

(ii) For the zero eigenvalue $\lambda = 0$, i.e. $s = 2$, we have as above that

$$\begin{aligned} h(\lambda) &= \frac{3\pi^2}{4\pi}\left(2X^2 + O(\log X + HX^2)\right)\left(\frac{2\cdot 2^2}{3!} + O(H^2)\right) \\ &= 2\pi X^2 + O(HX^2 + \log X). \end{aligned}$$

(iii) At $\lambda = s = 1$, we use definition (B.3) to see that

$$\begin{aligned} h(1) = &\frac{(4\pi)^2}{\mathrm{vol}(B(0,H))}\Big((\operatorname{arccosh} X + H)\cosh(\operatorname{arccosh} X + H) \\ &- \sinh(\operatorname{arccosh} X + H)\Big)(H\cosh H - \sinh H) \\ \ll &\, X\log X. \end{aligned}$$

(iv) Finally, and most importantly to us, by (6.33) we have for the embedded eigenvalues $\lambda > 1$ that

$$h(\lambda) = a(H,t_j)X^{1+it_j} + b(H,t_j)X^{1-it_j},$$

where

$$a(H,t),\, b(H,t) \ll \frac{1}{t^2}\min(1,(Ht)^{-3}).$$

The form of the main term in (B.1) follows from the above computations. We can now carry out the computation of the large sieve with this spectral expansion in an analogous way to the proof of Proposition 6.15. The only modification, apart from the estimates above, comes from the fact that in the spectral expansion we now have an actual Maaß form instead of the period integral. We use the local Weyl law (Lemma 6.12) to estimate this term. This leads to

$$\sum_{k=1}^{R} |E(p,q,X_k)|^2 \ll X^2 H^{-2} \log X + X^3 \epsilon^{-1} H^{-1} \log X + RX^4 H^2,$$

which is in the exact same form as before. Hence, after balancing,

$$\sum_{k=1}^{R} |E(p,q,X_k)|^2 \ll X^{3+2/3} \epsilon^{-1} \log X + RX^{2+2/3}.$$

The results claimed in Theorem B.1 now follow as in the proof of Theorem 6.13. □

Any improvement on the radial average (B.2) is still an open problem in three and higher dimensions.

Bibliography

[1] Asai, T., On a certain function analogous to $\log\eta(z)$, *Nagoya Math. J.* **40** (1970), pp. 193–211.

[2] Bauer, P. J., Zeros of Dirichlet L-series on the critical line, *Acta Arith.* **93**(1) (2000), pp. 37–52.

[3] Beardon, A. F., *The geometry of discrete groups*, vol. 91, Graduate Texts in Mathematics, Springer-Verlag, New York, 1983, pp. xii+337.

[4] Berndt, B. C. and Knopp, M. I., *Hecke's theory of modular forms and Dirichlet series*, vol. 5, Monographs in Number Theory, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008, pp. xii+137.

[5] Booker, A. R., Simple zeros of degree 2 L-functions, (preprint) arXiv:1211.6838, 2012.

[6] Booker, A. R., Uncovering a new L-function, *Notices Amer. Math. Soc.* **55**(9) (2008), pp. 1088–1094.

[7] Booker, A. R. and Strömbergsson, A., Numerical computations with the trace formula and the Selberg eigenvalue conjecture, *J. Reine Angew. Math.* **607** (2007), pp. 113–161.

[8] Bourgain, J., Decoupling, exponential sums and the Riemann zeta function, (preprint) arXiv:1408.5794, 2014.

[9] Bunimovič, L. A., The ergodic properties of certain billiards, *Functional Anal. Appl.* **8** (1974), pp. 268–269.

[10] Chamizo, F., Some applications of large sieve in Riemann surfaces, *Acta Arith.* **77**(4) (1996), pp. 315–337.

[11] Chamizo, F., The large sieve in Riemann surfaces, *Acta Arith.* **77**(4) (1996), pp. 303–313.

[12] Chamizo, F., Topics in Analytic Number Theory, PhD thesis, Universidad Autónoma de Madrid, 1994.

[13] Chandee, V. et al., Simple zeros of primitive Dirichlet L-functions and the asymptotic large sieve, *Q. J. Math.* **65**(1) (2014), pp. 63–87.

