NUMBER THEORY AND GEODESICS IN HYPERBOLIC SPACES

WILLIAM H. CHUANG

Abstract. **This note was a final project of a Number Theory class that was taught by Prof. Paul Zeitz. Since it's a final project, so the length and the contents must meet the requirements of that final project. Hence three sections of our works that based on the knowledge of this note were removed, but more examples were added. To keep the size down, I have decided to retain somewhat informal style, and this note introduced two connections between prime numbers and geometry:**
(1) Quadratic forms and the norms of closed geodesics.
(2) Prime Number Theorem and the Prime Geodesics Theorem.

Contents

Date: June 16, 2017.

1. Binary Quadratic Forms

1.1. **The Motivations.** The main reason for studying the geometry of numbers by studying the distribution of geodesics in hyperbolic spaces is based on a close connection between the geodesic flow for the modular surface and the continued fraction transformation on the unit interval. The relationship between the two is a well-known result of the fact that the Fuchsian group associated to the modular surface is the modular group $SL(2, \mathbb{Z})$ and the action of this group on the boundary of Beltrami upper half-plane induces the continued fraction transformation (e.g., Möbius transformations). Furthermore, the distribution of geodesics in hyperbolic spaces can give us useful information on the arithmetic of continue fractions, quadratic forms, and the number of prime geodesics is corresponding to the Prime Number Theorem, all of these are important branches of Number Theory.

In order to start from the basic, and scratch, we choose to let quadratic forms, and continued fractions, to be our departure points, because these are common topics in Number Theory, and our goal here is to build some intuitions so that when we extend this geometrical understanding of numbers to the second part we can build-up our intuitions there based on the intuitions we have built in the first part of this note.

In the second part, our focus turn to the works done by Huber, Selberg, and recently Margulis. Based on hyperbolic geometry they proved in three different generality of the understanding of prime distribution, for every hyperbolic surface, an approximation of $\Pi(t)$ is the function of e^t/t. The main technique to understand the behavior of $\Pi(t)$ is Selberg trace formula which is a generalization of the result that has the root in linear algebra that the trace of a symmetric matrix can be computed in two different methods: firstly, we can sum the diagonal entries; secondly, we sum the eigenvalues. In Selberg trace formula, we calculate the trace for infinite matrices (they are Laplace-Beltrami operators on Hilbert spaces). The norms of the closed geodesics correspond to these diagonal terms. The asymptotic behavior of $\Pi(t)$ is derived by focusing on the smallest eigenvalue (so-called the basic frequency) is 0. All of the eigenvalues[1] of the Laplace-Beltrami operator are the harmonics frequencies of the Riemannian manifold M, and can be heard when M was played as a drum.

For every closed geodesic on a manifold, M, we can assign its homology $\phi(\gamma)$, which is a pair of integers describing how many times it wraps around the hole of the manifold. In general, for a surface with genus g, the homology $H_1(M, \mathbb{Z})$ can be represented by points $(n_1, n_2, ..., n_{2g})$ with integer coordinates. Furthermore, a question can be posed: How are the norms of geodesics distributed, if the homology

[1] aka the principal frequencies.

they have are restricted in a given subset of $H_1(M, \mathbb{Z})$? For geodesics that all have the same homology, the answer is given by Phillips and Sarnak[4]. Their work was focusing on: $\Pi(t, \beta) = \{\gamma \in \Pi(t), \phi(\gamma) = \beta\}$. The result they proved is $\Pi(t, \beta) \sim (g-1)^g \frac{e^t}{t^{g+1}}$. This fascinating result turns out only depends on the topological invariant g. Our main task will be to present the big picture; whenever possible, my basic point of view will be that of number theory, since this can shed more lights on the key ideas.

1.2. **Pell's Equation.** One of the quadratic form examples from Number Theory is Pell's equation. It goes back to Proclus (410-485 A.D.) as he noticed that Pythagoreans developed an algorithm for solving the nonlinear diophantine equation (see Ch.15 in [29]), and when it has the following form it's called Pell's equation:

$$x^2 - dy^2 = \pm 1 \tag{1}$$

In order to solve it, we need to find the general solution for this equation. That is to find all pairs of integers x and y that satisfy this equation.
Example. $x^2 - 2y^2 = \pm 1$. The algorithm starts at the smallest solution $(x, y) = (1, 1)$, for $x^2 - 2y^2 = -1$. Likewise, for $x^2 - 2y^2 = 1$, $(x_3, y_3) = (3, 2)$ Give a solution (x_n, y_n), the number pairs $(x_{n+1}, y_{n+1}) = (2y_n + x_n, y_n + x_n)$ are the general solution. We can rewrite the above Pell's equation as $(x + y\sqrt{d})(x - y\sqrt{d}) = 1$. Then the general solution is $(x_n + \sqrt{2}y_n) = (1 + \sqrt{2})^n$, and (x_1, y_1) is the **fundamental solutions**. In other words, to find a solution become to find a nontrivial unit of the unit ring $\mathbb{Z}(\sqrt{d})$ of norm 1.

We can view the solvability of Pells equation as a special case of Dirichlets unit theorem that gives the structure of the group of units of a general ring of algebraic integers. In particular, for the ring $\mathbb{Z}(\sqrt{d})$, it's an infinite cyclic group, and the product of $\{\pm 1\}$.

Example. $d = 14$. We obtain $\sqrt{14} = 3 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{3 + \sqrt{14}}}}}$ Thus the continued fraction expansion of $3 + \sqrt{14}$ is periodic with period length 4.

1.3. **Diophantine Approximation.** Why is 22/7 and 355/113 are chosen as good approximation to π? If we look at 355/113, we can see that $\frac{355}{113} = 3 + \cfrac{1}{7 + \frac{1}{16}}$ approximate π to six decimal places. They are examples of continued fractions, which are used to derive the best approximations to an irrational number for a given upper bound on the denominator, and this is called Diophantine approximation.

Example. Continue the example in the previous subsection. We truncate the expansion of the continued fraction at the end of the first period, then we can get $\sqrt{14} \sim 3 + \cfrac{1}{1+\frac{1}{1+\frac{1}{1}}} = \frac{15}{4}$ Mind that the the denominator and the numerator is the fundamental solution $(x_1 = 15, y_1 = 4)$.

Furthermore, if we let $N \in \mathbb{N}$, and $Ax^2 + Bxy + Cy^2 = N$, then we can see Pell's equation is a special case of quadratic form.

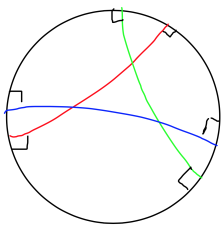


Figure 1: Prime geodesics are those arcs of circles that meet $\partial\mathbb{D}$ orthogonally, and the diameter of $\mathbb{D}$.

1.4. **Recall: Hyperbolic Geometry.** We consider $\mathbb{H}$ is mapped to a unit disk $\mathbb{D} = \{w = u + iv \in \mathbb{C}\,|u^2 + v^2 < 1\}$, and $\mathbb{D}$ has an induced metric $ds^2 = \frac{4(du^2+dv^2)}{(1-(u^2+v^2))^2}$.

$SL(2,\mathbb{R})/\{\pm I\}$ is the group of conformal transformations of $\mathbb{H}$, and the group of orientation preserving isometries of $\mathbb{H}$ w.r.t. the invariant metric[2], $ds^2 = \frac{dx^2+dy^2}{y^2}$. This metric is hyperbolic, that is the Gauβian curvature is -1. We have a corresponding measure $dA(z) = \frac{dxdy}{y^2}$, volume element $dv = \frac{dx\wedge dy}{y^2}$, and distance function $d(z,w) = 2\tanh^{-1}\frac{|z-w|}{|z-\overline{w}|} = \log\frac{|z-\overline{w}|+|z-w|}{|z-\overline{w}|-|z-w|}$, for $z, w \in \mathbb{H}$. Because the hyperbolic metric $ds^2_{\mathbb{H}}$ is conformal[3] to the Euclidean metric, $ds^2_{\mathbb{R}}$, so angles in $\mathbb{H}$ are computed as in Euclidean geometry $\mathbb{R}^2$. Geodesics in $\mathbb{H}$ are arcs of generalized circles orthogonally intersecting the boundary $\partial\mathbb{H} := \mathbb{R}\cup\{\infty\}$.

