L-FUNCTIONS

YIANNIS N. PETRIDIS

1. From the divisor problem to the Riemann zeta function

1.1. **The divisor function.** In number theory we encounter arithmetic sequences that behave rather irregularly, e.g. they are not increasing or decreasing themselves. Their values and growth quite often depends on the divisibility properties of the index. The simplest such function is the divisor function $d(n) = |\{a \in \mathbb{N}, a|n\}|$. If $n = p$, a prime number, then $d(p) = 2$, while $d(p^k) = k + 1$, i.e. it is constant on the prime numbers, while of the powers of a prime p^k it increases roughly like $\log_p(p^k)$. Even though we cannot see a specific order of growth in the divisor function by looking at the individual terms, we may ask about the average order of growth of the divisor function, i.e. we can investigate

$$\frac{1}{x}\sum_{n\le x} d(n).$$

Writing $n = ab$ for every divisor a of n, we calculate

$$\sum_{n\le x} d(n) = \sum_{n\le x}\sum_{a|n} 1 = \sum_{ab\le x} 1 = \sum_{a\le x}\sum_{b\le x/a} 1 = \sum_{a\le x}\left[\frac{x}{a}\right] = \sum_{a\le x}\left(\frac{x}{a} + O(1)\right),$$

as the fractional part of a number is in $[0, 1)$. Recall that the notation $f(x) = O(g(x))$ means that there exists a constant K such that $|f(x)| \le Kg(x)$ for all x. Clearly $f_i(x) = O(g_i(x))$ for $i = 1, 2$ gives $f_1(x) + f_2(x) = O(g_1(x) + g_2(x))$. We get now

$$\sum_{n\le x} d(n) = x\sum_{a\le x}\frac{1}{a} + O(x).$$

An standard idea from analysis is to compare the sum with the integral $\int_1^x dt/t$. By using left-hand sums and right-hand sums we see that, for any positive continuous decreasing function $f(x)$:

$$\sum_{n=2}^{N} f(n) \le \int_1^N f(t)\,dt \le \sum_{n=1}^{N-1} f(n)$$

Date: June 30, 2008.

2000 *Mathematics Subject Classification.* Primary 11F67; Secondary 11F72, 11M36.

The author was partially supported by NSF grant DMS-0401318, and PSC CUNY Re- search Award, No. 69288-00-38.

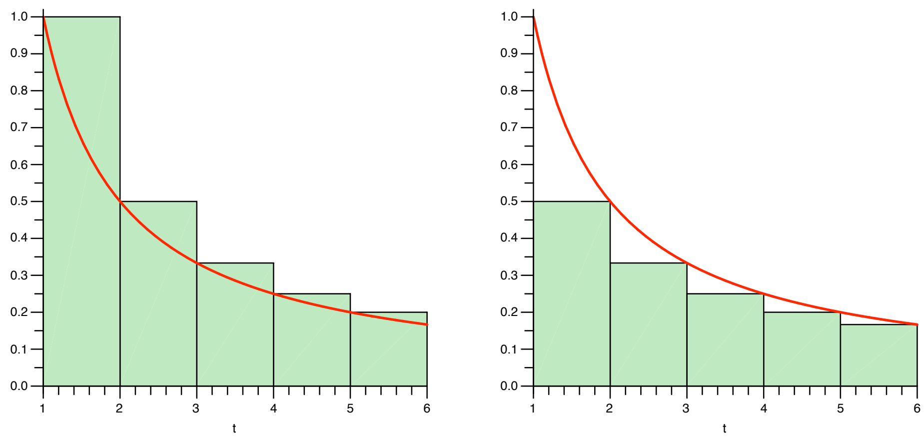


FIGURE 1. Riemann sums and the integral $\int_1^6 dt/t$

This is the main ingredient in the proof of the integral test for series. This estimate shows that up to $O(1)$ (i.e. bounded error) the sum $\sum_{n=1}^N f(n)$ and the integral $\int_1^N f(t)\,dt$ are the same. This gives

$$\sum_{n\le x} d(n) = x\sum_{a=1}^{[x]}\frac{1}{a}+O(x) = x\int_1^{[x]}\frac{dt}{t}+O(x) = x\int_1^{x}\frac{dt}{t}+O(x) = x\ln x+O(x). \tag{1.1}$$

A technique invented by Dirichlet (hyperbola principle) allows to estimate

$$\sum_{n\le x} d(n) = x\log x + (2\gamma-1)x + O(\sqrt{x}),$$

where γ is the Euler constant

$$\gamma = \lim_N\left(1+\frac{1}{2}+\cdots+\frac{1}{N}-\ln N\right).$$

For details look [5, Exercise 2.4.2]. A second arithmetic function one meets is the divisor sum $\sigma(n)=\sum_{a|n} a$, or, even, $\sigma_k(n)=\sum_{a|n} a^k$. In principle, we can follow similar techniques. There is, however, something unsatisfactory in these calculations. They seem easy but apply to the specific arithmetic function at hand. It would be nice to have a more general technique that can apply to estimate the order of growth of an arithmetic function.

In analytic number theory we associate generating functions to interesting sequences that behave irregularly. In the absence of special structure of the sequence

a_n we associate with it the Taylor series

$$\sum_n a_n x^n,$$

which has positive radius of convergence if a_n do not grow too quickly e.g. $a_n = O(k^n)$ for some k. In the case where a_n have multiplicative properties, it is more natural to associate the Dirichlet series

$$D(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} \tag{1.2}$$

since the function n^s satisfies the obvious relation $n_1^s n_2^s = (n_1 n_2)^s$. Hopefully the series $D(s)$ will converge for certain s. The exact notion of multiplicative properties usually takes one of the following two forms:

(1) $a_{mn} = a_m \cdot a_n$ for $(n, m) = 1$, where (n, m) is the greatest common divisor of m, n,
(2) $a_{mn} = a_m \cdot a_n$ for all n, m.

