THE PRIME GEODESIC THEOREM

HADI RIHAWI

Abstract. We prove the Wiener-Ikehara Tauberian theorem and use it to prove the prime number theorem. We utilize the Selberg trace formula to prove the prime geodesic theorem, the prime number theorem analog for primitive closed hyperbolic geodesics. We discuss improvements made to the bound, especially that of Iwaniec.

Contents

1. Introduction

In this expository paper, we discuss the tools needed to understand the proofs of the prime number theorem and the prime geodesic thoerem. To that end, we begin in Section 2 by giving a brief catalog of definitions and theorems from complex analysis which are crucial in the proof of the Wiener-Ikehara theorem, the manipulations performed on the functions introduced in Section 4, and the description of geodesics. The content of this section can be found in [4]. In Section 3, we introduce integral transforms on the upper half-plane and define geodesic in the modular group. In Section 5, we give the most important tools towards proving the prime geodesic theorem and Iwaniec's improvement to the bound. In Section 6, we prove the Wiener-Ikehara theorem using complex analysis and use it to directly prove the prime number theorem following the proof given in [7]. Finally, in Section 7, we state and prove the prime geodesic theorem and discuss the various improvements to the bound, with special focus on Iwaniec's proof given in [6].

We introduce some of the notation which will be used throughout the paper below.

Definitions 1.1. When we write

$$f(x) \ll g(x),$$

Date: August 18, 2023.

we mean that there exists some constant C such that $f(x) \leq Cg(x)$.

When we write
$$f(x) = O(g(x)),$$
we mean that there exists some constant C such that $|f(x)| \leq Cg(x)$ as $x \to \infty$.

When we write
$$f(x) \sim g(x),$$
we mean that $f(x)/g(x) \to 1$ as $x \to \infty$.

2. Complex Analysis

We state in this section a few of the definitions and theorems from complex analysis necessary to understand the language and proofs of the upcoming sections. These statements and their proofs can be found in [4].

Definition 2.1. A function
$$f\colon D \to \mathbb{C},\ D \subset \mathbb{C},$$
is *complex differentiable* at $a \in D$ if and only if
$$\lim_{z \to a} \frac{f(z) - f(a)}{z - a}$$
exists. We denote this limit by $f'(a)$ if it exists.

Definition 2.2. A function is *analytic* or *holomorphic* in its domain if it is complex diffentiable at every point in its domain. A function is *entire* if it is complex differentiable at every point in $\mathbb{C}$.

Definition 2.3. Let $D \subset \mathbb{C}$ be open and $f\colon D \to \mathbb{C}$ be an analytic function. a is a *singularity* of f if a is not in D but has the property that, for some $r > 0$, the *punctured disk*
$$\dot{U}_r(a) := \{z \in \mathbb{C} \big| 0 < |z - a| < r\}$$
is contained in D.

Definitions 2.4. Let f and a be as in the previous definition. a is:

- *removable* if and only if f can be analytically extended to $D \cup \{a\}$.
- *non-essential* if and only if there exists an $m \in \mathbb{Z}$ such that a is a removable singularity of $g(z) = (z - a)^m f(z)$.
  - a non-essential singularity which is not removable is called a *pole*.
- *essential* if and only if it is not non-essential.

Theorem 2.5 (Maximal modulus principle)**.** *Let $D \subset \mathbb{C}$ be a domain. Let $f\colon D \to \mathbb{C}$ be an analytic function. If its modulus $|f|\colon D \to \mathbb{R}_+$ reaches the maximum on D, then f is constant.*

In the discussion of Iwaniec's proof, we will make use of the following generalization of the maximal modulus principle. The *Phragmén-Lindelöf principle* is a technique used to prove the boundedness of analytic functions on unbounded domains.

Theorem 2.6 (Phragmén-Lindelöf principle)**.** *Let $D \subset \mathbb{C}$ be a domain with boundary ∂D. Let $f\colon D \to \mathbb{C}$ be an analytic function. If $|f(z)| \leq M$ on ∂D, then $|f(z)| \leq M$ everywhere in D.*

We introduce here the standard Poincaré model of the hyperbolic plane.

$$\mathbb{H} := \{z = x + iy \in \mathbb{C} : y > 0\}, \;\; ds^2 = \frac{dx^2 + dy^2}{y^2}$$

The geodesics on $\mathbb{H}$ are the arcs of circles that intersect $\partial\mathbb{H}$ orthogonally, i.e., semicircles with endpoints on $\partial\mathbb{H}$ and straight lines orthogonal to $\partial\mathbb{H}$. For our purpose, we care about orientation-preserving isometries of $\mathbb{H}$, i.e., the projective special linear group.

Definition 2.7. The *projective special linear group* is given by

$$\mathrm{PSL}_2(\mathbb{R}) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\}$$

Definition 2.8. A *Möbius transformation* or a *homographic transformation* is a rational function giving a bijective map from $\overline{\mathbb{C}} \to \overline{\mathbb{C}}$:

$$z \mapsto \frac{az + b}{cz + d}, \;\; a, b, c, d \in \mathbb{C} \text{ and } ad - bc \neq 0. \tag{2.9}$$

Definition 2.10. The *modular group* Γ is the group of Möbius transformations of the upper half of the complex plane, i.e.,

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

and for a transformation $P \in \Gamma$,

$$z \mapsto Pz := \frac{az + b}{cz + d}.$$

3. Harmonic Analysis on $\mathbb{H}$

We begin this section by classifying elements of Γ. A more detailed account of this work can be found in [3]. We classify the non-identity members in Γ by their fixed points. Finding the fixed points of a transformation requires solving the equation $Pz = z$:

$$Pz = \frac{az + b}{cz + d} = z \implies cz^2 + (d - a)z - b = 0.$$

The value of the discriminant of this quadratic in z will give the number of solutions it has, and thus the number and location of the fixed points.

$$\begin{aligned}(d - a)^2 + 4bc &= d^2 + a^2 - 2ad + 4bc \\ &= d^2 + a^2 + 2ad - 4(ad - bc) \\ &= (a + d)^2 - 4 = \mathrm{tr}(T)^2 - 4.\end{aligned}$$

We thus define the classes of transformations as follows:

Definition 3.1. A transformation $P \in \Gamma$ is:

- *elliptic* if $|\mathrm{tr}(P)| < 2$, i.e., P has one fixed point in $\mathbb{H}$.
- *parabolic* if $|\mathrm{tr}(P)| = 2$, i.e., P has one double-root fixed point in $\partial\mathbb{H}$.
- *hyperbolic* if $|\mathrm{tr}(P)| > 2$, i.e., P has two fixed points in $\partial\mathbb{H}$.