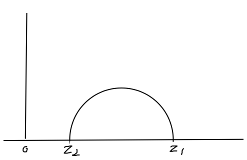


Figure 2: Geodesics are those semicircles, and straight lines that meet $\partial\mathbb{H}$ orthogonally.

1.5. **From Quadratic Forms to Hyperbolic Geometry.** Gauss was the first mathematician to study modular groups after he figured out that the reduction and equivalence of binary quadratic forms. One of the most important objects of study in number theory and geometry is the modular group $\Gamma = SL(2,\mathbb{Z}) = \left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix} : a,b,c,d \in \mathbb{Z}, \text{ and } ad - bc = 1\right\}$.

which acts on the upper half-plane $\mathbb{H} := \{z \in \mathbb{C} : \Im(z) > 0\}$, via a large group of conformal automorphisms, a linear fractional transformation[4], that is: $z \mapsto gz = \frac{az+b}{cz+d}, g = \begin{pmatrix} a & b \\ c & d\end{pmatrix}$.

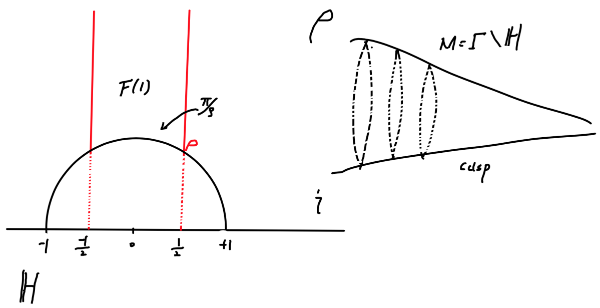
Then, the orbits of $\mathbb{H}$ under this action form a quotient surface which has fundamental domain

$$\Gamma \setminus \mathbb{H} := \{z \in \mathbb{C} : |z| > 1 \text{ and } |\Re(z)| < 1/2\}.$$

[2]Hence, under the $PSL(2,\mathbb{R})$ action, $\mathbb{H}$ has an invariant metric.

[3]$ds^2_{\mathbb{H}} = \frac{ds^2_{\mathbb{R}}}{y^2}$

[4]aka Möbius transformation.



Conversely, if we started from $\mathbb{H}$, and consider the fundamental domain $\mathcal{F}(1)$, which is a compact triangle on Beltrami disk, and is a triangle of finite area (by Gauβ-Bonnet theorem). Then, in $\mathcal{F}(1)$, we can glue $x_1 = \frac{1}{2}$ with $x_2 = \frac{-1}{2}$, using the isometry of $\mathbb{H}, z \mapsto z+1$, and glue $-\rho$ to i with i to ρ by using isometry $z \mapsto 1/z$, then these isometries generate the modular group: $\Gamma = SL(2,\mathbb{Z})$. On the geometry side, the resulting hyperbolic space (aka Riemann surface) $M = \Gamma \setminus \mathbb{H}$ is a modular surface with a cusp.

Gauβ might have been noticed the connection here between number theory and geometry:

- definite forms may be explained as points in $\mathbb{H}$
- indefinite forms may be interpreted as geodesic semicircles on $\mathbb{H}$

These geometrical viewpoint also show us why it's more difficult to study indefinite forms than definite forms (points versus curves).

To understand Γ better, it helps to understand what kind of elements can be found in this group. One natural way of classifying these transformations is to look at their fixed-point sets. So, for the Möbius transformation, we want to find the fixed point of $z \mapsto (az+b)/(cz+d)$, we need

$$\gamma \in \Gamma : z \mapsto (az+b)/(cz+d) \Leftrightarrow cz^2 + (d-a)z - b = 0. \tag{2}$$

If $c \neq 0$, we can use the quadratic formula to solve this equation:

$$z = \frac{a-d \pm \sqrt{(d-a)^2 + 4ac}}{2c} \tag{3}$$

Then, we want to understand the discriminat better so that we can see what kinds of points the Möbius transformation can fix. Since $ad - bc = 1$, hence $(d-a)^2 + 4bc = d^2 + a^2 - 2ad + 4(ad-1) = d^2 + a^2 + 2ad - 4 = (d+a)^2 - 4$. Thus, the type of fixed points depends only on $(d+a)^2$. We define the trace of $\gamma \in \Gamma$ to be $Tr(\gamma) = |a+d|$, and we classify γ based on its trace. If $Tr(T) = 2$, then T fixes exactly one point in $\mathbb{R}$. If $Tr(\gamma) < 2$, then γ fixes two points (that are conjugate) in $\mathbb{H}$. If $Tr(\gamma) > 2$, then γ fixes two points in $\mathbb{R}$. On the other hand, if $c = 0$, then we have $\frac{1}{d} = a \Rightarrow z = a(az - b)$. If $a = 1$, there there is only one solution. If $a \neq 1$, we obtain $z = \frac{ba}{1-a^2}$, which is fixed. Hence, if $Tr(\gamma) = 2$ we fix ∞, and if $Tr(\gamma) = |a + \frac{1}{a}| > 2$, then γ fixes ∞, and one point in $\mathbb{R}$. In summary, we have three classes:

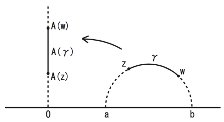


Figure 3: An oriented geodesic from a to b is called the ***axis*** of γ.

- if $Tr(\gamma) > 2$, and fixes two points in $\mathbb{R} \cup \{\infty\}$, then γ is hyperbolic.
- if $Tr(\gamma) < 2$, and fixes one point in $\mathbb{H}$, then T is elliptic.
- if $Tr(\gamma) 2$, and fixes one point in $\mathbb{R} \cup \{\infty\}$, then γ is parabolic.

Now, if $\gamma \in \Gamma$ and $|tr(\gamma)| > 2$, then γ is hyperbolic and determines two distinct fixed points in $\mathbb{R}$, and the two points are joining on $\mathbb{H}$ by the semi-circle geodesics; furthermore, these geodesics are projected to $\Gamma \setminus \mathbb{H}$. In other words, a segment of geodesic semi-circle joining a point p, to another point γp, and it's projected to a closed geodesic on $\Gamma \setminus \mathbb{H}$. To be more clear, consider there are two points z, and w in $\mathbb{H}$, and $\gamma(w) = \frac{1}{(cw+d)^2} < 1$ is attracting. $\gamma(u) > 1$ is repelling. Then an oriented geodesic from u to w is called the ***axis*** of γ. If $\gamma \in \Gamma$, and Γ is a Fuchsian group, then its axis is an ***oriented closed geodesic*** in the quotient-space $\Gamma \setminus \mathbb{H}$. If two elements γ_1, and γ_2 are conjugate in Γ, that is $\gamma_1 = \gamma\gamma_2\gamma^{-1}$, for some $\gamma \in \Gamma$, then γ maps the axis of γ_2 to the axis of γ_1, therefore, they represent the same oriented closed geodesic in $\Gamma \setminus \mathbb{H}$. In the opposite direction, each geodesics in $\Gamma \setminus \mathbb{H}$ represents the conjugacy class of a primitive[5] hyperbolic transformation in Γ. A ***prime geodesic*** on $\Gamma \setminus \mathbb{H}$ is a closed geodesic that traces out its image exactly once.

Moreover, there is a one-to-one correspondence between a closed geodesics on $\Gamma \setminus \mathbb{H}$, and a hyperbolic conjugacy class in Γ. Again, suppose Γ is a Fuchsian group whose elements have matrix entries locate in the ring of integers O_F of a number field F. For a prime geodesic $\wp$ on $\Gamma \setminus \mathbb{H}$, and a prime ideal Q of O_F, the Frobenius conjugacy class $Frob(\wp)$ is a conjugacy class derived by reducing the associated hyperbolic conjugacy class $\{\gamma\}$ modulo Q. This Forbenius map is analogous to the map when taking a prime ideal to its Frobenius conjugacy class in the Galois group of an extension of number fields. This, again, shows the connection between number theory and geometry. Based on this result, we can determine how the geodesic $\wp$ lifts to $\Gamma(Q) \setminus \mathbb{H}$, where $\Gamma(Q)$ is the congruence subgroup of all matrices in Γ, and they are equivalent to $\pm I$ mod Q. These symmetries form the projective special linear group $PSL(2, \mathbb{Z}) := SL(2, \mathbb{Z})/\{\pm I\}$, so that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ are considered equivalent.