In the first case we say that the sequence a_n is multiplicative and in the second that it is completely multiplicative. While it may seem that the second is more desirable, there are many arithmetic functions that do not satisfy the second, e.g. the divisor function is not completely multiplicative: $d(8) = 4 \neq d(2)d(4) = 2 \cdot 3$. In this case the associated Dirichlet series is

$$D(s) = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}. \tag{1.3}$$

It turns up that this function can be factored! In fact, it is exactly $\zeta(s)^2$, where

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{1.4}$$

is the celebrated Riemann zeta function. This series converges absolutely for $\sigma = \Re(s) > 1$. We first use the comparison test, since

$$\left| \frac{1}{n^s} \right| = \frac{1}{n^\sigma}.$$

Then we use the integral test:

$$\int_1^{\infty} \frac{dt}{t^\sigma} = \left[\frac{t^{-\sigma+1}}{1-\sigma} \right]_1^{\infty} = \frac{1}{\sigma - 1}.$$

(For $0 < s \leq 1$ the same calculation shows that the series diverges, while for $\Re(s) \leq 0$ the general term does not even tend to 0). The calculation $D(s) = \zeta(s)^2$ is actually very easy (formally):

$$D(s)=\sum_{n=1}^{\infty}\frac{d(n)}{n^s}=\sum_{n=1}^{\infty}\sum_{ab=n}\frac{1}{a^sb^s}=\sum_{a,b=1}^{\infty}\frac{1}{a^sb^s}=\sum_{a=1}^{\infty}\frac{1}{a^s}\sum_{b=1}^{\infty}\frac{1}{b^s}=\zeta(s)\zeta(s)=\zeta(s)^2.$$

A careful eye may notice that we used a rearrangement of the double series $\sum_{a,b}a^{-s}b^{-s}$, so that we order the pairs (a,b) according to their product $n=ab$. This is allowed as a consequence of the Fubini theorem for absolutely convergent series. See [9, 7.50]. A question that arises is for which s the Dirichlet series (1.3) converges. We include a general result that applies to a wide range of Dirichlet series. We define $A_n=\sum_{j=1}^n a_n$.

Theorem 1.1. [8, 9.12–9.14] *Assume that $\sum a_n$ is divergent. Define*

$$\sigma_0=\limsup\frac{\log|A_n|}{\log n}.$$

Then the Dirichlet series (1.2) converges for $\Re(s)>\sigma_0$ and diverges for $\Re(s)<\sigma_0$.

It follows that, if

$$\bar{\sigma}=\limsup\frac{\log(|a_1|+|a_2|+\ldots+|a_n|)}{\log n},$$

then the series (1.2) converges absolutely for $\Re(s)>\bar{\sigma}$ and is not absolutely convergent for $\Re(s)<\bar{\sigma}$. The number σ_0 is the abscissa of convergence and $\bar{\sigma}$ is the abscissa of absolute convergence (clearly $\sigma\le\bar{\sigma}$). For (1.3) we conclude that it converges absolutely for $\Re(s)>1$, using (1.1):

$$\lim_n\frac{\log\sum_1^n d(k)}{\log n}=\lim_n\frac{\log(n\log n)}{\log n}=1.$$

Recall that the radius of convergence of the power series $\sum a_nz^n$ is

$$R=\frac{1}{\limsup\sqrt[n]{a_n}}.$$

1.2. **The Riemann zeta function.** So far it all seems to involve calculations with series. What is not obvious is (i) what properties of $\zeta(s)$ are important in the asymptotics of the divisor function, (ii) what is the relation of $\zeta(s)$ with prime numbers.

For the first we first explain the notion of analytic continuation in a simple example. The geometric series

$$f(z)=\sum_{n=0}^{\infty}z^n$$

converges for $|z|<1$ and is equal in this region to the function $g(z)=1/(z-1)$. The function $g(z)$ is analytic on $\mathbb{C}$ with the exception of a simple pole at $z=1$. The function $f(z)$, which was initially defined in the open unit disc only, is said to have an analytic continuation in $\mathbb{C}$ with simple pole at $z=1$. The analytic continuation of $f(z)$ is certainly not given by the series $\sum z^n$ when $|z| \geq 1$. It cannot, as the series does not converge in this region. It is given by the equation $f(z) = g(z)$. What is nice in this example is that we can write a simple formula for the analytic continuation. But this is not always possible and certainly it is not necessary. The Gamma function is a good example of this. Moreover, it is important for the study of $\zeta(s)$. The Gamma function $\Gamma(s)$ is the generalisation of the factorial ($\Gamma(n) = (n-1)!$, $n \in \mathbb{N}$). It is given by

$$\Gamma(s) = \int_0^\infty e^{-t} t^s \frac{dt}{t}, \quad \Re(s) > 0. \tag{1.5}$$

One need to prove that the integral converges at $t = \infty$ and $t = 0$ and it is at the second point, where the condition $\Re(s) > 0$ is used together with integration by parts:

$$\Gamma(s) = \left[\frac{t^s}{s} e^{-t}\right]_0^\infty + \frac{1}{s} \int_0^\infty e^{-t} t^{s+1} \frac{dt}{t} = \frac{1}{s}\Gamma(s+1).$$

This method proves also the functional equation:

$$\Gamma(s) = \frac{1}{s}\Gamma(s+1). \tag{1.6}$$

The analytic i.e. meromorphic continuation of the Gamma function to $\mathbb{C}$ follows from this as follows: The right-hand side of the equation (1.6) makes sense for $\Re(s+1) > 0$, i.e. $\Re(s) > -1$. So we can define $\Gamma(s)$ in the strip $0 \geq \Re(s) > -1$. Then we continue the process to define $\Gamma(s)$ in the strip $-1 \geq \Re(s) > -2$, using the right-hand side of the equation (1.6). The process continues to extend $\Gamma(s)$ on successive vertical strips to the left, covering $\mathbb{C}$. Moreover, this process gives us where the poles of $\Gamma(s)$ occur: at $s = 0$ we have the first pole and by this process we find poles at all negative integers. We can calculate the residues e.g.

$$\mathrm{Res}_{s=0}\Gamma(s) = \Gamma(1) = \int_0^\infty e^{-t} dt = 1.$$

We summarize the analytic i.e. meromorphic behaviour of $\zeta(s)$ in the following theorem.