The names of the classes are borrowed from the shape of the conic $ax^2 + bxy + cy^2 = 1$ in terms of its discriminant $d = b^2 - 4ac$. For our purposes in relating the proof of [6], we focus on the hyperbolic transformations. We characterize the hyperbolic conjugacy class to define the norm of a transformation.

Definition 3.2. A hyperbolic transformation $P \in \Gamma$ is *primitive* if it is not a non-trivial power in Γ.

This definition implies that every hyperbolic element of Γ is a power of a unique primitive hyperbolic in Γ.

Proposition 3.3. *Let P be a hyperbolic transformation in Γ. Then P is conjugate to a matrix of the form* $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$.

Proof. We solve for the eigenvalues of P:

$$\begin{aligned} \det(P - xI) &= \det \begin{bmatrix} a-x & b \\ c & d-x \end{bmatrix} \\ &= (a-x)(d-x) - bc = ad - ax - dx + x^2 - bc \\ &= x^2 - (a+d)x + (ad-bc) = x^2 - \operatorname{tr}(P)x + 1 \\ x^2 - \operatorname{tr}(P)x + 1 = 0 \end{aligned}$$

Since $|\operatorname{tr}(P)| > 2$, this polynomial has two distinct real roots, q and r. By Vieta's formulas, we have that

$$\operatorname{tr}(P) = q + r \quad \text{and} \quad q = \frac{1}{r}$$

Let $q = \lambda$ so that $r = 1/\lambda$ and $\operatorname{tr}(P) = \lambda + \lambda^{-1}$. Then λ and $1/\lambda$ are the eigenvalues of P. Thus, $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$ for $|\lambda| > 1$ is the representative for the conjugacy class of hyperbolic transformations P in Γ. $\square$

The conjugate matrix found above acts as multiplication by λ^2. We use this to define the norm of the hyperbolic transformation.

Definition 3.4. The *norm* of a hyperbolic element P is given by

$$NP := \lambda^2$$

with λ as defined in 3.3.

Notice that we must have that $\operatorname{tr}(P) = \lambda + 1/\lambda$ is an integer $n > 2$. Hence, NP also has the form

$$\begin{aligned} NP &= \left(\frac{n + \sqrt{n^2 - 4}}{2} \right)^2 = \frac{1}{4} \left(n^2 + n^2 - 4 + 2n\sqrt{n^2 - 4} \right) \\ &= \frac{1}{2}n^2 - 1 + \frac{1}{2}n \sum_{k=0}^{\infty} \frac{\Gamma(\frac{1}{2}+1)}{k!\Gamma(\frac{1}{2}+1-k)} (-4)^k (n^2)^{\frac{1}{2}-k} \\ &= \frac{1}{2}n^2 - 1 + \frac{1}{2}n \left(n - \frac{2}{n} + O(n^{-3}) \right) = n^2 - 2 + O(n^{-2}). \end{aligned}$$

We can also show that primitive hyperbolic conjugacy classes correspond to equivalence classes of primitive indefinite binary quadratic forms as demonstrated by Sarnak in [10]. This correspondence provides another way of interpreting the norms NP. Given a primitive ($\gcd(a,b,c) = 1$) binary quadratic form $ax^2 + bxy + cy^2$ with discriminant $d = b^2 - 4ac > 0$, let $h(d)$ denote the number of inequivalent classes with discriminant d and let $\epsilon_d = \frac{1}{2}\left(n + \sqrt{d}m\right) > 1$ be the smallest solution of Pell's equation $n^2 - dm^2 = 4$. Then the norms of primitive classes are ϵ_d^2 with multiplicity $h(d)$. In fact, there is a one-to-one correspondence between the classes of forms and hyperbolic elements.

We define below a few integral transforms that we will encounter as we build up to the prime number theorem and the prime geodesic theorem.

Definitions 3.5. The *convolution* of two functions f and g is given by

$$(f * g)(x) := \int_{-\infty}^{\infty} f(y)g(x-y)dy. \tag{3.6}$$

The *Fourier transform* of a function f is given by

$$\hat{f}(\xi) := \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx. \tag{3.7}$$

The *Laplace transform* of a function f is given by

$$\mathcal{L}f(s) := \int_{0}^{\infty} f(t)e^{-st}dt. \tag{3.8}$$

The *Mellin transform*, often regarded as the multiplicative version of the Laplace transform, of a function f is given by

$$\mathcal{M}f(s) := \int_{0}^{\infty} f(x)x^{s-1}dx. \tag{3.9}$$

And finally, we define the hyperbolic Laplacian, which is closely related to the zeros of the Selberg zeta function (Definition 4.16).

Definition 3.10. The *hyperbolic Laplace operator* or *hyperbolic Laplacian* on $\mathbb{H}$ is given by

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

The Laplacian has eigenvalues

$$\lambda_j = \frac{1}{4} + t_j^2 = z_j(1-z_j),\ z_j = \frac{1}{2} + it_j.$$

4. Special Functions

Definition 4.1. The *logarithmic integral function* is given by

$$\mathrm{li}(x) := \int_{2}^{\infty} \frac{dt}{\log t} \tag{4.2}$$

for $x \geq 2$.