[5]That is not a power of another.

1.6. **Indefinite Quadratic Forms.** Now, let's go back to (2), and consider a more general quadratic form:

$$Q(x,y) = Ax^2 + Bxy + Cy^2 = \begin{pmatrix} x \\ y \end{pmatrix}^T M \begin{pmatrix} x \\ y \end{pmatrix} \tag{4}$$

where $A, B, C \in \mathbb{Z}$, $x, y \in \mathbb{C}$, $Q(x,y) = N, N \in \mathbb{N}$, and $M = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$.

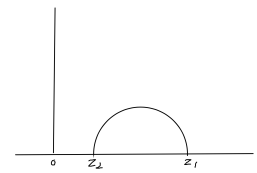


Figure 4: Suppose $z_1 = \frac{-b+\sqrt{d}}{2a}$, $z_2 = \frac{-b-\sqrt{d}}{2a}$. Note that they are connected by a geodesic.

Suppose

$$Q(x,y) = 0, d = B^2 - 4AC, \text{ then } z_1 = \frac{-b+\sqrt{d}}{2a}, z_2 = \frac{-b-\sqrt{d}}{2a}$$

Definition

Two quadratic forms $Q_1(x,y)$ and $Q_2(x,y)$ are equivalent if $\exists \alpha, \beta, \gamma, \delta \in \mathbb{Z}$ such that $\alpha\delta - \beta\gamma = 1$, and $Q_1(x,y) = Q_2(\alpha x + \beta y, \gamma x + \delta y)$.

Example. $Q_1(x,y) = x^2 - 2y^2$ is equivalent to $Q_2(x,y) = -2x^2 + y^2$. However, forms can be equivalent to a more general way: $Q_3(x,y) = x^2 + 2xy - y^2$ is equivalent to $Q_1(x,y)$, and $Q_2(x,y)$. The reasons are: $(x+y)^2 - 2y^2 = x^2 + 2xy - y^2$, and $(x-y)^2 + 2(x-y)y - y^2 = x^2 - 2y^2$. This is why we define in this way.

Consider the following two matrices: $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. These two linear transformations $S(x,y) = (-y, x)$, and $T(x,y) = (x+y, y)$ show the equivalence of $Q_1 \cong Q_3$, and $Q_2 \cong Q_3$.

Moreover, a fundamental quantity in number theory is linked to this which is the ***Class Number***.

Definition

Let $d > 0$ be given, the class number $h(d)$ is the number of nonequivalent quadratic forms with given discriminant $d := B^2 - 4AC$.

If $x = z, y = 1, B = (d-a)$, then we derive the (2) again. Now, if $(d-a)^2 + 4ac > 0$, then $Q(x,y)$ is called ***indefinite***. In particular, (2) determine a unique geodesic γ on $\mathbb{H}$ with $\gamma(+\infty) = x$, and $\gamma(-\infty) = x'$ that quotients down to a closed geodesic γ on the modular Riemann surface $M = \Gamma \setminus \mathbb{H}$ that we made by gluing method, and the modular surface$M$ has a cusp. *Again, the focus of this note is only concentrating on those image geodesics are closed.*

Theorem 1. *Equivalence classes $[Q]$ of quadratic forms $Q(x,y) = Ax^2 + Bxy + Cy^2$ with discriminant d correspond to all closed geodesics γ of length $\ell(\gamma)$.*

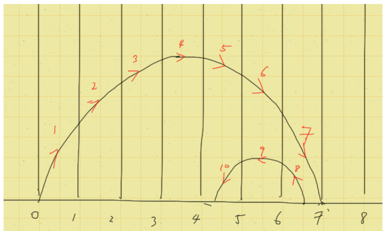
$$\ell(\gamma) = 2\log(\epsilon_d) = 2\log\left(\frac{x_n + y_n\sqrt{d}}{2}\right) \tag{5}$$

This equation also bridges the fundamental solution to the lengths of closed geodesics. *This length of the all closed geodesics in a certain arithmetic hyperbolic surface (associated with the matrices T and S) are identified with $\ell(\gamma) = 2\log\epsilon_d$, with each length appearing $h(d)$ times,* ***where d is restricted to be not a square and*** *$d \equiv 0$* ***or*** *$1(mod4)$.*

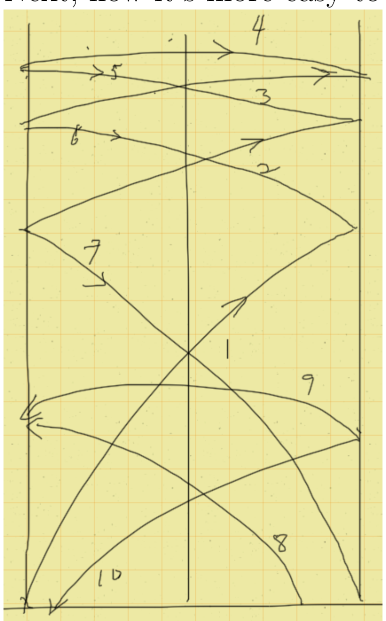
If we look at the lifts γ' of closed geodesics γ, where $\gamma'(\infty) > 1$, and $0 < \gamma'(-\infty) < 1$. Then we can have a connection between continued fractions and closed geodesics, and this connection is called Gauβ map (aka Gauβ continued fraction map). For $\alpha = n_0 + \cfrac{1}{n_1 + \frac{1}{n_2 + \frac{1}{\frac{1}{n_3 + \dots}}}} = [n_0; n_1, n_2, n_3, ...]$ The values n_i have a geometric meaning that they are the number of times the geodesic wraps around the cusp, **the number of times the geodesic that links two points on real line over** $\mathbb{H}$, so once we glue these tiles and get the cusp, we can see it wraps around the cusp on each excursion into the cusp.

Example. If we consider Zagier's "minus" convention, we denote a real number with $[n_0; n_1, n_2, n_3, ...] = n_0 - \cfrac{1}{n_1 - \frac{1}{n_2 - \frac{1}{\frac{1}{n_3 - \dots}}}}$ Then we can represent a fixed point $w = 4 + 2\sqrt{3}$ on real line by joining the origin to w. Thus, $w = [\overline{8,2}] = 8 - \cfrac{1}{2 - \frac{1}{8 - \frac{1}{\frac{1}{2 - \dots}}}}$ Furthermore, we can only consider this in quotient space. We can consider it takes 7 crossings to the right, then it goes back with 3 crossings, that is it hits 8 boundaries of the triangle when it heads to the right, then it hits another 2 times when it moves backward.

First, we start to think of the whole upper-half plane tiled in the following Farey tessellations.



Next, now it's more easy to understand the following idea:



This is why the each digit in the code $[n_0; n_1, n_2, n_3, ...]$ tells how far it should wrap the cusp. Further, the idea is since the period is two, so we can denote $w = [\overline{8, 2}]$, and the closed geodesic in F is determine by the quadratic form: $Q = \begin{pmatrix} 15 & -8 \\ 2 & -1 \end{pmatrix}$.

The above figure has ten closed geodesic segments. Number 1 to 7 are from left to right (w.r.t. to the vertex, on real line, of the triangle it crosses), and can be denoted by an operator (it's also a quadratic form matrix) L. Number 8 to 10 is in the opposite, and denoted as R. Hence, $w = [\overline{8, 2}] = 8 - \cfrac{1}{2 - \cfrac{1}{8 - \cfrac{1}{2 - \cdots}}} \longleftrightarrow [\overline{L^8, R^2}]$

where $[\overline{L^8, R^2}]$ is called the cutting sequence. The idea goes back to Artin, and was firstly developed by Caroline Series[27] in "plus" sign convention (which has the same correspondence, and the only difference is the signs in writing the continued fractions), and a nice $\Gamma(2)$[6] example, and the introduction of the original idea is in[30].