[14] Chatzakos, D. and Petridis, Y., The Hyperbolic Lattice Point Problem in Conjugacy Classes, (submitted) arXiv:1504.01307, 2015.

[15] Chazarain, J., Formule de Poisson pour les variétés riemanniennes, *Invent. Math.* **24** (1974), pp. 65–82.

[16] Cherubini, G., PhD thesis, University of Copenhagen, 2016.

[17] Cohen, H., *Number theory. Vol. I. Tools and Diophantine equations*, vol. 239, Graduate Texts in Mathematics, Springer, New York, 2007, pp. xxiv+650.

[18] Cohen, H., *Number theory. Vol. II. Analytic and modern tools*, vol. 240, Graduate Texts in Mathematics, Springer, New York, 2007, pp. xxiv+596.

[19] Coleman, M. D., A zero-free region for the Hecke L-functions, *Mathematika* **37**(2) (1990), pp. 287–304.

[20] Conrey, B., Iwaniec, H., and Soundararajan, K., Critical zeros of Dirichlet L-functions, (preprint) arXiv:1105.1177, 2011.

[21] Conrey, J. B., Ghosh, A., and Gonek, S. M., Simple zeros of the zeta function of a quadratic number field. I, *Invent. Math.* **86**(3) (1986), pp. 563–576.

[22] Davenport, H., *Multiplicative number theory*, Second, vol. 74, Graduate Texts in Mathematics, Revised by Hugh L. Montgomery, New York: Springer-Verlag, 1980, pp. xiii+177.

[23] Duke, W., Rudnick, Z., and Sarnak, P., Density of integer points on affine homogeneous varieties, *Duke Math. J.* **71**(1) (1993), pp. 143–179.

[24] Dyatlov, S., Microlocal limits of Eisenstein functions away from the unitarity axis, *J. Spectr. Theory* **2**(2) (2012), pp. 181–202.

[25] Elstrodt, J., Grunewald, F., and Mennicke, J., *Groups acting on hyperbolic space*, Springer Monographs in Mathematics, Harmonic analysis and number theory, Springer-Verlag, Berlin, 1998, pp. xvi+524.

[26] Eskin, A. and McMullen, C., Mixing, counting, and equidistribution in Lie groups, *Duke Math. J.* **71**(1) (1993), pp. 181–209.

[27] Fujii, A., On the zeros of Dirichlet L-functions. V, *Acta Arith.* **28**(4) (1976), pp. 395–403.

[28] Garunkštis, R. and Kalpokas, J., The discrete mean square of the Dirichlet L-function at nontrivial zeros of another Dirichlet L-function, *Int. J. Number Theory* **9**(4) (2013), pp. 945–963.

[29] Garunkštis, R., Kalpokas, J., and Steuding, J., Sum of the Dirichlet L-functions over nontrivial zeros of another Dirichlet L-function, *Acta Math. Hungar.* **128**(3) (2010), pp. 287–298.

[30] Garunkštis, R. and Steuding, J., On the roots of the equation $\zeta(s) = a$, *Abh. Math. Semin. Univ. Hambg.* **84**(1) (2014), pp. 1–15.

[31] Gelbart, S., An elementary introduction to the Langlands program, *Bull. Amer. Math. Soc. (N.S.)* **10**(2) (1984), pp. 177–219.

[32] Gonek, S. M., An explicit formula of Landau and its applications to the theory of the zeta-function, in *A tribute to Emil Grosswald: number theory and related analysis*, vol. 143, Contemp. Math. Providence, RI: Amer. Math. Soc., 1993, pp. 395–413.

[33] Gonek, S. M., Mean values of the Riemann zeta-function and its derivatives, *Invent. Math.* **75**(1) (1984), pp. 123–141.

[34] Gonek, S. M., Lester, S. J., and Milinovich, M. B., A note on simple a-points of L-functions, *Proc. Amer. Math. Soc.* **140**(12) (2012), pp. 4097–4103.

[35] Good, A., *Local analysis of Selberg's trace formula*, vol. 1040, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1983, pp. i+128.