Lemma 4.3.

$$\frac{x}{\log x} \sim \mathrm{li}(x).$$

Proof. Integrating $\mathrm{li}(x)$ by parts, we get

$$\begin{aligned}\mathrm{li}(x) &= \frac{x}{\log x} + \frac{2}{\log 2} + \int_2^x \frac{dt}{(\log t)^2} \\ &= \frac{x}{\log x} + \int_e^x \frac{dt}{(\log t)^2} + \left(\frac{2}{\log 2} + \int_2^e \frac{dt}{(\log t)^2}\right).\end{aligned}$$

Since $1/(\log t)^2$ is decreasing, we have that

$$\begin{aligned}\int_e^x \frac{dt}{(\log t)^2} &= \int_e^{\sqrt{x}} \frac{dt}{(\log t)^2} + \int_{\sqrt{x}}^x \frac{dt}{(\log t)^2} \\ &\leq \sqrt{x} + x\left(\frac{1}{\log \sqrt{x}}\right)^2 = \sqrt{x} + \frac{4x}{(\log x)^2}\end{aligned}$$

Thus

$$\frac{\log x}{x}\int_e^x \frac{dt}{(\log t)^2} \leq \frac{\log x}{\sqrt{x}} + \frac{4}{\log x}$$

so that

$$\frac{\log x}{x}\int_e^x \frac{dt}{(\log t)^2} \to 0 \text{ as } x \to \infty.$$

Hence

$$\frac{\log x}{x}\mathrm{li}(x) = 1 \text{ as } x \to \infty.$$

□

Definition 4.4. A *Dirichlet series* is any series of the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^z},$$

where $z \in \mathbb{C}$ and $\{a_n\} \subset \mathbb{C}$.

Definition 4.5. The *Riemann zeta function* is given by

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

for $\Re(z) > 1$.

Zeta functions are a special class of generating functions which are defined in terms of a countable collection of numbers. These functions can keep track of these values, and the study of these functions allows one to extract information about their associated values. The Riemann zeta function carries information related to the distribution of prime numbers.

Proposition 4.6. *$\zeta(z)$ is absolutely convergent for $\Re(z) > 1$.*

Proof. Let $z = x + iy$. For all $\delta > 0$ we have

$$\left|\frac{1}{n^z}\right| = \frac{1}{n^x} \leq \frac{1}{n^{1+\delta}}$$

□

Corollary 4.7 (Euler product formula)**.**

$$\zeta(z) = \prod_{p \text{ prime}} \left(1 - p^{-z}\right)^{-1} \tag{4.8}$$

Proof. $\zeta(z)$ is absolutely convergent for $\Re(z) > 1$ by the previous Proposition.

$$\begin{aligned}\zeta(z) &= 1 + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \cdots \\ \frac{1}{2^z}\zeta(z) &= \frac{1}{2^z} + \frac{1}{4^z} + \frac{1}{6^z} + \cdots\end{aligned}$$

Subtracting the second equation from the first removes all terms containing a factor of 2.

$$\left(1 - \frac{1}{2^z}\right)\zeta(z) = 1 + \frac{1}{3^z} + \frac{1}{5^z} + \frac{1}{7^z} + \cdots$$
$$\frac{1}{3^z}\left(1 - \frac{1}{2^z}\right)\zeta(z) = \frac{1}{3^z} + \frac{1}{9^z} + \frac{1}{15^z} + \cdots$$

Subtracting the second equation from the first removes all terms containing a factor of 3. Since $\zeta(z)$ converges absolutely, we can repeat this process for every prime to get

$$\zeta(z) \prod_{p \text{ prime}} (1 - p^{-z}) = 1$$

□

Definition 4.9. The *prime counting function* is given by

$$\pi(x) \coloneqq \#\{p \text{ prime} \colon p \le x\} = \sum_{p \le x} 1$$

Definition 4.10. The *von Mangoldt function* is given by

$$\Lambda(n) \coloneqq \begin{cases} \log p & \text{if } n = p^m \text{ for some } m \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

The following special functions closely relate to the prime counting function π.

Definition 4.11. The *Chebyshev psi function* is given by

$$\psi(x) \coloneqq \sum_{n \le x} \Lambda(n)$$

or equivalently,

$$\psi(x) = \sum_{p \le x} \left[\frac{\log x}{\log p}\right] \log p.$$

Definition 4.12. The *Chebyshev theta function* is given by

$$\vartheta(x) \coloneqq \sum_{p \le x} \log p.$$

Theorem 4.13.

$$-\frac{\zeta'(z)}{\zeta(z)} = z \int_1^\infty \psi(t) t^{-z-1} dt \tag{4.14}$$

for $\Re(z) > 1$.

Proof. Let p and q be primes. By Corollary 4.7,

$$\begin{aligned}\zeta'(z) &= \frac{d}{dz}\left(\prod_p (1-p^{-z})^{-1}\right)\\ &= \sum_p \frac{-p^{-z}\log p}{(1-p^{-z})^2}\prod_{q\neq p}(1-q^{-z})^{-1}\\ &= \sum_p \frac{-p^{-z}\log p}{(1-p^{-z})^2}\zeta(z)(1-p^{-z})\\ &= \zeta(z)\sum_p \frac{-p^{-z}\log p}{1-p^{-z}}.\end{aligned}$$

Thus

$$\begin{aligned}-\frac{\zeta'(z)}{\zeta(z)} &= \sum_p \frac{p^{-z}}{1-p^{-z}}\log p = \sum_p\sum_{n=1}^{\infty} p^{-nz}\log p\\ &= \sum_k k^{-z}\Lambda(k) = \sum_{k=1}^{\infty} k^{-z}(\psi(k)-\psi(k-1)) \qquad (4.15)\end{aligned}$$

where $k = p^n$ for some n. By partial summation, taking $a_k = k^{-z}$ and $b_{k+1} = \psi(k)$,

$$\sum_{k=1}^{M} k^{-z}(\psi(k)-\psi(k-1)) = (M+1)^{-z}\psi(M) + \sum_{k=1}^{M}\psi(k)(k^{-z}-(k+1)^{-z}).$$