Based on theorem 1; moreover, Gauβ's one more observation at here is he noticed the dependence $h(d)$ on d is very irregular. Then, Siegel showed that

$$\sum_{\epsilon_d \leq x, d \leq x} h(d) \log \epsilon_d \sim \frac{\pi^2 x^{3/2}}{18\zeta(3)}, x \to \infty.$$

2. Prime Number Theorem

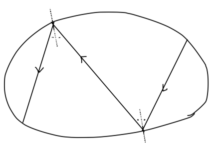


Figure 5: A good example to dive in is the planar billiard.

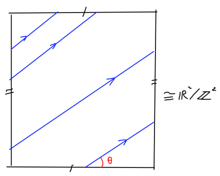


Figure 6: $\theta \in \mathbb{Q}$, if and only if the trajectory is periodic. If $\theta \not\in \mathbb{Q}$, then the orbit is equidistributedl, and dense within $\partial\Omega$.

2.1. **Einstein's Problem of 1917.** The question he posed: "how can classical mechanics give us any hints about the quantum mechanical energy levels when the classic system is ***ergodic***?"[19]. This is the departure point of the studies of the modern quantum chaos in complex system. A good example to dive in is the planar billiard. Consider a point particle moves on a flat surface, and the motion is confined to a compact domain $\Omega \in \mathbb{R}^2$ with boundary $\partial\Omega$, and $\partial\Omega$ is a hard wall (all collisions are elastic), that is a flat torus $\mathbb{R}^2 \setminus \mathbb{Z}^2$ with genus $g = 1$, so this space is not hyperbolic. It turns out the motion is sensitive to the geometry of the boundary. If the boundary is a rectangle, an ellipse, a circle, the system is integrable. However, if the shape is a stadium, then it becomes a strongly chaotic system. The study of the billiard problem become influential for the development of modern ergodic theory, and the breakthrough was made by Artin in 1924. Artin's billiard is a two-dimensional non-Euclidean billiard, that is a non-compact Riemann surface of constant negative Gauβian curvature[7] $K = -1$. This can be made by removing one point on a torus to make it be punctured. Some further details of this billiard are covered in the following subsections. Artin used an idea which goes back to Gauβ so that he could formulate the geodesic motion as a map in terms of ***continued fractions***, which we just constructed some basic understandings in the previous section. ***It can be shown that the geodesic flows on compact symmetric Riemann spaces are Anosov system, and all Anosov systems are ergodic.*** Furthermore, for a general Anosov flow, a homology class in $H_1(M, \mathbb{Z})$ may contain only a finite number of periodic orbits[9].

The next breakthrough was done by Yakov Sinai in 1963[26], he proved that the ideal gas billiard with Boltzmann-Gibbs statistics is ergodic. After Sinai, in 1970, the next milestone was set by Gutzwiller. Based on Feynman's idea that "sum over histories" as a third approach, so-called Feynman's path integrals, to recreate the Quantum Mechanics, and Quantum Field Theory, Gutzwiller pioneered the idea

[6]A congruence subgroup of $SL(2, \mathbb{Z})$.

[7]Geodesics diverge exponentially fast in t[17].

of summing over all classical solutions in semiclassical quantization[8] expansions of Greens functions and chaotic spectral problems. **The observation he made is the trace of the energy-dependent Green's function[9] is given by a formal sum over all classical *closed orbits* in phase space, that is, to sum over all periodic orbits. If we apply Gutzwiller's method to Artin's billiard, then the growth rate of this sum is in a number that we will prove in the next two subsections. Furthermore, this can show that $K = -1$, the Gauβian curvature, is the key source for entropy and chaos in billiards in hyperbolic polygons. That is one of the physical meanings of the prime number theorem (PNT), and the physical meaning of t, or $\log x$, in (36), and in(28) are the most important global property of a strongly chaotic system, and it's called *topological entropy*.** The geodesic (Hamilton) folw on negative curvature Riemann surfaces (hyperbolic spaces) are the best understood fully chaotic flows (Hopf, Morse, Sinai, and so on).

2.2. A counting function for prime numbers.

Definition

$\pi : \mathbb{R} \longrightarrow \mathbb{N}$. Consider the following arithmetic function:

$$\pi(x) := \#\{p \mid p \leq x\} \tag{6}$$

where p stands for prime number.

Theorem 2. *The prime number theorem says that*

$$\pi(x) \sim \frac{x}{\log x} \tag{7}$$

as x approaches to ∞.

The first proof was independently given by Hadamard and de la Vallée Poussin in 1896, and analytic properties of the Riemann zeta function were used in their proofs. Our proof of PNT will base on the a Tauberian theorem introduced by Ikehara (1931). The following definition is our departure point to prove Theorem 2.

[8]The WKB method.

[9]The Fourier transform of the time-evolution operator.

Definition

The original zeta function. Consider the open half-plane $A = \{s = \mathbb{C} \mid \Re(s) > 1\} \subset \mathbb{C}$. Then we can define a map by using the set A as the domain to the complex plane as its range:

$$\zeta : A \longrightarrow \mathbb{C}$$

then we have the following Euler zeta function[a]

$$s \longmapsto \zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{8}$$

[a] The sum (aka Dirchlet sum) converges on the domain A(see lemma 1), and it is holomorphic on the complex plane $\mathbb{C}$ (by Morera's theorem, and then we can show it's analytic on the domain A, by using Weierstrass M-test and Cauchy's integral theorem).

The following (10) will be used in our proof of PNT.

- First, we need to define Euler gamma function: $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ (which can be proved that is also well-defined, and holomorphic, if $s \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$).
- then we change variable for t to $n\alpha$, and multiply both sides by $\frac{1}{n^s}$, and then take the summation, $\sum_{n=1}^{\infty}$, on both sides. Then, we will reach two cases for α, which are $\alpha \in [0,1]$, and $\alpha \in [1,\infty)$ which give us nonnegative Lebesgue integrable functions.
- Then, the following Mellin transform expression was firstly derived by Abel in 1823 (and then Riemann (1859) went further):

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty \frac{\alpha^{s-1}}{e^\alpha - 1} d\alpha. \tag{9}$$

- By using Gamma function $\Gamma(s) = \int_0^\infty \frac{\alpha^{s-1}d\alpha}{e^\alpha} \Rightarrow \frac{\Gamma(s)}{n^s} = \int_0^\infty \frac{t^{s-1}}{e^{tn}}dt$. Riemann (1859) obtained: $\Gamma(s)\zeta(s) = \int_0^\infty \frac{t^{s-1}}{e^{t-1}}dt$.
- To see a deeper connection between the above ζ-function and primes, Riemann used (5) and (10) to derive the below equation that will also be applied in our proof of PNT:

$$\frac{-\zeta'(s)}{\zeta(s)} = \sum_{p_i} \frac{\log p_i}{p_i^{ms}} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = \int_1^\infty x^{-s} d\psi(x) = \int_0^\infty e^{-st} d\psi(e^t) = s\int_0^\infty e^{-st}\psi(e^t)dt. \tag{10}$$

where $\Lambda(n)$ is von Mangoldt's function, and $\psi(x)$ is Chebyshev function.

Now, let's introduce Ikehara-Wiener theorem which was revised by Ikehara from his mentor Wiener's Tauberian theorem:

Theorem 3.
Let the real valued function f be non negative and non-decreasing on $[0,\infty)$, and apply the Mellin transform

$$g(s) := s\int_1^\infty f(x)x^{-s-1}dx \tag{11}$$

Then by using Stieltjes integral, and integration by parts, the following expression exists for $\Re(s) > 1$, we have $g(s) = f(1) + \int_1^\infty x^{-s}df(x)$. Furthermore, if for some constant α, the holomorphic function $g(s) - \frac{\alpha}{s-1}$ has a continuous extension to the closed half-plane with $\Re(s) \geq 1$, then $\lim_{x\to\infty} \frac{f(x)}{x} = \alpha$, whenever $\alpha \geq 0$.