Theorem 1.2. *The Riemann zeta function, given by Eq. 1.4 for $\Re(s) > 1$ is holomorphic in this region. It has analytic continuation to the whole complex plane $\mathbb{C}$ with the exception of one simple pole at $s = 1$. The residue at the pole is also* 1. *The zeta function satisfies the functional equation*

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

Remark. It is more convenient to define $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$, so that the functional equation takes the form $\xi(s) = \xi(1-s)$.

There are many proofs of the analytic continuation of $\zeta(s)$. Titchmarsh [7] lists seven methods and there are variants within each! Our choice of proof is due to two facts: (i) it is one of the original proofs of Riemann, (ii) it generalises to number fields, see Lang, Algebraic Number Theory, p. 252–258.

We work through the various parts of the proof. The holomorphic nature of $\zeta(s)$ for $\Re(s) > 1$ is obvious, as it is the sum of uniformly convergent series of holomorphic functions $1/n^s$. Uniformity of convergence on compact sets of the region follows by comparison with $\zeta(\sigma)$.

The proof of the analytic continuation and functional equation we will give uses the Poisson summation formula, which roughly says that for sufficiently smooth and decaying functions:

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \hat{f}(m). \tag{1.7}$$

Here we define the Fourier transform by

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

For a precise statement see Exercise 8. We actually use it only for $f(x) = e^{-\pi x^2}$ the Gaussian function. This is a function which is equal to its Fourier transform. We have:

$$e^{-\pi\xi^2} = \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i\xi x}\, dx. \tag{1.8}$$

If you have never seen this, here is a proof: use contour integration for $e^{-\pi z^2}$ on the rectangle $[-R, R] \times [0, \xi]$, for $\xi > 0$ and use the standard integral from calculus

$$\int_{-\infty}^{\infty} e^{-\pi x^2}\, dx = 1.$$

For details look [6, p. 42–43]. We substitute $x = \sqrt{t}x'$ and $\xi\sqrt{t} = \xi'$ to (1.8) to get that the Fourier transform of $f(x) = e^{-\pi x^2 t}$ is (t is a positive parameter)

$$\hat{f}(\xi) = \frac{1}{\sqrt{t}} e^{-\xi^2\pi/t}.$$

We apply the Poisson summation formula (1.7) to this pair to get for the theta function

$$\theta(t) = \sum_{n\in\mathbb{Z}} e^{-n^2\pi t} = \frac{1}{\sqrt{t}} \sum_{n\in\mathbb{Z}} e^{-\pi n^2/t} = \frac{1}{\sqrt{t}}\theta(t^{-1}).$$

We go back to the Gamma function and substitute in the Euler integral (1.5) $t = n^2\pi t'$ to get

$$\Gamma(s/2) = \int_0^{\infty} e^{-\pi n^2 t}(\pi n^2 t)^{s/2} \frac{dt}{t} \Longrightarrow \pi^{-s/2}\Gamma(s/2)n^{-s} = \int_0^{\infty} e^{-\pi n^2 t} t^{s/2} \frac{dt}{t}.$$

We sum over $n \in \mathbb{N}$ and set $\psi(t) = \sum_1^\infty e^{-\pi n^2 t}$ to get

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

$$= \int_0^1 \frac{1}{2}(t^{-1/2}\theta(t^{-1}) - 1)t^{s/2}\frac{dt}{t} + \int_1^\infty \psi(t)t^{s/2}\frac{dt}{t} = \int_1^\infty \frac{1}{2}(u^{1/2}\theta(u) - 1)u^{-s/2}\frac{du}{u} + \int_1^\infty \psi(t)t^{s/2}\frac{dt}{t}$$

(with the change of variables $u = t^{-1}$)

$$= \int_1^\infty \psi(u)(u^{(1-s)/2} + u^{s/2})\frac{du}{u} + \frac{1}{2}\int_1^\infty u^{(1-s)/2} - u^{-s/2}\frac{du}{u} = \int_1^\infty \psi(u)(u^{(1-s)/2} + u^{s/2})\frac{du}{u}$$

$$+\frac{1}{s-1} - \frac{1}{s}.$$

The improper integral here converges for all $s \in \mathbb{C}$: for $u \geq 1$ we have

$$\psi(u) \leq \sum_1^\infty e^{-n\pi u} = \frac{e^{-\pi u}}{1 - e^{-\pi u}}.$$

Such a convergent integral with integrand depending holomorphic on the complex parameter s defines a holomorphic function, see [8, 2.83–2.84]. Moreover, we see that $\xi(s)$ has poles at 0 and 1. Therefore, $\zeta(s)$ has a meromorphic continuation with simple pole at 1 with residue 1 (here we need that $\Gamma(1/2) = \pi^{1/2}$, which follows from $\int_0^\infty t^{-1/2}e^{-t}dt = \int_0^\infty u^{-1}e^{-u^2}2udu = \int_{-\infty}^\infty e^{-u^2}du = \sqrt{\pi}$). This is the only pole as $\Gamma(s/2)$ has no zeros (well-known property of the Gamma function). We also see the functional equation as the right-hand side is invariant under $s \to 1-s$. The poles of $\Gamma(s)$ at the negative integers force $\zeta(s)$ to have zeros at $-2, -4, \ldots$. These are called the trivial zeros. However, the pole of $\Gamma(s)$ at 0 does not force a zero, because of the $-1/s$ in the equation above. In fact, $\zeta(0) = -1/2$.