[36] Gradshteyn, I. S. and Ryzhik, I. M., *Table of integrals, series, and products*, Seventh, Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, Elsevier/Academic Press, Amsterdam, 2007, pp. xlviii+1171.

[37] Griffiths, D., *Introduction to Quantum Mechanics*, Second, Pearson Education International, 2005.

[38] Hardy, G. H. and Littlewood, J. E., Contributions to the theory of the riemann zeta-function and the theory of the distribution of primes, *Acta Math.* **41**(1) (1916), pp. 119–196.

[39] Harish-Chandra, Spherical functions on a semisimple Lie group. I, *Amer. J. Math.* **80** (1958), pp. 241–310.

[40] Harish-Chandra, Spherical functions on a semisimple Lie group. II, *Amer. J. Math.* **80** (1958), pp. 553–613.

[41] Hassell, A., What is quantum unique ergodicity?, *Austral. Math. Soc. Gaz.* **38**(3) (2011), pp. 158–167.

[42] Heath-Brown, D. R., The growth rate of the Dedekind zeta-function on the critical line, *Acta Arith.* **49**(4) (1988), pp. 323–339.

[43] Hecke, E., Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung, *Math. Ann.* **112**(1) (1936), pp. 664–699.

[44] Heitkamp, D., *Hecke-Theorie zur* SL(2; $\mathfrak{o}$), vol. 5, Schriftenreihe des Mathematischen Instituts der Universität Münster, 3. Serie, Universität Münster, Mathematisches Institut, Münster, 1992, pp. iv+143.

[45] Hejhal, D. A., *The Selberg trace formula for* PSL(2, $\mathbf{R}$). *Vol. 2*, vol. 1001, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1983, pp. viii+806.

[46] Herrmann, O., Über die Verteilung der Längen geodätischer Lote in hyperbolischen Raumformen, *Math. Z.* **79** (1962), pp. 323–343.

[47] Hill, R. and Parnovski, L., The variance of the hyperbolic lattice point counting function, *Russ. J. Math. Phys.* **12**(4) (2005), pp. 472–482.

[48] Holowinsky, R. and Soundararajan, K., Mass equidistribution for Hecke eigenforms, *Ann. of Math. (2)* **172**(2) (2010), pp. 1517–1528.

[49] Huber, H., Ein Gitterpunktproblem in der hyperbolischen Ebene, *J. Reine Angew. Math.* **496** (1998), pp. 15–53.

[50] Huber, H., Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I, *Comment. Math. Helv.* **30** (1956), 20–62 (1955).

[51] Ingham, A. E., Mean-Value Theorems in the Theory of the Riemann Zeta-Function, *Proc. London Math. Soc.* **S2-27**(1) (1928), pp. 273–300.

[52] Iwaniec, H., Prime geodesic theorem, *J. Reine Angew. Math.* **349** (1984), pp. 136–159.

[53] Iwaniec, H., *Spectral Methods of Automorphic Forms*, second, American Mathematical Society, Providence, RI, 2002.

[54] Iwaniec, H. and Kowalski, E., *Analytic number theory*, vol. 53, American Mathematical Society Colloquium Publications, American Mathematical Society, Providence, RI, 2004, pp. xii+615.

[55] Jakobson, D., Quantum unique ergodicity for Eisenstein series on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathrm{PSL}_2(\mathbf{R})$, *Ann. Inst. Fourier (Grenoble)* **44**(5) (1994), pp. 1477–1504.

[56] Kalpokas, J., Korolev, M. A., and Steuding, J., Negative values of the Riemann zeta function on the critical line, *Mathematika* **59**(2) (2013), pp. 443–462.

[57] Katok, S., *Fuchsian groups*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992, pp. x+175.

[58] Koyama, S.-y., Quantum ergodicity of Eisenstein series for arithmetic 3-manifolds, *Comm. Math. Phys.* **215**(2) (2000), pp. 477–486.

[59] Laaksonen, N., *http://nikolaaksonen.fi/research*, 2014.

[60] Landau, E., Über die Nullstellen der Zetafunktion, *Math. Ann.* **71**(4) (1912), pp. 548–564.

[61] Landau, E., Über die Wurzeln der Zetafunktion, *Math. Z.* **20**(1) (1924), pp. 98–104.