By Definition 4.11, it is clear that $\psi(x) \le x\log x$. So for $\Re(z) > 1$, $\psi(M)(M+1)^{-z} \le M\log M(M+1)^{-z} \to 0$ as $M\to\infty$. Thus, we have that

$$\begin{aligned}-\frac{\zeta'(z)}{\zeta(z)} &= \sum_{k=1}^{\infty}\psi(k)(k^{-z}-(k+1)^{-z})\\ &= \sum_{k=1}^{\infty}\psi(k)z\int_k^{k+1} t^{-z-1}dt\\ &= \sum_{k=1}^{\infty} z\int_k^{k+1}\psi(t)t^{-z-1}dt\\ &= z\int_1^{\infty}\psi(t)t^{-z-1}dt.\end{aligned}$$

□

Definition 4.16. The *Selberg zeta function* is given by

$$Z(z) \coloneqq \prod_{\{P_0\}}\prod_{n=0}^{\infty}(1-(NP_0)^{-z-n}) \qquad (4.17)$$

for $\Re(z) > 1$, where $\{P_0\}$ is the set of all primitive hyperbolic classes of conjugate elements in Γ and NP_0 is the norm of P_0.

The *Riemann hypothesis* is the conjecture that the nontrivial zeros of the Riemann zeta function all lie on the line of complex numbers with real part $1/2$. Unlike the Riemann zeta function, the Riemann hypothesis equivalent for the Selberg zeta function is known to be true. Thus, the zeros and singularities of this function are well-characterized:

i) a simple zero at $z = 1$,
ii) nontrivial zeros at $z = z_j = \frac{1}{2} \pm it_j$ where $z_j(1 - z_j)$ are eigenvalues of the hyperbolic Laplacian Δ corresponding to cusp forms,
iii) nontrivial zeros at $z = \frac{1}{2}\rho_j$ where ρ_j are complex zeros of the Riemann zeta function,
iv) simple poles at $s = \frac{1}{2} - n$, for $n \in \mathbb{N}$.

Finally, we introduce the Bessel functions, and define a related transform and prove an asymptotic for them that appeared in [6]. These functions relate to Fourier transforms of powers of quadratic functions.

Definition 4.18. The *Bessel functions* are generalizations of the trigonometric function, and they are canonical solutions $y(x)$ of

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - \alpha^2) y = 0$$

for a complex number α. The Bessel functions are given by

$$J_\alpha(z) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m + \alpha + 1)} \left(\frac{z}{2}\right)^{2m+\alpha}.$$

For real t, the *Bessel-Kuznetsov transform* of a function f is given by

$$\tilde{f}(t) = \frac{\pi i}{2 \sinh \pi t} \int_0^\infty (J_{2it}(x) - J_{-2it}(x)) f(x) \frac{dx}{x}. \tag{4.19}$$

And for integer k, the transform is given by

$$\tilde{f}(k) = \int_0^\infty J_k(x) f(x) \frac{dx}{x}.$$

Proposition 4.20. *When α is not a negative integer,*

$$J_\alpha(z) = \frac{1}{\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha \left(1 + O\left(\frac{|z|^2}{|\alpha| + 2}\right)\right).$$

as $z \to 0$.

Proof.

$$\begin{aligned}
J_\alpha(z) &= \frac{1}{\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha - \frac{1}{\Gamma(\alpha + 2)} \left(\frac{z}{2}\right)^{2+\alpha} + \sum_{m=2}^{\infty} \frac{(-1)^m}{m!\Gamma(m + \alpha + 1)} \left(\frac{z}{2}\right)^{2m+\alpha} \\
&= \frac{1}{\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha - \frac{z^2}{4(\alpha + 1)\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha + \sum_{m=2}^{\infty} \frac{(-1)^m}{m!\Gamma(m + \alpha + 1)} \left(\frac{z}{2}\right)^{2m+\alpha} \\
&= \frac{1}{\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha + \frac{1}{\Gamma(\alpha + 1)} \left(\frac{z}{2}\right)^\alpha O\left(\frac{|z|^2}{|\alpha| + 2}\right).
\end{aligned}$$

□

5. Trace and Summation Formulas

The main tool in the proof of the prime geodesic theorem is the Selberg trace formula applied to compact hyperbolic surfaces, or the space $\Gamma\backslash\mathbb{H}$. We simply provide the formula here without proof. See [12] and [13] for derivations and proofs of the formula. The Selberg trace formula provides an infinite dimensional representation homomorphism of the group to matrices.

Theorem 5.1 (Selberg trace formula)**.** *Suppose that there exists $\delta > 0$ such that*

$$\begin{aligned} &h(r) \text{ is analytic on } |\Im(r)| \leq \frac{1}{2} + \delta, \\ &h(-r) = h(r), \\ &|h(r)| \ll (1 + |\Re(r)|)^{-2-\delta}. \end{aligned} \tag{5.2}$$

Let g be the Fourier transform of h such that

$$h(r) = \int_{-\infty}^{\infty} g(\xi) e^{ir\xi} d\xi$$

Then

$$\sum_{n=0}^{\infty} h(r_n) = \frac{\mu(\Gamma\backslash\mathbb{H})}{4\pi} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr + \sum_{\{P\}} \frac{\log NP_0}{NP^{\frac{1}{2}} - NP^{-\frac{1}{2}}} g(\log NP),$$

where r_n correspond to the eigenvalues $1/4 + r_n^2$ of Δ and the right hand side is a sum over conjugacy classes of Γ. The integral corresponds to identity elements and the sum corresponds to the hyperbolic conjugacy classes $\{P\}$. NP gives the norm of the element P.

We now introduce Kloosterman sums and their associated trace formula which connects them to the spectral theory of automorphic forms and the prime geodesic theorem as will be seen in Iwaniec's proof [6].