2.3. **Wiener and Ikehara's proof of Prime Number Theorem.**
Motivation. The reason why we focus on Wiener and Ikehara's approach in this note is due to the reason that the notions we developed in this proof will be used in learning Huber's proof of Prime Number Theorem of Closed Geodesics.[10]

Proof. From equation (8), we let $x = e^t$, then we have $\frac{-\zeta'(s)}{\zeta(s)} = s\int_1^\infty \psi(x)x^{-(s+1)}dx$. Then, since $\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}$ has a holomorphic extension to the open half-plane $\Re(s) > 0$, hence by Ikehara-Wiener theorem we have $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$. Now we have $\psi(x) \sim x$, as $x \longrightarrow \infty$. *The following derivations are elementary, but will be used as a connection with the closed geodesic.*

We can notice that $\psi(x)$ doesn't grow as fast as $\vartheta(x) := \sum_{p_i\leq x}\log p_i$. Furthermore, $\psi(x) - \vartheta(x) = \sum_{m=2}^{m_x}\vartheta(x^{1/m})$ where m_x is the greatest integer such that $2^m \leq x$. Since $m_x \leq \log_2 x$ and $\vartheta(x^{1/m}) \leq \vartheta(x^{1/2}) \leq \psi(x^{1/2}) = O(x^{1/2})$, hence

$$\vartheta(x) \sim x, \text{ as } x \longrightarrow \infty.$$

Lastly, integrating by parts[3], we can obtain

$$\pi(x) = \sum_{p\leq x} 1 = \int_2^x \frac{d\vartheta(u)}{\log u} = \frac{\vartheta(u)}{\log u} + \int_2^x \frac{\vartheta(u)}{u(\log u)^2}du \tag{12}$$

Furthermore, from (9), and use the result we just obtain $\vartheta(x) \sim x$, we have the following

$$\frac{\pi(x)}{x/\log x} = \frac{\vartheta(x)}{x} + o(1) \Rightarrow \pi(x) \sim \frac{x}{\log x}. \tag{13}$$

□

[10]Selberg's elementary proofs, with one version that he used Erdös' result, can be reached at[1].

3. Prime Number Theorem for the Compact Riemann Surfaces

The main idea in this section is counting points in lattice orbits on the hyperbolic upper half-plane $\mathbb{H}$.

3.1. Huber's proof on PNT of closed geodesics.

Before asking a practical question, let's define some more new notations, and notions.

- dist(,) denotes the distance function.
- $\ell(\gamma)$ denotes the length of a curve γ.
- A curve $\gamma : [a, b] \to M$ is called simple, if γ is an injective mapping.
- Suppose M be a topological space, then two **closed curves** $\gamma_1, \gamma_2 : S^1 \to M$ are called **homotopic**, if there is a continuous map $\psi : [0, 1] \times S^1 \to M$ such that $\psi(0, t) = \gamma_1(t)$, $\psi(1, t) = \gamma_2(t)$, where $t \in S^1 = \mathbb{R}/[t \mapsto t + 1]$. (That is we can freely deform one curve to the other.)
- Two parametrized closed geodesics $\gamma_1, \gamma_2 : S^1 \to M$ (at unit speed) are **equivalent** if there is a homeomorphism $h : S^1 \to S^1$ of the form $h(t) = t+$ const such that $\gamma' = \gamma \circ h$. Then, a **closed geodesic** is an **equivalence class** of closed parametrized geodesics.
- *Let γ_1, γ_2 are closed geodesics, and let $m \in \mathbb{Z}\setminus\{0\}$. $\gamma_1 = \gamma_2^m$ denotes γ_1 is the m-fold iterate of γ_2, if $\gamma_1(t) = \gamma_2(mt), t \in S^1$.*

Definition: Γ-primes

Let $M = \Gamma\backslash\mathbb{H}$ be a compact Riemann surface of genus $g \geq 2$. Then $\forall\gamma \in \ell(M)$, where γ is a closed geodesic, $\exists\gamma_0 \in \wp(M)$, where γ_0 is a unique prime geodesic, and $m \geq 1$ is a unique exponent such that $\gamma = \gamma_0^m$.

	N-primes	Γ-primes
bounded numbers of primes	$\pi(x) := \#\{p \mid p \leq x\}$	$\Pi(t) = \#\{\gamma \in \wp(\Gamma \setminus \mathbb{H}) \vert \ell(\gamma) \leq t\}$ $\phi(t) = \#\{\gamma \in \ell(\Gamma \setminus \mathbb{H}) \vert \ell(\gamma) \leq t\}$
von Mangoldt's func.	$\Lambda(n) = \log p$, if $n = p^m$, for some prime $p, m \geq 1$, otherwise, $\Lambda(n) = 0$	$\Lambda(\gamma) = \ell(\gamma_0) = \log N_{\gamma_0}$
Chebyshov func.	$\vartheta(x) = \sum_{p\leq x} \log p$ $\psi(x) = \sum_{n\leq x} \Lambda(n)$	$\theta(t) = \sum_{\ell(\gamma)\leq t} \ell(\gamma) = \int_0^t \tau d\Pi(\tau)$ $\Psi(t) = \sum_{\ell(\gamma)\leq t} \Lambda(\gamma)$

For Γ-primes, # means cardinality; the prime in the definition of Chebyshov θ is the restricted summation of prime geodesics. Now, we can rewrite lengths to norms[11]

[11] $N_{\gamma^k} = (N_\gamma)^k$. This implies the norm of geodesics is a **completely multiplicative function**. Instead of considering lengths, it's more general to consider the norms of the prime geodesics. If f is a completely multiplicative function, then f satisfies the requirement: $f(mn) = f(m)f(n)$,

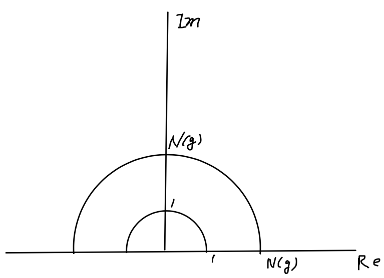
(so that we can obtain a more generic framework) by substituting $x = e^t$, and adding a subscript N to each N-prime counting functions.

According to Theorem 1, we can define the norm of a closed geodesic by taking the length as it's power, that is: $e^{\ell(\gamma)} = \left(\frac{t+u\sqrt{d}}{2}\right)^2$**.**, where (t, u) is the fundamental solution of Pell's equation, $\ell(\gamma)$ is the eigenvalue of $g \in \Gamma$, $g = \begin{pmatrix} \frac{t-Bu}{2} & -Cu \\ Au & \frac{t+Bu}{2} \end{pmatrix}$, and $t^2 - du^2 = 4$ (it's a Pell's equation). Note that A, B, and C are from the quadratic form which can satisfies:

$$Q' \sim Q \Leftrightarrow M = \gamma^T M \gamma, \gamma \in \Gamma,$$

and $Q \to g \in \Gamma$.

Mostly, for $g \in SL(2, \mathbb{R})$, $g \sim \begin{pmatrix} N(g)^{1/2} & 0 \\ 0 & N(g)^{-1/2} \end{pmatrix}$. Furthermore, $\epsilon_d = \frac{t+u\sqrt{d}}{2}$ is the eigenvalue of g.



From Pell's Equation to lengths of closed geodesics

Proof.

$$\ell(\gamma) = \int_1^{N(g)} \frac{1}{y} dy = \ln N(g) = \ln(\epsilon_d^2) \tag{14}$$

$$\Rightarrow e^{\ell(\gamma)} = \epsilon_d^2. \tag{15}$$

where ϵ_d is from $t^2 - du^2 = 4$, and (t, u) is the fundamental solution. □

$\forall m, n \in \mathbb{N} - \{0\}$. If f is only **multiplicative function** (not completely multiplicative), then $m \perp n$, i.e., m, n must be **coprime**.