[62] Lavrik, A. F., An Approximate Functional Equation for the Dirichlet L-function, *Trans. Moscow Math. Soc.* **18** (1968), pp. 101–115.

[63] Lax, P. D. and Phillips, R. S., The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces, in *Functional analysis and approximation (Oberwolfach, 1980)*, vol. 60, Internat. Ser. Numer. Math. Birkhäuser, Basel-Boston, Mass., 1981, pp. 373–383.

[64] Li, X. and Radziwiłł, M., The Riemann Zeta Function on Vertical Arithmetic Progressions, *Int. Math. Res. Not.* **2015**(2) (2015), pp. 325–354.

[65] Lindenstrauss, E., Invariant measures and arithmetic quantum unique ergodicity, *Ann. of Math. (2)* **163**(1) (2006), pp. 165–219.

[66] Luo, W. Z. and Sarnak, P., Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$, *Inst. Hautes Études Sci. Publ. Math.* (81) (1995), pp. 207–237.

[67] Luo, W., Nonvanishing of L-values and the Weyl law, *Ann. of Math. (2)* **154**(2) (2001), pp. 477–502.

[68] Martin, G. and Ng, N., Nonzero values of Dirichlet L-functions in vertical arithmetic progressions, *Int. J. Number Theory* **9**(4) (2013), pp. 813–843.

[69] Mezzadri, F. and Snaith, N. C., eds., *Recent perspectives in random matrix theory and number theory*, vol. 322, London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 2005, pp. x+518.

[70] Mueller, J., On the difference between consecutive zeros of the Riemann zeta function, *J. Number Theory* **14**(3) (1982), pp. 327–331.

[71] Müller, W., Weyl's law in the theory of automorphic forms, in *Groups and analysis*, vol. 354, London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2008, pp. 133–163.

[72] Murty, M. R. and Murty, V. K., Strong multiplicity one for Selberg's class, *C. R. Acad. Sci. Paris Sér. I Math.* **319**(4) (1994), pp. 315–320.

[73] Odlyzko, A., *Tables of zeros of the Riemann zeta function*, http://www.dtc.umn.edu/~odlyzko/zeta_tables/, 2014.

[74] Parkkonen, J. and Paulin, F., Counting arcs in negative curvature, (preprint) arXiv:1203.0175v2, 2013.

[75] Parkkonen, J. and Paulin, F., On the hyperbolic orbital counting problem in conjugacy classes, *Math. Z.* **279**(3-4) (2015), pp. 1175–1196.

[76] Patterson, S. J., *An introduction to the theory of the Riemann Zeta-Function*, vol. 14, Cambridge Studies in Advanced Mathematics, Cambridge: Cambridge University Press, 1988, pp. xiv+156.

[77] Petridis, Y. and Risager, M. S., Local Average in Hyperbolic Lattice Point Counting, (submitted) arXiv:1408.5743, 2015.

[78] Petridis, Y. N., Perturbation of Scattering Poles for Hyperbolic Surfaces and Central Values of L-series, *Duke Math. J.* **103**(1) (2000), pp. 101–130.

[79] Petridis, Y. N., Raulf, N., and Risager, M. S., Erratum to "Quantum limits of Eisenstein series and scattering states" [MR3121690], *Canad. Math. Bull.* **56**(4) (2013), pp. 827–828.

[80] Petridis, Y. N., Raulf, N., and Risager, M. S., Quantum limits of Eisenstein series and scattering states, *Canad. Math. Bull.* **56**(4) (2013), pp. 814–826.

[81] Petridis, Y. N. and Risager, M. S., Dissolving of cusp forms: higher-order Fermi's golden rules, *Mathematika* **59**(2) (2013), pp. 269–301.

[82] Petridis, Y. N. and Sarnak, P., Quantum unique ergodicity for $\mathrm{SL}_2(\mathcal{O})\backslash\mathbf{H}^3$ and estimates for L-functions, *J. Evol. Equ.* **1**(3) (2001), Dedicated to Ralph S. Phillips, pp. 277–290.