Recall the notation $h(x) \sim g(x)$, as $x \to \infty$: it means that $f(x)/g(x) \to 1$, as $x \to \infty$. There is a general technique in analytic number theory to study the distribution on average of a sequence from its generating function. This comes under the subject of tauberian theorems. A simple and very useful one is the following:

Theorem 1.3 (Ikehara-Wiener [5])**.** *Let*

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

with $a_n \geq 0$. Assume that the Dirichlet series converges absolutely for $\Re(s) > 1$ and has an analytic continuation on $\Re(s) \geq 1$ with a simple pole at $s = 1$ with residue a and is holomorphic on other points on $\Re(s) = 1$. Then

$$\sum_{n \leq x} a_n \sim ax, \quad x \to \infty.$$

If the pole is of order $k > 1$ and the leading term in the Laurent expansion is $c_{-k}(s-1)^{-k}$, then

$$\sum_{n\le x} a_n \sim \frac{1}{(k-1)!} c_{-k} x (\log x)^{k-1}.$$

Remark. The function $f(s)$ has analytic continuation to the points on $\Re(s) = 1$ means that for each point s_0 with $\Re(s_0) = 1$, $s_0 \neq 1$ there is a neighborhood on which $f(s)$ has an analytic continuation. These neighborhoods may be shrinking as $\Im(s) \to \pm\infty$.

Remark. This theorem may seem to lack intuition. However, according to the Perron formula (see the exercises below) we have

$$\sum_{n\le x} a_n = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(s)\frac{x^s}{s} ds.$$

If we can deform the contour as in this exercise to include the pole at $s = 1$, the residue is exactly ax for a simple pole of $f(s)$. The contribution by the deformed contour 'should' be smaller. For this one needs to understand the order of growth of $f(s)$ to the left of $\Re(s) = 1$. This is tricky even for $\zeta(s)$. However, the theorem as stated does not require any knowledge to the left, only holomorphicity up to $\Re(s) = 1$ with the given pole at $s = 1$.

With this knowledge we can recover the asymptotics of the divisor function: $\zeta^2(s)$ has pole of order 2 at $s = 1$ with leading singularity $1/(s-1)^2$. This gives $\sum_{n\le x} d(n) \sim x \log x$.

Actually more refined information can be recovered by using contour integration and computing lower terms in the asymptotics of the form $x(\log x)^j$, $j < k-1$.

There is another very important property of the Riemann zeta function. It has an Euler product:

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

where the produce extends over all primes. Formally this is proves as follows for $\Re(s) > 1$:

$$\frac{1}{1-p^{-s}} = \sum_{k=0}^{\infty} p^{-ks}$$

by expanding the geometric series. Multiplying over all primes we get

$$\prod_p \frac{1}{1-p^{-s}} = \prod_p \sum_{k=0}^{\infty} p^{-ks} = \sum_n n^{-s} = \zeta(s),$$

because every integer n has a unique factorisation into prime powers, and distributing the product above gives all such products (to the exponent $-s$). The definition of the infinite product $\prod b_n$ is similar to infinite series:

$$\prod_{n=1}^{\infty} b_n = \lim_N \prod_1^N b_n,$$

provided that the limit is nonzero. As far as a rigorous proof of the infinite product of $\zeta(s)$, we fix P. Then

$$\prod_{p\le P}\left(1+p^{-s}+p^{-2s}+\cdots\right) = 1+\frac{1}{n_1^s}+\frac{1}{n_2^s}+\cdots,$$

where on the right-hand side we have the integers with prime factors $\le P$. All numbers $\le P$ are included in this list, so that

$$\left|\zeta(s)-\prod_{p\le P}\frac{1}{1-p^{-s}}\right| \le \sum_{n>P}\frac{1}{n^\sigma},$$

which is the tail of $\zeta(s)$. Consequently it tends to 0 for $\Re(s)>1$.

Remark. A slightly more sophisticated point is that an infinite product $\prod(1+a_n)$ converges iff $\sum\log(1+a_n)$ does. This last series converges absolutely iff $\sum|a_n|$ does. See [1, p. 191–192]. For $\Re(s)>1$ we have $|\sum_{k\ge1}p^{-ks}|\le 2p^{-\sigma}$, so in this case $\sum_p|\sum_{k\ge1}p^{-ks}|\le 2\zeta(\sigma)$.

The infinite product of the Riemann zeta function lies at the heart of its relation to the prime numbers and the prime number counting function

$$\pi(x) = |\{p, p\le x\}|.$$

We differentiate the logarithmic derivative of $\zeta(s)$ given as the Euler product to get:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{p^{-s}\log p}{1-p^{-s}} = \sum_p \log p \sum_{m\ge1} p^{-ms}.$$

The last series counts the prime powers p^m with weight $\log p$, i.e. it is the Dirichlet series associated to the von Mangoldt function

$$\Lambda(n) = \begin{cases} \log p, & n=p^m,\\ 0, & \text{otherwise.}\end{cases}$$

So

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_n \frac{\Lambda(n)}{n^s}.$$

Complex analysis tells us that the left-hand side has a pole of order 1 with residue 1 (notice the minus sign), as a pole of $f(s)$ of order k contributes a simple pole of

$f'(s)/f(s)$ with residue $-k$. If we can apply the Tauberian theorem, we immediately get

$$\sum_{n\leq x}\Lambda(n)\sim x,\quad \text{i.e.}\ \sum_{p^m\leq x}\log p\sim x. \tag{1.9}$$

It is not obvious why the conditions of the theorem apply, in particular, one needs to know that $\zeta(s)$ does not have zeros on $\Re(s)=1$, so that these do not give poles of $-\zeta'(s)/\zeta(s)$. This is due to de la Vallée Poussin. Eq. (1.9) is two steps away from the proof of the Prime Number Theorem (PNT). We first remove the powers of primes with $m>1$, as all such numbers p^m have $p\leq\sqrt{x}$ and their count $O(\sqrt{x}^{1+\epsilon})$, as $\log p=O(x^\epsilon)$. Then we have to remove the weight $\log p$ from the sum

$$\sum_{p\leq x}\log p.$$

This is done by a summation by parts. We get a simple version of PNT:

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

Remark. The argument that $\zeta(1+it)\neq 0$ is based of the inequality

$$\zeta(\sigma)^3|\zeta^4(\sigma+it)\zeta(\sigma+2it)|\geq 1,$$

for $\sigma>1$. This follows from the inequality

$$3+4\cos\theta+\cos(2\theta)=2(1+\cos\theta)^2\geq 0.$$

1.3. **The Riemann hypothesis.** It follows from the functional equation that if ρ is a nontrivial zero, so is $1-\rho$. Moreover, since $\overline{\zeta(\bar{s})}=\zeta(s)$, which follows from the fact that $\zeta(\sigma)\in\mathbb{R}$, $\sigma>1$, we have that $\bar{\rho}$, $1-\bar{\rho}$ are also nontrivial zeros.