Definition 5.3. A *Kloosterman sum* is given by

$$S(m, n, c) := \sum_{\substack{x \leq m \\ \gcd(x,c)=1}} e^{\frac{2\pi i}{c}(mx + n\overline{x})},$$

where $x\overline{x} \equiv 1 \pmod{c}$.

Definition 5.4. A general *Kuznetsov trace formula* for a function f is given by

$$\sum_{c \equiv 0 \bmod N} c^{-r} S(m, n, c) f\left(\frac{4\pi\sqrt{mn}}{c}\right) = \text{Integral transform } + \text{ Spectral terms.}$$

Iwaniec applies the Kuznetsov trace formula to bound

$$\sum_{c} \frac{1}{c} S(n, n, c) f\left(\frac{4\pi n}{c}\right). \tag{5.5}$$

The details of the Kuznetsov trace formula can be found in [8].

6. The Prime Number Theorem

We begin by stating and proving a specific case of a Tauberian theorem of Wiener and Ikehara. The general Wiener-Ikehara theorem omits (6.4). Both proofs following Newman's method can be found in [7].

Theorem 6.1 (Wiener-Ikehara theorem)**.** *Let the Dirichlet series*

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n}{n^z}, \text{ with } a_n \geq 0 \tag{6.2}$$

converge for $\Re(z) > 1$*.* $f(z)$ *is thus analytic for* $\Re(z) > 1$*. Suppose that there is a constant* A *such that*

$$g(z) := f(z) - \frac{A}{z-1} \tag{6.3}$$

has an analytic or continuous extension for $\Re(z) \geq 1$*. Suppose that there is a constant* C *such that*

$$s_n = \sum_{k \leq n} a_k \leq Cn \text{ for all } n. \tag{6.4}$$

Then

$$s_n \sim An \text{ as } n \to \infty. \tag{6.5}$$

Proof. Let the conditions of Theorem 6.1 be satisfied. Define

$$s(v) = \sum_{k \leq v} a_k$$

so that $s(v)$ is nondecreasing, $s(v) = s_n$ for $n \leq v < n+1$ and $s(v) = 0$ for $v < 1$. Applying partial summation to (6.2) shows that for $\Re(z) > 1$,

$$\begin{aligned} f(z) &= \sum_{n=1}^{\infty} \frac{s_n - s_{n-1}}{n^z} = \sum_{n=1}^{\infty} s_n \left(\frac{1}{n^z} - \frac{1}{(n+1)^z} \right) \\ &= \sum_{n=1}^{\infty} s_n z \int_n^{n+1} v^{-z-1} dv = z \int_1^{\infty} s(v) v^{-z-1} dv. \end{aligned} \tag{6.6}$$

Now, by (6.3), write

$$\begin{aligned} g(z) - A &= f(z) - \frac{A}{z-1} - A = f(z) - \frac{Az}{z-1} \\ &= z \int_1^{\infty} \left(\frac{s(v)}{v} - A \right) v^{-z} dv \end{aligned} \tag{6.7}$$

We can now substitute $v = e^t$ so that

$$\frac{s(v)}{v} - A = e^{-t} s(e^t) - A =: \rho(t) \tag{6.8}$$

and we set $\rho(t) = 0$ for $t < 0$. By (6.4), ρ is bounded above by $C - A$. A function ρ is *slowly decreasing* if it satisfies a relation

$$\rho(t) - \rho(u) \geq -\eta(t, u), \text{ where } \eta(t, u) \to 0 \tag{6.9}$$

as $u \to \infty$ and $0 < t-u \to 0$. Notice that this definition is very loose; an increasing f is also slowly decreasing. For $0 \le u < t$,

$$\begin{aligned}
\rho(t) - \rho(u) &= e^{-t}s(e^t) - e^{-u}s(e^u) \ge \left(e^{-t} - e^{-u}\right) s(e^u) \\
&= -\left(1 - e^{-(t-u)}\right) e^{-u}s(e^u) = -\left(1 - e^{-(t-u)}\right)(\rho(u) + A) \\
&\ge -(t-u)\,(\rho(u) + A) \qquad (6.10)
\end{aligned}$$

Since the expression in (6.9) goes to 0 as $t-u$ goes to 0, ρ is slowly decreasing. Now consider the Laplace transform of ρ:

$$\begin{aligned}
\mathcal{L}\rho(z) &= \int_0^\infty \rho(t)e^{-zt}dt = \int_1^\infty \left(\frac{s(v)}{v} - A\right) v^{-z-1}dv \\
&= \frac{g(z+1)}{z+1} - \frac{A}{z+1}. \qquad (6.11)
\end{aligned}$$

By the conditions of Theorem 6.1, $\mathcal{L}\rho(z)$ is analytic for $\Re(z) > 0$ and has an analytic extension for $\Re(z) \ge 0$. It is left to prove that $\rho(t) \to 0$ as $t \to \infty$. To that end, we state and prove, as a lemma, a general theorem on the Laplace transforms of bounded functions.

Lemma 6.12. *Let $\rho(t) = 0$ for $t < 0$ and $|\rho(t)| \le M < \infty$ for $t \ge 0$. Then the Laplace transform*

$$G(z) = \mathcal{L}\rho(z) = \int_0^\infty \rho(t)e^{-zt}dt,\ z = x + iy, \qquad (6.13)$$

defines an analytic function for $x > 0$. Suppose that for $|y| \le R$, the function $G_x(iy) = G(x+iy)$ converges uniformly to a limit function $G(iy)$ as $x \searrow 0$. Then for every positive T and δ,

$$\left|\int_T^{T+\delta} \rho(t)dt\right| \le \frac{4M}{R} + \frac{1}{2\pi}\left|\int_{-R}^{R} G(iy)\frac{e^{i\delta y} - 1}{y}\left(1 - \frac{y^2}{R^2}\right)e^{iTy}dy\right|. \qquad (6.14)$$