	N-primes	Γ-primes
norms	$\|p\|$	$N_\gamma := e^{l(\gamma)}$
bounded numbers of primes	$\pi_N(x) := \#\{p \mid p \leq x\}$	$\pi(x) = \#\{\gamma \in \wp(\Gamma \setminus \mathbb{H}) \mid N_\gamma \leq x\}$ $\varphi(x) = \#\{\gamma \in \ell(\Gamma \setminus \mathbb{H}) \mid N_\gamma \leq x\}$
von Mangoldt's func.	$\Lambda_N(n) = \log p$, if $n = p^m$, for some prime $p, m \geq 1$, otherwise, $\Lambda_N(n) = 0$	$\Lambda(\gamma) = \ell(\gamma_0) = \log N_{\gamma_0}$
Chebyshov func.	$\vartheta_N(x) = \sum_{p \leq x} \log p$ $\psi_N(x) = \sum_{n \leq x} \Lambda(n)$	$\vartheta(x) = \sum_{N_\gamma \leq x} \ell(\gamma) = \int_1^x \log \xi d\pi(\xi)$ $\psi(x) = \sum_{N_\gamma \leq x} \Lambda(\gamma)$

This prime geodesic theorem (PGT) can be proved independently by using Selberg's trace formula; however, we cannot use that approach to prove PNT, so instead, we dive into Huber's proof which can be bridged to PNT. Suppose $s \in \mathbb{C}, \Re(s) > 1$, and suppose $K(\rho) = (\cosh(\rho))^{-s}$ is a generating function that satisfies lemma 5.

Let $M = \Gamma \backslash \mathbb{H}$ be a given *orbifold*, where Γ is a congruence subgroup of $PSL(2, \mathbb{R})$, and it acts *freely discontinuously*[12] on $\mathbb{H}$[13]

Furthermore, K generates a smooth heat kernel $\mathcal{K}$ on $\Gamma \setminus \mathbb{H}$, and its Γ-bi-invariant lift[14], $\mathcal{K}_\Gamma(z, w)$ is Dirichlet series $G(s; z, w)$ that has the following expression:

$$\mathcal{K}_\Gamma(z, w) = G(s; z, w) = \sum_{T \in \Gamma} (\cosh(dist(z, Tw)))^{-1}, \tag{16}$$

which is a Dirichlet series, and it can be proved to be a holomorphic function of s for $\Re(s) > 1$.

Claim: Our goal here is to prove that this can be extended to a bit larger region $\{\Re(s) > 1 - \delta\}$, once we get there, this will be an entry of the Wiener-Ikehara Theorem, and can let us bridge Number Theory.

Proof. We can rewrite the Dirichlet series and the heat kernel by using the lemma 9.3.5 in [5]. The function $h = h_s$ is the transform of $K = \cosh^{-s}$.

[12]That is a free regular set. So, $\exists$ a neighborhood U of $z \in \mathbb{H}$ such that $\forall T \in \Gamma \setminus \{I\}$, we have $T \circ U \cap U = \emptyset$.

[13]It's the same requirement for defining Dirichlet polygons.

[14]A differential operator on a Lie group G is said to be bi-invariant if it is both right and left invariant. Here the Dirichlet series $G(s; z, w)$ is a natural Γ-lift on $\mathbb{H}$.

$$h(r) := \int_{\mathbb{H}} k(z,i)\Omega(z)d\,\mathbb{H}(z) \tag{17}$$

(18)
$$= \int_{-\infty}^{+\infty}\int_0^{+\infty} y^{1/2+ir} L\left(\frac{1+x^2+y^2}{2y}\right)\frac{dy}{y^2}dx = \sqrt{2}\int_{-\infty}^{+\infty} e^{iru}\int_{|u|}^{\infty}\frac{K(\rho)\sinh\rho}{\sqrt{\cosh\rho-\cosh\rho}}d\rho du.$$

If we let $s_n = \frac{1}{2} \pm ir_n$, then for $n > 0$, and $0 \leq \lambda_n \leq \frac{1}{4}$, s_n are spectral zeros of the Selberg Zeta function (which is based on Dirichlet L-function when L-function is written as an Euler product in the half-plane of absolute convergence)[9].

$$\mathcal{K}_M(x,y) = \sum_{n=0}^{\infty} h_s(r_n)\varphi_n(x)\varphi_n(y), \tag{19}$$

$$G(s;z,w) = \sum_{n=0}^{\infty} h_s(r_n)\phi_n(z)\phi_n(w), \tag{20}$$

$x, y \in M$, and $z, w \in \mathbb{H}$. n is running from $0, 1, ...$ the function ϕ_n is lifting from eigenfunction φ_n of the Laplacian on $M = \Gamma \setminus \mathbb{H}$.

$$r_n = \begin{cases} i\sqrt{\frac{1}{4}-\lambda_n} & \text{if } 0 \leq \lambda_n \leq \frac{1}{4}, \\ \sqrt{\frac{1}{4}-\lambda_n} & \text{if } \lambda_n \geq \frac{1}{4}. \end{cases}$$

That is, r_n is mapped from eigenvalues λ_n.

In the following derivations, Γ temporarily denotes *Gamma* functions unless we emphasize that it denotes the congruence subgroup of $PSL(2,\mathbb{R})$. If $\Re(\eta) > -1, \Re(w) > \frac{1}{2} + \frac{1}{2}\Re(\eta), a > 0, b > 0,$
then[23]

Γ is holomorphic in $\mathbb{C}$, except at: $z = 0, -1, -2, ...,$ these are poles with residues $\frac{(-1)^n}{n!}$.

$$\int_0^{\infty}\frac{\eta dt}{(a+bt^2)^{\omega}} = \frac{a^{(\eta+1-2\omega)/2}}{2b^{(\eta+1)/2}}\frac{1}{\Gamma(\omega)}\Gamma\left(\frac{1}{2}(\eta+1)\right)\Gamma(\omega-\frac{1}{2}(\eta+1)). \tag{21}$$

For $|\Im(r)| \leq \frac{1}{2}$ and $\Re(s) > 1$, if $z = x + iy \in \mathbb{H}$, then
(22)
$$\int_0^{\infty} y^{ir-3/2}\left(\frac{1+x^2+y^2}{2y}\right)^{-s} = 2^{s-1}(1+x^2)^{(-1/2+ir-s)/2}\frac{1}{\Gamma(s)}\Gamma(\frac{1}{2}(s+ir-\frac{1}{2}))\Gamma(\frac{1}{2}(s-ir+\frac{1}{2})).$$

Compare equation (18), and (22), and recall $K(\rho) = (\cosh\rho)^{-s}$, and let $L(\tau) = \tau^{-s}$, we can see

$$h_s(r_n) = 2^{s-1}\pi^{1/2}\frac{1}{\Gamma(s)}\Gamma(\frac{1}{2}(s-\frac{1}{2}+ir_n))\Gamma(\frac{1}{2}(s-\frac{1}{2}-ir_n)). \tag{23}$$

By using an identity[15] in [23], and since as $n=0$, we have $r_0=\frac{i}{2}$, so we can substitute r_0 into (23), and derive

$$h_s(r_0)=\frac{2\pi}{s-1}. \tag{24}$$

If $n\leq 1$, we have $\lambda_n\geq\lambda_1>0$, thus, $\Re(\frac{1}{2}-ir_n)\leq\delta$, for some $\delta>0$. Furthermore, since $\Gamma\neq 0\forall z\in\mathbb{C}$, so for each $n,n\geq 1$, the function $s\mapsto h_s(r_n)$ is meromorphic with poles that are contained in $\{\frac{1}{2}\pm ir_n-2k|k=0,1,2,...\}$, and for $n>0$, $h_s(r_n)$ is holomorphic in the half-plane $\Re(s)\geq(1-\delta)$.

□

Claim: Our next goal is to study the growth rate of $h_s(r_n)$ as $n\to\infty$, for some $s\in B\subset\{s\in\mathbb{C}\,|\Re(s)\geq(1-\delta)\}$, whenever B is a compact domain.

Proof. Since B is compact, so we can apply Stirling formula to (23), and reach $|\Gamma(\xi-i\rho)\Gamma(\xi+i\rho)|\leq c_1\rho^{2\Re(\xi)-1}e^{-x\rho}$, where $\xi\in B$, $c_1\in\mathbb{R}$, and $\forall\rho\in\mathbb{R},\rho\geq 1$. Moreover, we can have $\|h_s(r_n)\|\leq c_2\lambda^{-2}$, where $c_2\in\mathbb{R},\forall s\in B$, and $n\geq 1$. Then substitute this $h_s(r_n)$ back into (20), so we obtain $\sum\lambda_n^{-2}\phi_n(z)\phi_n(w):=G_0$.