The Riemann hypothesis (RH) is the statement that all the nontrivial zeros ρ of $\zeta(s)$ lie on the critical line $\Re(s)=1/2$. It gives the most symmetric location for the zeros. It is important for many reasons. The most obvious has to do with the distribution of the prime numbers. The smaller the order of growth of $\pi(x)-\mathrm{li}(x)$ the 'smoother' the approximation of the discontinuous function $\pi(x)$ by $\mathrm{li}(x)$. The easiest way to see this is the following theorem

Theorem 1.4. [4] *Suppose that $\Re(\rho)\leq\theta$ for all nontrivial zeros and $\theta<1$. Then*

$$\sum_{n\leq x}\Lambda(n)=x+O(x^\theta(\log x)^2)$$

Conversely, if for $\alpha<1$ we have

$$\sum_{n\leq x}\Lambda(n)=x+O(x^\alpha)$$

then all the nontrivial zeros has $\Re(\rho)\leq\alpha$.

If RH is true we can take $\theta = 1/2$.

Quite often RH is a working hypothesis. We try to prove a statement assuming it is true to see how far the results can go, even if RH is still unproven. Sometimes one can prove the results afterwards without assuming RH.

2. Dirichlet L-series

To capture divisibility properties of integers one can introduce

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

where $\chi(\cdot)$ is a multiplicative character modulo N, i.e.

$$\chi(ab) = \chi(a)\chi(b), \quad (a, N) = 1, (b, N) = 1,$$

while we set $\chi(a) = 0$, if $(a, N) > 1$. Notice that this is a completely multiplicative function. In practice the construction of χ can be done as follows: Let $N = q$ be a prime, then the multiplicative group modulo q, i.e $(\mathbb{Z}/q\mathbb{Z})^*$ is cyclic with generator, say g, called a primitive root modulo p, then we pick a complex number $\chi(g)$ with $\chi(g)^{q-1} = 1$. Then we define $\chi(g^m) = \chi(g)^m$. The same can be done if q is a prime power p^l for $p \neq 2$, as there is a primitive root in this case (substitute $q-1$ with $\phi(q)$). For 2 and 4 we also easily identify the characters. But 2^l, $l > 2$ has no primitive root. For details look at [2, p. 28]. The simplest character is the trivial character: $\chi_0(n) = 1$, $(n, q) = 1$.

The convergence of $L(s,\chi)$ for the other characters can be determined as follows: since

$$\sum_{m=1}^{q-1} \chi(g)^m = \frac{1 - \chi(g)^{q-1}}{1 - \chi(g)} = 0$$

as a geometric sum. Consequently we have $\sum_{n \le N} \chi(n) = O(1)$, i.e. is bounded (in fact by $q-1$). Using Theorem 1.1, we see that the Dirichlet series converges for $\Re(s) > 0$. By sticking absolute values, we get that the domain of absolute convergence is the same as for $\zeta(s)$, i.e. $\Re(s) > 1$. The Dirichlet L-series is holomorphic for $\Re(s) > 0$ as a consequence. In particular there is no pole at $s = 1$. Such Dirichlet L-series were introduced by Dirichlet to prove the infinity of primes in an arithmetic progression

$$a, a + q, a + 2q, \dots,$$

where we have to assume $(a, q) = 1$. Let us assume that q is prime for simplicity. He proved that

$$\sum_{p \equiv a \,(\mathrm{mod}\ q)} \frac{1}{p^s} = \frac{1}{q-1} \sum_{\chi \,(\mathrm{mod}\ q)} \overline{\chi(a)} \log L(s,\chi) + O(1).$$

The important ingredient here is the use of orthogonality of characters

$$\frac{1}{q-1}\sum_{\chi \,(\mathrm{mod}\ q)} \overline{\chi(a)}\chi(n) = \begin{cases} 1, & n \equiv a \pmod q \\ 0, & \text{otherwise.} \end{cases}$$

The first line is obvious since $n \equiv a \pmod q$ means that $\chi(a) = \chi(n) \Longrightarrow \overline{\chi(a)} = \chi(n)^{-1}$. For the second line we write $\bar{a}$ the multiplicative inverse of $a \,(\mathrm{mod}\ q)$ and notice that $\bar{a}n \neq 1 \,(\mathrm{mod}\ q)$. We set $k = g^m = \bar{a}n$. We can assume also that $(n, q) = 1$. Then we can find a character $\psi(\cdot)$ with $\psi(\bar{a}n) \neq 1$. Just choose a $q-1$ root of 1, say ω with $\omega^m \neq 1$ and set $\psi(g) = \omega$. Then as χ ranges over all the characters $\pmod q$, so does $\psi\chi$. As a result

$$\sum_\chi \chi(k) = \sum_\chi (\psi\chi)(k) = \sum_\chi \psi(k)\chi(k) = \psi(k)\sum_\chi \chi(k) \Longrightarrow \sum_\chi \chi(k) = 0.$$

Dirichlet considered what can happen to $\log L(s, \chi)$ as $s \to 1$. The character χ_0 gives $L(s, \chi_0) = (1 - q^{-s})\zeta(s)$ and we know that this tends to ∞ as $s \to 1$. The rest of the characters are separated into the ones that take complex values (and they come in pairs) and the real character with $\chi(g) = -1$. Dirichlet could easily consider the complex characters and see that $L(1, \chi) \neq 0$. The real one, given by

$$\chi(n) = \left(\frac{n}{q}\right)$$

using the quadratic residue symbol, is much harder to treat.