If R may be taken arbitrarily large, and the function ρ is slowly decreasing, then

$$\rho(T) \to 0 \ \textit{as}\ T \to \infty. \qquad (6.15)$$

Proof. The proof of (6.14) will be omitted here; a proof of it is given in [7] using Cauchy's theorem and the residue theorem. Using this, we prove (6.15).
For fixed δ and R, the last integral in (6.14) goes to 0 as T goes to ∞ since $G(z)$ is analytic on $[-iR, iR]$ and integration by parts yields

$$e^{iTy}dy = \frac{1}{iT}d(e^{iTy})$$

We thus conclude that

$$\limsup_{T\to\infty}\left|\int_T^{T+\delta}\rho(t)dt\right| \le \frac{4M}{R} \qquad (6.16)$$

and assuming that R may be taken arbitrarily large, for every $\delta > 0$ we have

$$\int_T^{T+\delta}\rho(t)dt \to 0 \ \text{ as } T \to \infty. \qquad (6.17)$$

Now, suppose also that $\rho(t)$ is slowly decreasing as defined in (6.9). Then taking

$$\int_T^{T+\delta} (\rho(t) - \rho(T))dt \geq -\int_T^{T+\delta} \eta(t,T)dt$$

and (6.17) shows that, for a given $\epsilon > 0$,

$$\limsup_{T\to\infty} \rho(T) \leq \epsilon.$$

Applying a similar method to $\int_{T-\delta}^{T} \rho(t)dt$ gives the inequality in the other direction. □

Since $\rho(t)$ satisfies the conditions of the lemma above, $\rho(t) \to 0$ as $t \to \infty$. □

We now present a theorem expressing the relationship between ψ and π, namely that the asymptotic of ψ to x is necessary and sufficient for the asymptotic of the prime counting function π to $x/\log x$.

Theorem 6.18. *As $x \to \infty$,*

$$\pi(x) \to \frac{x}{\log x} \quad \Longleftrightarrow \quad \psi(x) \to x.$$

Proof. By Definition 4.11,

$$\begin{aligned}\psi(x) &= \sum_{p\leq x} \left[\frac{\log x}{\log p}\right] \log p \leq \sum_{p\leq x} \frac{\log x}{\log p} \log p \\ &= \log x \sum_{p\leq x} 1 = \pi(x)\log x \qquad (6.19)\end{aligned}$$

If $1 < y < x$, we have that

$$\begin{aligned}\pi(x) = \pi(y) + \sum_{y<p\leq x} 1 &\leq \pi(y) + \sum_{y<p\leq x} \frac{\log p}{\log y} \\ &\leq y + \frac{1}{\log y} \sum_{y<p\leq x} \log p \leq y + \frac{1}{\log y}\psi(x).\end{aligned}$$

Taking $y = x/(\log x)^2$, we get

$$\pi(x) \leq \frac{x}{(\log x)^2} + \frac{1}{\log x - 2\log\log x}\psi(x). \qquad (6.20)$$

Thus, combining (6.19) and (6.20), we have

$$\frac{\psi(x)}{x} \leq \frac{\log x}{x}\pi(x) \leq \frac{1}{\log x} + \frac{\log x}{\log x - 2\log\log x}\frac{\psi(x)}{x}.$$

As $x \to \infty$,

$$\frac{1}{\log x} \to 0 \quad \text{and} \quad \frac{\log x}{\log x - 2\log\log x} \to 1,$$

so we are done. □

Theorem 6.21 (Prime number theorem)**.**

$$\pi(x) \sim \frac{x}{\log x}.$$

Proof. Consider the Dirichlet series given in (4.15):

$$f(z) := -\frac{\zeta'(z)}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^z}.$$

Since $\zeta(z) \neq 0$ for $\Re(z) = 1$, $f(z)$ is analytic there except for $z = 1$. Since $\zeta(z)$ behaves like $1/(z-1)$ near the point $z = 1$, $f(z)$ also behaves in the same manner. Thus, we have that

$$g(z) := f(z) - \frac{1}{z-1}$$

is analytic for $\Re(z) \geq 1$. $a_n = \Lambda(n)$, f, and g satisfy the conditions of Theorem 6.1 with $A = 1$. Using Chebyshev's estimate, that $\pi(n) \log n \leq Cn$, the partial sums $s_n = \psi(n)$ are bounded by Cn. Hence, Theorem 6.1 implies $\psi(x) \to x$. Thus, the prime number theorem holds by Theorem 6.18. □

The use of Chebyshev's estimate can be avoided if the general Wiener-Ikehara theorem is proven instead.

While $x/\log x$ has been shown to be a good approximation for the number of primes less than or equal to x, numerical calculations have shown that the logarithmic integral function $\mathrm{li}(x)$ is a better approximation. We state that here as a corollary.

Corollary 6.22.

$$\pi(x) \sim \mathrm{li}(x).$$

Proof. Combine Theorem 6.21 and Lemma 4.3. □

7. The Prime Geodesic Theorem

We focus our efforts on primitive closed geodesics, or closed curves that trace out their image exactly once. These geodesics are also called *prime geodesics* because they asymptotically obey a distribution law similar to that found for prime numbers, as seen in Theorem 6.21. We first define the prime geodesic counting function.

Definition 7.1. Let P_0 be a primitive element in Γ. Then

$$\pi_\Gamma(x) := \#\{\{P_0\} : NP_0 \leq x\} = \sum_{N\{P_0\} \leq x} 1$$

gives the number of primitives in Γ with norm less than or equal to x.

We also define the Chebyshev function (as given in Definitions 4.11 and 4.12) equivalents for geodesics.

Definition 7.2.

$$\vartheta_\Gamma(x) = \sum_{N\{P_0\} \leq x} \log NP_0$$

$$\psi_\Gamma(x) = \sum_{N\{P\} \leq x} \Lambda P$$

where $\Lambda P = \log NP_0$ if $\{P\}$ is a power of the primitive hyperbolic class $\{P_0\}$, similar to the von Mangoldt function.

Theorem 6.18 also holds if we take π to be π_Γ and ψ to be ψ_Γ. We now move on to the main theorem of the paper. We provide the proof given in [12].