[83] Phillips, R. and Sarnak, P., Automorphic spectrum and Fermi's golden rule, *J. Anal. Math.* **59** (1992), Festschrift on the occasion of the 70th birthday of Shmuel Agmon, pp. 179–187.

[84] Phillips, R. S. and Sarnak, P., On cusp forms for co-finite subgroups of $\mathrm{PSL}(2,\mathbf{R})$, *Invent. Math.* **80**(2) (1985), pp. 339–364.

[85] Phillips, R. and Rudnick, Z., The circle problem in the hyperbolic plane, *J. Funct. Anal.* **121**(1) (1994), pp. 78–116.

[86] Raulf, N., Traces of Hecke operators acting on three-dimensional hyperbolic space, *J. Reine Angew. Math.* **591** (2006), pp. 111–148.

[87] Riemann, B., Über die Anzahl der Primzahlen unter einer gegebenen Grosse, *Ges. Math. Werke und Wissenschaftlicher Nachlaß* **2** (1859), pp. 145–155.

[88] Rubinstein, M. and Sarnak, P., Chebyshev's bias, *Experiment. Math.* **3**(3) (1994), pp. 173–197.

[89] Rudnick, Z. and Sarnak, P., The behaviour of eigenstates of arithmetic hyperbolic manifolds, *Comm. Math. Phys.* **161**(1) (1994), pp. 195–213.

[90] Sarnak, P., The arithmetic and geometry of some hyperbolic three-manifolds, *Acta Math.* **151**(3-4) (1983), pp. 253–295.

[91] Sarnak, P., Arithmetic quantum chaos, in *The Schur lectures (1992) (Tel Aviv)*, vol. 8, Israel Math. Conf. Proc. Bar-Ilan Univ., Ramat Gan, 1995, pp. 183–236.

[92] Sarnak, P., Recent progress on the quantum unique ergodicity conjecture, *Bull. Amer. Math. Soc. (N.S.)* **48**(2) (2011), pp. 211–228.

[93] Shnirelman, A. I., Ergodic properties of eigenfunctions, *Uspehi Mat. Nauk* **29**(6(180)) (1974), pp. 181–182.

[94] Soundararajan, K. and Young, M. P., The prime geodesic theorem, *J. Reine Angew. Math.* **676** (2013), pp. 105–120.

[95] Soundararajan, K., Quantum unique ergodicity for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, *Ann. of Math. (2)* **172**(2) (2010), pp. 1529–1538.

[96] Stein, E. M. and Shakarchi, R., *Complex analysis*, Princeton Lectures in Analysis, II, Princeton University Press, Princeton, NJ, 2003, pp. xviii+379.

[97] Steuding, J., *Value-distribution of L-functions*, vol. 1877, Lecture Notes in Mathematics, Berlin: Springer, 2007, pp. xiv+317.

[98] Strömbergsson, A. and Booker, A., http://www2.math.uu.se/research/archive/astrombe/mweff, 2014.

[99] Then, H., http://modular.math.washington.edu/home/then/browse.html, 2014.

[100] Titchmarsh, E. C., *The theory of the Riemann zeta-function*, Second, Edited and with a preface by D. R. Heath-Brown, The Clarendon Press, Oxford University Press, New York, 1986, pp. x+412.

[101] Truelsen, J. L., Quantum unique ergodicity of Eisenstein series on the Hilbert modular group over a totally real field, *Forum Math.* **23**(5) (2011), pp. 891–931.

[102] Tsuzuki, M., Spectral square means for period integrals of wave functions on real hyperbolic spaces, *J. Number Theory* **129**(10) (2009), pp. 2387–2438.

[103] Verdière, Y. Colin de, Ergodicité et fonctions propres du laplacien, *Comm. Math. Phys.* **102**(3) (1985), pp. 497–502.

[104] Young, M. P., The quantum unique ergodicity conjecture for thin sets, (submitted) arXiv:1306.1554, 2013.

[105] Zagier, D., Newman's short proof of the prime number theorem, *Amer. Math. Monthly* **104**(8) (1997), pp. 705–708.

[106] Zelditch, S., Uniform distribution of eigenfunctions on compact hyperbolic surfaces, *Duke Math. J.* **55**(4) (1987), pp. 919–941.