As far as the analytic properties of the Dirichlet L-series we have the following theorem:

Theorem 2.1. [2, p. 68–71] *The L-functions have analytic continuation on $\mathbb{C}$ with* no *poles, unless $\chi = \chi_0$ the trivial character. Define the Gauss sum*

$$\tau(\chi) = \sum_{n=1}^{q} \chi(n) e^{2\pi i n/q}.$$

(a) If $\chi(-1) = 1$, then

$$q^{s/2}\pi^{-s/2}\Gamma(s/2)L(s, \chi) = \frac{\tau(\chi)}{\sqrt{q}} q^{(1-s)/2}\pi^{(1-s)/2}\Gamma((1-s)/2)L(1-s, \bar{\chi}).$$

The trivial zeros are now $0, -2, -4, \ldots$.

(b) If $\chi(-1) = -1$, then

$$q^{(s+1)/2}\pi^{-(s+1)/2}\Gamma((s+1)/2)L(s, \chi) = \frac{\tau(\chi)}{i\sqrt{q}} q^{(2-s)/2}\pi^{-(2-s)/2}\Gamma((2-s)/2)L(1-s, \bar{\chi}).$$

The trivial zeros are at $-1, -3, -5, \ldots$.

Remark. Concerning the sum $\sum_{n\le N}\chi(n)$ the bound $O(q)$ is far from optimal (in the q aspect). The Polya–Vinogradov inequality gives

$$\sum_{n=M+1}^{M+N}\chi(n) = O(q^{1/2}\log q).$$

See [2, p. 135]

The Dirichlet L-series also have an infinite product, much like $\zeta(s)$:

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

The proof is essentially the same as for $\zeta(s)$, using the complete multiplicativity of $\chi(\cdot)$.

3. The Gauss circle problem

The third sequence we would like to study is

$$r(n) = \#\{(a,b)\in\mathbb{Z}^2, n = a^2+b^2\}$$

which counts the number of ways of representing n as sum of two squares. At first glance it is not obvious that it is a multiplicative function. However, the Lagrange identity

$$(a^2+b^2)(c^2+d^2) = (ac+bd)^2 + (ad-bc)^2$$

(a simple consequence of $|z_1z_2| = |z_1|\,|z_2|$ with $z_1 = a+ib$, $z_2 = c+id$) shows that the product of two numbers which are sums of squares is a sum of squares. For a proof that $r(n)$ is multiplicative, look at the exercises. We are forced to consider

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

It becomes clear that we are interested in the norms of points in the complex plane with integer coordinates: $\mathbb{Z}^2$. This is a lattice in $\mathbb{R}^2$, i.e. a $\mathbb{Z}$-module in $\mathbb{R}^2$ of rank 2. There is an easy elementary estimate on $r(n)$ on average (this argument is due to Gauss): $\sum_{n\le x} r(n)$ counts the number of lattice points in a disc of radius $\sqrt{x}$ centered at the origin. Put a square of size 1 centered at any one of these points. These squares are quite often entirely inside the disc but some of them extend off the disc. However, they are all contained in a slightly larger disc of radius $\sqrt{x}+1/\sqrt{2}$, as follows from the triangle inequality: for a point (x,y) inside the square centered at (a,b) we have (notice that the diagonals of the squares have length $\sqrt{2}$)

$$|(x,y)-(a,b)| \le \frac{\sqrt{2}}{2}, \quad |(a,b)| \le \sqrt{x} \Longrightarrow |(x,y)| \le \sqrt{x} + \frac{1}{\sqrt{2}}.$$

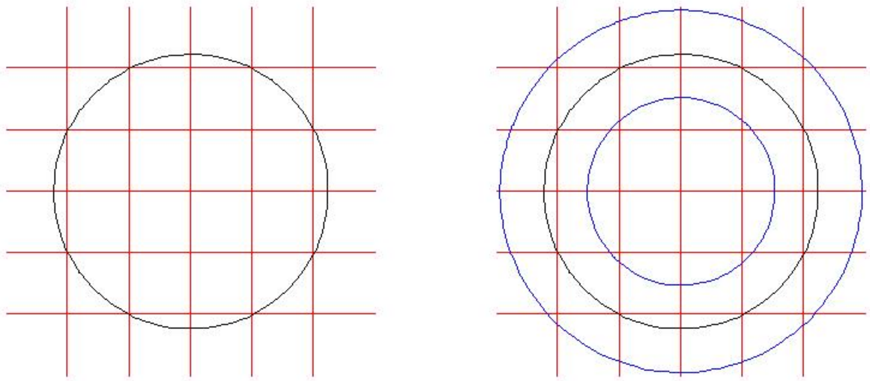


FIGURE 2. The Gauss circle problem

Since the area of a disc of radius R is πR^2, we get

$$\sum_{n\leq x} r(n) = \sum_{|(a,b)|\leq\sqrt{x}} 1 \leq \pi\left(\sqrt{x}+\frac{1}{\sqrt{2}}\right)^2 = \pi\left(x+\sqrt{2x}+\frac{1}{2}\right).$$

Actually one can go one step further. The disc of radius $\sqrt{x}-\frac{1}{\sqrt{2}}$ is contained entirely in the union of the squares defined above. Let (x,y) be such a point, i.e. $|(x,y)| \leq \sqrt{x}-1/\sqrt{2}$. Let $a=[x]$, $b=[y]$, so that (a,b) is a lattice point. The triangle inequality gives

$$|(a,b)| \leq |(x,y)| + \frac{1}{\sqrt{2}} \leq \sqrt{x}.$$

This gives that

$$\sum_{n\leq x} r(n) = \sum_{|(a,b)|\leq\sqrt{x}} 1 \geq \pi\left(\sqrt{x}-\frac{1}{\sqrt{2}}\right)^2 = \pi\left(x-\sqrt{2x}+\frac{1}{2}\right).$$