Theorem 7.3 (Prime geodesic theorem)**.**

$$\pi_\Gamma(x) = \mathrm{li}(x) + O\left(\frac{x^{3/4}}{\log x}\right) \tag{7.4}$$

Proof. Given a $x > 0$, define h_x to be the function with Fourier transform

$$\hat{h}_x(r) = 2\cosh\left(\frac{r}{2}\right)\chi_{[-\log x, \log x]}(r)$$

Let φ be a mollifier (a smooth bump function) supported in $[-1,1]$ such that $\int_{-\infty}^{\infty}\varphi(r)dr = 1$. Given $\epsilon > 0$, let

$$\varphi_\epsilon(r) = \frac{1}{\epsilon}\varphi\left(\frac{r}{\epsilon}\right)$$

which is supported in $[-\epsilon, \epsilon]$ and satisfies $\int_{-\infty}^{\infty}\varphi_\epsilon(r)dr = 1$. Notice that h_x does not satisfy all of the conditions necessary to apply the Selberg trace formula given in (5.4). Define $h_{x,\epsilon}$ to be such that $\hat{h}_{x,\epsilon} = \hat{h}_x * \varphi_\epsilon$ Then,

$$\begin{aligned} h_x(t) &= \int_{-\log x}^{\log x}\left(e^{\frac{\xi}{2}} + e^{-\frac{\xi}{2}}\right)e^{it\xi}d\xi = \int_{-\log x}^{\log x}\left(e^{\xi(\frac{1}{2}+it)} + e^{-\xi(\frac{1}{2}-it)}\right)d\xi \\ &= \frac{x^z - x^{-z}}{z} + \frac{x^{1-z} - x^{-(1-z)}}{1-z} \\ h_{x,\epsilon}(t) &= h(t)\hat{\varphi}_\epsilon(t) \end{aligned}$$

where $z = 1/2 + it$. The Laplacian Δ is a symmetric, non-negative operator, so its eigenvalues $\lambda = 1/4 + t^2 = z(1-z)$ are real and non-negative. Then, it must be the case that either $1/2 < z \le 1$ or $\Re(z) = 1/2$. Since $\hat{\varphi}(t) = 1 + O(t)$, for $1/2 < z \le 1$,

$$h_x(t) = \frac{x^z}{z} + O(\sqrt{x}) \quad \text{and} \quad h_{x,\epsilon}(t) = \frac{x^z}{z} + O(\sqrt{x} + \epsilon x). \tag{7.5}$$

And since φ is smooth, $|\hat{\varphi}(t)| \ll 1/(1+|t|)^2$, so for $\Re(z) = 1/2$,

$$|h_x(t)| \ll \frac{\sqrt{x}}{1+|t|} \quad \text{and} \quad |h_{x,\epsilon}(t)| \ll \frac{\sqrt{x}}{(1+|t|)(1+\epsilon|t|)^2}.$$

By the latter estimate,

$$\int_0^\infty t|h_{x,\epsilon}(t)|dt \ll \epsilon^{-1}\sqrt{x}. \tag{7.6}$$

We now apply Theorem 5.1 to the function $h_{x,\epsilon}$ and the estimates in (7.5) and (7.6) to get

$$\begin{aligned} \sum_{n=0}^{\infty} h_{x,\epsilon}(t_n) &= \frac{\mu(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{\infty} r h_{x,\epsilon}(r)\tanh(\pi r)dr + \sum_{\{P\}}\frac{\hat{h}_{x,\epsilon}(\log NP)\log NP_0}{NP^{1/2} - NP^{-1/2}} \\ &= O\left(\int_{-\infty}^{\infty} r|h_{x,\epsilon}(r)|dr\right) + x + O(\epsilon x + \epsilon^{-1}\sqrt{x}) \\ &= O(\epsilon^{-1}\sqrt{x}) + x + O(\epsilon x + \epsilon^{-1}\sqrt{x}) = x + O(\epsilon x + \epsilon^{-1}\sqrt{x}). \end{aligned}$$

More specifically, we have that

$$H_\epsilon(x) := \sum_{\{P\}}\frac{\hat{h}_{x,\epsilon}(\log NP)\log NP_0}{NP^{1/2} - NP^{-1/2}} = x + O(\epsilon x + \epsilon^{-1}\sqrt{x}). \tag{7.7}$$

We now let

$$H(x) = \sum_{\{P\}} \frac{\hat{h}_x(\log NP) \log NP_0}{NP^{1/2} - NP^{-1/2}}$$

and would like to show that $H(x) = x + O(x^{3/4})$. Notice that

$$\hat{h}_{xe^{-\epsilon},\epsilon}(r) \leq \hat{h}_x(r+\epsilon) \leq e^{\epsilon/2}\hat{h}_x(r) \text{whenever } r \geq 0$$
$$\hat{h}_{xe^{\epsilon},\epsilon}(r) \leq \hat{h}_x(r-\epsilon) \leq e^{-\epsilon/2}\hat{h}_x(r) \text{whenever } r \geq \epsilon$$

For sufficiently small ϵ, we have that

$$e^{-\epsilon/2} H_\epsilon(xe^{-\epsilon}) \leq H(x) \leq e^{\epsilon/2} H_\epsilon(xe^{\epsilon}).$$