By Gauβ-Bonnet theorem, we know $area(M)=2\pi(2g-2)$, where g is the genus of $M=\Gamma\setminus\mathbb{H}$. Thus, $\phi_0(z)=constant=(\text{area}(M))^{\frac{-1}{2}}$. It follows that as this identity is substituted into (20), and $\phi_0(z)=\text{constant}=(\text{area}(M))^{\frac{-1}{2}}$, so it gives us

$$G(s;z,w)=G_0+\frac{1}{2(g-1)(s-1)}. \tag{25}$$

Now, recall the counting function $N(t)$, and apply this definition to Γ. We are counting the group elements that satisfy the condition:

$$N(t;z,w)=\#\{T\in\Gamma|dist(z,Tw)\leq t\}. \tag{26}$$

By (16),

(27)

$$G(s;z,w)=l_0+\int_0^\infty(\cosh t)^{-s}dN(t)=l_0+\int_0^\infty e^{-st}dN(t(\tau))=s\int_0^\infty e^{-st}N(t(\tau))d\tau,$$

where l_0 is the number of group elements $T\in\Gamma$ such that $dist(z,Tw)=0$, in the second equality the variable $\cosh t$ is changed to e^τ, and then we can derive the third equality by using integration by parts.

[15] See [23], Section 1.1 *The Gamma Function: functional equations*, page 3. In a subsection: *the multiplication theorems*, as $m=3$, we can obtain $\Gamma(2z)=\pi^{-1/2}2^{2z-1}\Gamma(z)\Gamma(\frac{1}{2}+z)$.

Moreover, let's apply Wiener-Ikehara's tauberian theorem, that we have used in proving PNT, on (25) and (27), which gives an asymptotic behavior that only depends on the topological invariant–genus, g:

$$N(t(\tau))(2g-2) \sim e^{\tau} = \cosh(t(\tau)) \tag{28}$$

as $\tau \to \infty$. Hence,

$$N(t) = N(t; z, w) \sim \frac{e^t}{2(2g-2)}, \text{ as } t \to \infty. \tag{29}$$

Now, if we use the notion in linear algebra, and take the trace on $\mathcal{K}_M$, that is $tr\mathcal{K}_M = \sum_{n=0}^{\infty} h(r_n)$. By using the length trace formula, theorem 9.2.10 in [5], which has a same derivation route we used in deriving (18).

(30)

$$tr\mathcal{K}_M = \sum_{n=0}^{\infty} h(r_n) = 4\pi(g-1) + \sum_{\gamma \in \ell(M)} \frac{\ell(\gamma_0)}{\sqrt{\cosh \ell(\gamma) - 1}} \int_{\ell(\gamma)}^{\infty} \frac{(\cosh \rho)^{-s} \sinh \rho d\rho}{\sqrt{\cosh \rho - \sinh \ell(\gamma)}}$$

where γ_0 is the unique primitive closed geodesic without lifting (so $\gamma = \gamma^n$, for some $n \geq 1$). We change the variable with $\cosh \rho := \cosh \ell(\gamma) + \frac{1}{2}x^2 e^{-\ell(\gamma)}$, and then apply (21) so that we can obtain:

$$tr\mathcal{K}_M = 4\pi(g-1) + \frac{\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} \sum_{\gamma \in \ell(M)} \Lambda(\Gamma)(\cosh \ell(\gamma))^{1/2-s}(\cosh \ell(\gamma))^{-1/2} \tag{31}$$

let $\Lambda^{\star}(\gamma) := \Lambda(\gamma)(\cosh \ell(\gamma))^{1/2}(\cosh \ell(\gamma) - 1)^{-1/2}$

$$= 4\pi(g-1) + \frac{\Gamma(s-\frac{1}{2})\sqrt{\pi}}{\Gamma(s)} \sum_{\gamma \in \ell(M)} \Lambda^{\star}(\Gamma)(\cosh \ell(\gamma))^{-s} \tag{32}$$

By (24),

$$\sum_{\gamma \in \ell(M)} \Lambda^{\star}(\Gamma)(\cosh \ell(\gamma))^{-s} = \frac{2}{s-1} + f(s)$$

where $f(s)$ is holomorphic on half-plane $\Re(s) > (1-\delta)$, and we have come back to what we did in proving Prime Number Theorem.

Let $\Psi^{\star}(t) = \sum_{\ell(\gamma) \leq t} \Lambda^{\star}(\gamma)$, then by lemma 6, we have $\Psi^{\star}(t) = O(te^t)$ as $t \to \infty$, and furthermore, $\Re(s) > 1 \Rightarrow t(\tau) = \cosh^{-1}(e^{\tau})$, which is useful when we apply the

Wiener-Ikehara theorem in the following integration.

$$\sum_{\gamma\in\ell(M)} \Lambda^\star(\gamma)(\cosh\ell(\gamma))^{-s} = \int_0^\infty (\cosh t)^{-s} d\Psi^\star(t) = \int_0^\infty e^{-s\tau} d\Psi^\star(t(\tau))$$

$$= s\int_0^\infty e^{-st}\Psi^\star(t(\tau))d\tau \tag{33}$$

After applying Wiener-Ikehara theorem as we did in the proof of PNT, here $\Psi^\star(t(\tau)) \sim 2e^\tau, \Rightarrow \Psi(t) = \sum_{\ell(\gamma)\leq t}\Lambda(\gamma) \sim e^t$ as $\tau\to\infty$.
The next step is to let $\eta > 0$ be the lower bound of the norms of the closed geodesics of M, and let $m(t) = \lfloor \frac{t}{\eta}\rfloor$. Thus, the two Chebyshev functions can be related by

$$\Psi(t) = \theta(t) + \sum_{m=2}^{m(t)} \theta(\frac{t}{m}) \leq m(t)\theta\left(\frac{t}{2}\right) \leq \frac{t}{\eta}\Psi\left(\frac{t}{2}\right) = O(te^{\frac{t}{2}}), t\to\infty \tag{34}$$

Hence,

$$\theta(t) \sim e^t, t\to\infty. \tag{35}$$

Lastly, we use the same method that we used to derive (12), and obtain the following

$$\Pi(t) = \int_{\frac{\eta}{2}}^t \frac{d\theta(\tau)}{\tau} \sim \frac{e^t}{t} \Rightarrow \pi(x) \sim \frac{x}{\log x}, \tag{36}$$

where $x = e^t$. □

4. Conclusion

In the previous two sections we have seen how number theory is connected to geometry. These two subjects is branched after Gauβ (actually there is one more great idea originated from him, that is the connections between linking numbers, closed geodesics (and this time we focus on knots), Legendre symbol and quadratic reciprocity, and the integration loop in Biot-Savart Law). Based on the study of linking numbers and closed geodesics, Gauβ developed now it's called Gauβ's law in electromagnetism and in vector analysis. This is also a good reason to seek the geometrical meaning of numbers which can help us to solve number theory problems, for example, the geometry of continued fractions can help us to solve Diophantine approximation. On the contrary, the notions in number theory can also help us to solve problems in geometry. For example, once we code the geometry by using continued fraction, then we can show that any two geodesics with the same coding are the same. Furthermore, this coding idea may become a foundation to help us to count the growth rate of non-simple closed geodesics directly. Moreover, once we get there, it would be interesting to look back to see what's the number theory interpretation on this.

Additionally, there is one more connection between PNT and PGT is the Möbius inversion function. We can use this function to reconstruct Witten index[28]. On one hand, Witten index is deriving by taking the trace of the Hamiltonian operator on Hilbert space, that is to sum over eigenvalues, and the meaning of it is the bose-fermi cancellation in non-singlet representation of the supersymmetry algebra. Thus, it can tell the difference of the two kinds of number: the number of bosons and the number of fermions. The essential difference of these two particles is parity. A fermion has odd parity, and a boson has even parity. Hence, we can use fermions to construct bosons, like to use primes to build composite numbers, but not in the other way around. On the other hand, Witten index is govern by Euler characteristic numbers, and it's a topological invariant that only depends on the system that describe by the Hamiltonian (for example, we can apply it to Laplace-Beltrami operators on Hilbert spaces as in billiard models) where we use Witten index to take the trace. *The connection between the PNT and PGT is: since the Atiyah-Singer index theorem is equal to the Witten index and thus relates index theorems to not only supersymmetry, but also Number Theory.* Hence, this is the fourth[16] point of view we can take when study the connections between PNT-PGT, and Number Theory-Geometry.