Together these results give the estimate, known to Gauss

$$\sum_{n\leq x} r(n) = \pi x + O(\sqrt{x}).$$

We recall the notation $f=O(g)$, which means that $|f(x)| \leq Kg(x)$ for some constant K and all x. The Gauss circle problem asks to estimate the remainder

$$R(x) = \sum_{n\leq x} r(n) - \pi x.$$

It is still open and the conjecture, due to Hardy, is that

$$R(x) = O_\epsilon(x^{1/4+\epsilon}).$$

Although it is not of direct importance to the theory of automorphic forms, which we wish to introduce, we remark that the first improvement to the Gauss estimate $R(x) = O(x^{1/2})$ is due to Sierpinski and Van der Corput (and uses crucially the Voronoi summation formula) and is $R(x) = O(x^{1/3+\epsilon})$. Many mathematicians worked in this problem. The best known result is due to Huxley [50] $R(x) = O(x^{131/416} \log^{18627/16640} x)$ and it is too complicated to explain here. Hardy's conjecture is not possible to improve, as $R(x) = \Omega(x^{1/4})$ (for the best Omega result look at Soundararajan).

4. Lattices and $\mathrm{SL}_2(\mathbb{R})$

The most general lattice in $\mathbb{R}^2$ is

$$L = \{n_1 w_1 + n_2 w_2, (n_1, n_2) \in \mathbb{Z}^2\},$$

where we have fixed two complex numbers w_1, w_2, which are linearly independent over $\mathbb{R}$. As far as the shape of the lattice is concerned, we can scale it and rotate it, so that $w_1 = 1$. Moreover, since $w \in L \Leftrightarrow -w \in L$, we can assume that $\Im(w_2) > 0$. We set $z = w_2$. The lattice is now

$$L = \{m + nz, (m, n) \in \mathbb{Z}^2\}.$$

A lattice or vector space has many bases and we learn early to switch from one to another. If the vectors w_1, w_2 are a basis of the lattice and w_1', w_2' is another basis, then they are related by an invertible 2×2 matrix A with integer coefficients

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

such that A^{-1} also has integer coefficients. Since $\det(A \cdot A^{-1}) = 1 = \det(A)\det(A^{-1})$ and these are integers, we must have $\det(A) = \pm 1$. If the bases have the same orientation $\det(A) = 1$. This leads us to consider the special linear group

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}.$$

In these lectures this ends up being the most important group and a prime example of the theory. We would like to visualize the change of bases in the shape of the lattice. We form the parallelogram with two adjacent sides w_1 and w_2 (or 1 and z). This is called the fundamental region. Two bases for the same lattice will produce different shapes of the fundamental region. If we insist that the first vector is 1, what happens to the second vector? To get the vector z from w_1 and w_2, we scaled the rotated the lattice and this is done by setting

$$z = \frac{w_2}{w_1}.$$

If we act on the basis $(w_2, w_1)^T$ by $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ to get $(w_2', w_1')^T$ then

$$w_2' = aw_2 + bw_1, \quad w_1' = cw_2 + dw_1$$

so that

$$\gamma \cdot z = z' = \frac{w_2'}{w_1'} = \frac{aw_2 + bw_1}{cw_2 + dw_1} = \frac{az+b}{cz+d}. \tag{4.1}$$

This is the action of a linear fractional transformation. The most general transformation of this has the form

$$\gamma z = \frac{az+b}{cz+d}, \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C}).$$

They have many well-known properties that are studied e.g. in complex analysis [1]. For instance, they are meromorphic functions with a simple pole at $-d/c$ and map lines and circles in the complex plane to lines and circles of the complex plane. They are conformal in the extended complex plane $\mathbb{C} \cup \{\infty\}$. Given two triples of points in the extended complex plane (z_1, z_2, z_3) and (w_1, w_2, w_3), there exists a unique l.f.t. T mapping $T(z_i) = w_i$, $i = 1, 2, 3$. We saw that in a lattice we can take a basis of 1 and z with $\Im(z) > 0$. Which linear fractional transformations preserve the positivity of the imaginary part of z, i.e. for $\Im(z) > 0$ we also have $\Im(\gamma z) > 0$. Such transformations should map the real line to itself, and, therefore can be written with real coefficients (see the exercises). Here is a good point to include the fundamental calculation for $a, b, c, d \in \mathbb{R}$:

$$\Im(\gamma z) = \frac{1}{2i}\left(\frac{az+b}{cz+d} - \frac{a\bar{z}+b}{c\bar{z}+d}\right) = \frac{1}{2i}\frac{ad(z-\bar{z}) - cd(z-\bar{z})}{|cz+d|^2} = \frac{(ad-bc)\Im(z)}{|cz+d|^2}. \tag{4.2}$$

In fact we can consider these transformations preserving $\Im(z) > 0$ as forming a group. Closely associated is the group

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\}.$$

We call the complex numbers with positive imaginary part the hyperbolic plane $\mathbb{H}$, i.e.

$$\mathbb{H} = \{z \in \mathbb{C}, \Im(z) > 0\}.$$

The group $\mathrm{SL}_2(\mathbb{R})$ acts on it by (4.1). Strictly speaking this is not the group of linear fractional transformations, which are mappings, because two matrices give the same l.f.t. if they are related by multiplication by $-I$. Automorphic forms and modular forms are really the study of functions that transform in a certain way under the action (4.1). We would like to study the geometry of this action. Before we do so, we can define an automorphic form of weight k for $\mathrm{SL}_2(\mathbb{Z})$ to be a function on lattices, which is homogeneous of degree $-k$. This means that

$$F(\lambda L) = \lambda^{-k} F(L), \quad \forall L.$$

We set $f : \mathbb{H} \to \mathbb{C}$, $f(z) = F(\langle 1, z\rangle)$, where $\langle , \rangle$ means the lattice generated by the vectors. We determine the behavior of f under the action of $\mathrm{SL}_2(\mathbb{Z})$. We have

$$\begin{aligned} f(\gamma z) = F(\langle 1, \gamma z\rangle) = F((cz+d)^{-1}\langle cz+d, az+b\rangle) = (cz+d)^k F(\langle cz+d, az+b\rangle) \\ = (cz+d)^k F(\langle 1, z\rangle) = (cz+d)^k f(z), \end{aligned}$$

since the lattice is determined by the two bases $1, z$ and $cz + d, az + b$.