Letting $\epsilon = x^{-1/4}$ in (7.7), we get $H(x) = x + O(x^{3/4})$. Now note that

$$H(x) = \psi_\Gamma(x) + O\left(\sum_{N\{P\} \leq x} \frac{\Lambda P}{NP}\right)$$

since finitely many $N\{P\}$ are less than a fixed constant and $H(x) \to \infty$. Hence, $\psi_\Gamma(x) \sim x$, yielding

$$\sum_{N\{P\} \leq x} \frac{\Lambda P}{NP} = \int_0^x \frac{1}{y} d\psi_\Gamma(y) = O(\log x),$$

and thus $\psi_\Gamma(x) = x + O(x^{3/4})$. By definition of ψ_γ and ϑ_Γ,

$$\psi_\Gamma(x) = \sum_{k=1}^{\lfloor \log x/\delta \rfloor} \vartheta(x^{1/k})$$

with δ the length of the shortest closed geodesic on $\Gamma\backslash\mathbb{H}$, so $\vartheta = x + O(x^{3/4})$ as well. Considering

$$\pi_\Gamma(x) = \sum_{N\{P_0\} \leq x} 1 = \int_{e^{\delta/2}}^x \frac{d\vartheta(y)}{\log y},$$

$$\pi_\Gamma(x) = \mathrm{li}(x) + O\left(\frac{x^{3/4}}{\log x}\right)$$

follows. □

At the time of Iwaniec's proof, many mathematicians (such as Huber, Hejhal, Venkov, and Kuznetsov) had studied the error term in the prime geodesic theorem. The results found were of the form

$$\pi_\Gamma(x) = \mathrm{li}(x) + O(x^{\frac{3}{4}} (\log x)^\alpha).$$

Since the Riemann hypothesis holds for $Z(z)$, the error term should be reducible down to $O(x^{1/2+\epsilon})$. But since $Z(z)$ has many more zeros than $\zeta(z)$, the immediate consequence of the Selberg trace formula is the error term $O(x^{3/4+\epsilon})$, as seen above. Iwaniec remarks that the error term can be refined to $O(x^{2/3+\epsilon})$ if the Generalized Lindelöf Hypothesis for Dirichlet L-functions holds.

In [6], Iwaniec becomes the first to break the 3/4 barrier, proving that

$$\pi_\Gamma(x) = \mathrm{li}(x) + O(x^{\frac{35}{48}+\epsilon}). \tag{7.8}$$

In order to get to this conclusion, Iwaniec needs to establish that, for $T \geq 2$ and $X \geq 47$,

$$\sum_{|t_j| \leq T} X^{it_j} \ll T X^{11/48+\epsilon}, \tag{7.9}$$

where t_j is as found in the eigenvalues $\frac{1}{4} + t_j^2$ of the Laplacian. Taking $T = X^{13/48}$, (7.8) immediately follows for $x \geq 2$. To prove (7.9), Iwaniec gives a smooth version. For $X \geq 2$ and $T \geq 2$,

$$\sum_j t_j^{-1} e^{-|t_j|/T} X^{it_j} \ll X^{11/48+\epsilon}. \tag{7.10}$$

By using a smooth function, the Fourier transform, and its inversion, Iwaniec establishes that (7.10) implies (7.9). In the following section, Iwaniec connects the sum in (7.10) to the sum of Kloosterman sums given in (5.5). The latter sum makes use of the Bessel-Kuznetsov transform defined in (4.19). Finally, Iwaniec proves the following mean value estimate for the Rankin zeta function for $\Re(z) = 1/2$:

$$\sum_{|t_j| \leq T} \frac{|R_j(z)|}{\cosh(\pi t_j)} \ll T^{5/2} |z|^A \log^2 T \tag{7.11}$$

where $R_j(z)$ is the Rankin zeta function given by

$$R_j(z) \coloneqq \sum_n \frac{|\rho_j(n)|^2}{n^z}$$

where $\rho_j(n)$ is the n-th Fourier coefficient of corresponding cusp form. This allows Iwaniec to prove (7.10) and thus (7.8).

Subsequently, many mathematicians worked on reducing the error term. In [9], Luo and Sarnak obtained an error term of $O(x^{7/10+\epsilon})$ by improving on Iwaniec's mean value estimate for the Rankin zeta function. They showed that (7.11) holds with the exponent $5/2$ replaced by $2 + \epsilon$. In [2], Cai improved on the error term, reducing it to $O(x^{71/102+\epsilon})$ by refining the steps in Iwaniec's proof that use the estimates for character sums. In [11], Soundararajan and Young further reduced the error term to $O(x^{25/36+\epsilon})$ by placing more emphasis on the connection between prime geodesics and Dirichlet L-functions.

Acknowledgments. I would like to thank my mentor, DeVon Ingram, for guiding me through the dense readings surrounding the prime geodesic theorem and assisting me in writing this paper. And, of course, thank you to Peter May for setting up the REU program and giving me and many others the opportunity to participate in enriching our knowledge of mathematics.

References

[1] Robert Ash and W. P. Novinger. Complex Variables. Chapter 7.
[2] Yingchun Cai. Prime geodesic Theorem. Journal de Théorie des Nombres de Bordeaux **14**, 2002. pp. 59-72.
[3] Keith Conrad. Decomposing $\mathrm{SL}_2(\mathbf{R})$.
[4] Eberhard Freitag and Rolf Busam. Complex Analysis. Springer, 2005. pp. 42-174.
[5] Henryk Iwaniec and Emmanuel Kowalski. Analytic Number Theory. American Mathematical Society Colloquium Publications **53**, 2004.
[6] Henryk Iwaniec. Prime geodesic theorem. 1983. pp. 136-159.

[7] Jaap Korevaar. The Wiener-Ikehara Theorem by Complex Analysis. Proceedings of the American Mathematical Society **134**. 2005. pp. 1107-1116.

[8] Nikolay Kuznetsov. Petersson's conjecture for cusp forms of weight zero and Linnik's conjecture. Sums of Kloosterman sums. Mathematics of the USSR-Sbornik **39**. 1981. pp. 299-342.

[9] Wenzhi Luo and Peter Sarnak. Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$. Publications mathématiques de l'I.H.É.S. **81**, 1995. pp. 207-237.

[10] Peter Sarnak. Class Numbers of Indefinite Binary Quadratic Forms. Journal of Number Theory **15**, 1982. pp. 229-247.

[11] Kannan Soundararajan and Matthew Young. The prime geodesic theorem. Journal für die reine und angewandte Mathematik **676**. 2013. pp. 105-120.

[12] Matt Tyler. The Prime Geodesic Theorem.

[13] Yihan Yuan. The Selberg Trace Formula & Prime Orbit Theorem. 2019.