5. Appendix

Lemma 1. *If $s \in A$, then the Dirchlet sum of the Euler zeta function converges.*

Proof. Consider $s = \sigma + it \in A$, and let $n \in \mathbb{N}$, then we can have $n^s = n^t n^{i\sigma} = n^t \exp\{i(\sigma \log n)\}$ where $\sigma \log n \in \mathbb{R}$. Hence, if $\sigma \in A$, we obtain $|\zeta(s)| = \left|\sum_{n=1}^{\infty} \frac{1}{n^s}\right| = \left|\sum_{n=1}^{\infty} \frac{\exp\{-i\sigma \log n\}}{n^t}\right| \leq \sum_{n=1}^{\infty} \left|\frac{\exp\{-i\sigma \log n\}}{n^t}\right| = \sum_{n=1}^{\infty} \frac{1}{n^t} \leq 1 + \int_1^{\infty} \frac{1}{a^t} da = 1 + \frac{1}{t-1}$. It follows that the summation converges, if $t > 1$, that is the open half-plane, and so the Euler zeta function is well-defined if s is in the domain set A. □

Lemma 2. ***Euler product formula***

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{i=1}^{\infty} \frac{1}{1 - p_i^{-s}} \tag{37}$$

---

[16]**The first is Huber's method, the second is by Selberg's original viewpoit, the third is to code closed geodesics by using Pell's fundamental solution, continued fractions, and ergodic billiards.**

where p_i are prime. The origin of this formula is in 1737, Euler discovered a theorem that if function $f : \mathbb{N} - \{0\} \to \mathbb{C}$ is a **completely multiplicative function**[17], and then he considered $f(m) = m^s, s \in \mathbb{C}$. Then if the series $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent, then we can prove the above claim by using the property of multiplicative function, geometric series, and then rearrange the terms on the right hand side by collecting the terms as the method used in Combinatorics when dealing with generating functions. After derive the Euler product formula, Euler derived the following:

$$\log \zeta(s) = \sum_p \log \left(1 - \frac{1}{p^s}\right)^{-1} = \sum_p \frac{1}{p^s} + \sum_p \sum_{m=2}^{\infty} \frac{1}{mp^{ms}} = \sum_p \frac{1}{p^s} + R(s) \tag{38}$$

where $0 < R(s) < \frac{1}{2}$.[18]

Lemma 3. *Let S be a compact hyperbolic surface and let γ_1 be a homotopically non-trivial closed curve on S, then γ_1 is freely homotopic to a* ***unique*** *closed geodesic γ.[4]*

Lemma 4. *Let S be a compact hyperbolic surface, and $L > 0$. Only finitely many closed geodesics on S have length $\leq L$.*

Proof. This can be proved by contradiction, by considering there are infinite pairwise different closed geodesics $\gamma_1, \gamma_2, ...$ on S of length $\ell \leq L$. Then by using compactness, we can have finitely many coordinate neighborhood for each point in S. So we can form a convex neighborhood, by considering points with distance less than a certain amount, say $4r$. Then, within that neighborhood, for two distinct geodesics, we can have $dist(\gamma_n(t), \gamma_k(t)) \leq r, \forall t \in [0, 1]$. Thus, γ_n, and γ_k are homotopic. By the previous lemma, $\gamma_n = \gamma_k$, and we reach a contradiction. □

Lemma 5. *Let B be a real or complex parameter space. Assume the generating function $K(\rho, b) : [0, \infty) \times B \to \mathbb{C}$ is an even function belonging to $C^{\nu,\lambda}([0, \infty), B; \mathbb{C})$ and $|K^{(n,l)}(\rho, b)| \leq ce^{-\rho(1+\delta)}$ on $[0, \infty) \times B$. $n = 0, ..., \nu$, $l = 0, ..., \lambda$, where c and $\delta \in \mathbb{R}^+$. Note that $C^{\nu,\lambda}$ means K is ν (or λ) times differentiable w.r.t. ρ (or b).*

Lemma 6. *Let $\Phi(L)$ be the number of closed geodesics of length $\ell \leq L$ on a compact Riemann surface of genus $g \geq 2$. Then Φ has growth rate $\Phi(L) = O(e^L)$ as $L \to \infty$.*

[17] f satisfies the requirement: $f(mn) = f(m)f(n)$, $\forall m, n \in \mathbb{N} - \{0\}$. If f is only multiplicative function (not completely multiplicative), then $m \perp n$, i.e., m, n must be coprime.

[18] $R(s) < \sum_{p_i} \sum_{m=2}^{\infty} \frac{1}{2p_i^m} = \frac{1}{2} \sum_{p_i} \frac{1}{p_i(p_i - 1)} = \frac{1}{2}$.

6. References

[1] Atle Selberg. "An Elementary Proof of the Prime-Number Theorem." (1949).
[2] G. H. Hardy, E. M. Wright, and A. Wiles."An Introduction to the Theory of Numbers." (2008).
[3] Bent E. Petersen. "Prime Number Theorem." (1996).
[4] William P. Thurston. "The Geometry and Topology of Three-Manifolds." (2002).
[5] P. Buser. "Geometry and Spectra of Compact Riemann Surfaces." (1992).
[6] Darin Brown. "Lifting Properties of Prime Geodesics." (2009).
[7] Mark Pollicott. "Dynamical Zeta Function." (2010).
[8] C. Walkden. "Hyperbolic Geometry: Lec 6, The Poincaŕe disc model." (2017).
[9] Toshikazu Sunada. "On the Number-Theoretic Method in Geometry: Geometric Analogue of Zeta and L-functions and its applications." (1994).
[10] Svetlana Katok. "Coding of Closed Geodesic After Gauss and Morse." (1996).
[11] Albert Chang. "Isometries of the Hyperbolic Plane." (2010).
[12] David Sprunger. "Fuchsian Groups: Intro."
[13] Yingchun Cai. "Prime Geodesic Theorem." (2002).
[14] C. Johson. "Continued fractions and geodesics on the modular surface." (2013).
[15] David Borthwick. "Spectral Theory on Hyperbolic Surfaces." (2010).
[16] Y. N. Petridis. "Geodesics in Hyperbolic Space and Number Theory." (2006).
[17] Peter Sarnak. "Chaos, Quantum Mechanics, and Number Theory." (2011).
[18] Daniel Bump. "Spectral Theory and the Trace Formula." (1997).
[19] Frank Steiner. "Quantum Chaos." (1994).
[20] Shyamoli Chaudhuri. "Ultraviolet Limit of Open String Theory." (2005).
[21] Jens Marklof. "Selbergs trace formula: an introduction." (2008).
[22] Benson Farb and Dan Margalit. "A Primer on Mapping Class Groups. Ch 1. Curves, Surfaces, and Hyperbolic Geometry." (2012).
[23] W. Magnus, F. Oberhettinger, and R. P. Soni. "Formulas and Theorems for the Special Functions of Mathematical Physics." (1966).
[24] Jeffrey Stopple. "A Reciprocity Law for Prime Geodesics." (1988).
[25] Masanori Morishita. "Knots and Primes." (2009).
[26] Y. G. Sinai. "On Higher Order Spectral Measures of Ergodic Stationary Processes ." (1963).
[27] Caroline Series. "Continued Fractions and Hyperbolic Geometry." (2015).
[28] Donald Spector. "Supersymmetry and the Mbius inversion function." (1990).
[29] Pommersheim, Marks, and Flapan. "Number Theory: A Lively Introduction with Proofs, Applications, and Stories." (2010).
[30] Caroline Series. "The Geometry of Markoff Numbers". (1985).