If we ask that f is holomorphic in $\mathbb{H}$, we are lead to the theory of classical modular forms. For the spectral theory of automorphic forms, we ask instead that f satisfies appropriate partial differential equations. This is not simply an effort to generalize. There is a theory of invariant differential operators out of the Lie algebra of $\mathrm{SL}_2(\mathbb{R})$ that justifies this. This will not be explained further in the notes.

5. Hyperbolic Geometry

Since for $\gamma \in \mathrm{SL}_2(\mathbb{R})$

$$\gamma'(z) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{1}{(cz+d)^2} \tag{5.1}$$

(a simple calculation, using that $\det(\gamma) = 1$), and the formula for computing the length of a curve $s(t)$, $t \in [a, b]$ is

$$l = \int_a^b |s'(t)|\, dt$$

we get by the formula for change of variables that the length of $\gamma(s(t))$ is

$$\tilde{l} = \int_a^b \frac{1}{|cs(t)+d|^2} |s'(t)|\, dt.$$

This shows that inside the circle $|cz + d| = 1$, given by $|cz + d| < 1$, the length is increased, while outside it is decreased. However, this does not remain true if we change the way we measure lengths. We can turn γ to be an isometry of $\mathbb{H}$, if we adjust the metric. The clue is in Eq. (4.2). We have (using the complex derivative in the form $df/dz = f'(z)$)

$$\frac{|d(\gamma z)|}{\Im(\gamma z)} = \frac{|\gamma'(z)|\,|dz|}{\Im(z)/|cz+d|^2} = \frac{|dz|}{\Im(z)}.$$

This means that if we set

$$ds^2 = \frac{|dz|^2}{(\Im z)^2} = \frac{dx^2 + dy^2}{y^2}$$

to be the metric, then $\gamma \in \mathrm{SL}_2(\mathbb{R})$ is an isometry, since it preserves the lengths (at least locally). This is the hyperbolic metric on $\mathbb{H}$, and $\mathbb{H}$ is called the hyperbolic plane. Notice that angle measurements between (tangent) vectors are the same as in the Euclidean case. This is seen by the absence of the $dx\,dy$ term. We can know find the hyperbolic length of the curve $s(t) = x(t) + iy(t)$, $t \in [a, b]$ by the formula

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

Once we know how to compute lengths of curves, we define the distance $d(p, q)$ between two points as the infimum of the lengths of the smooth curves connecting p and q. Curves which locally minimize the distance are called geodesics. We compute the geodesics in hyperbolic space:

Theorem 5.1. *The geodesics in $\mathbb{H}$ are the half circles with center on the x-axis and the half-lines parallel to the imaginary axis.*

Proof. We begin by considering $z = ia$, $w = ib$ with $b > a > 0$. Let $s : [0, 1] \to \mathbb{H}$ be any curve with $s(0) = z$, $s(1) = w$. Then for its length we have

$$L = \int_0^1 \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} dt \geq \int_0^1 \frac{y'(t)}{y(t)} dt = \ln\left(\frac{y(1)}{y(0)}\right) = \ln(b/a).$$

On the other hand the most obvious choice of curve between these points is the segment on the imaginary axis $z(t) = i((1-t)a + tb)$, $t \in [0, 1]$ with length

$$\int_0^1 \frac{b-a}{t(b-a)+a} dt = \int_a^b \frac{du}{u} = \ln(b/a).$$

This means that the imaginary axis minimizes the distance between any of its points and it, therefore, a geodesic.

In general z and w are arbitrary. We can find a l.f.t. g that maps them on the imaginary axis. The geodesic through $g \cdot z$ and $g \cdot w$ is the imaginary axis. Since l.f.t's are isometries, the image of the imaginary axis by g^{-1} is a geodesic. We claim that it is one of the two kinds described in the theorem.

Case 1. $\Re(z) = \Re(w)$, so that $g(u) = u - \Re(z)$. Then the geodesic is the vertical ray through z and w and is the image under g^{-1} of the imaginary axis. Case 2. $\Re(z) \neq \Re(w)$. By drawing the perpendicular bisector of the segment through z and w, which is not parallel to the real axis, we find a point of its intersection with the real axis. Make this a center for a circle passing through z and w. We claim this is the image of the imaginary axis under g^{-1}, or, equivalently, that g maps this circle to the imaginary axis. We can specifically write down g as follows: Let α and β be the points of intersection of the circle with the real axis. Then

$$g(u) = \frac{u - \beta}{u - \alpha}.$$

This is seen as follows: g maps α to ∞ and β to 0. Since it preserves angles and the circle is perpendicular to the real axis, its image will be perpendicular to the image of the real axis, which is the real axis again (g has real entries). The only line-circle through ∞ and 0 perpendicular to the real axis is the imaginary axis. $\square$

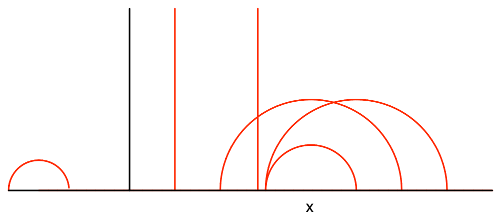


FIGURE 3. Various geodesics for $\mathbb{H}$

In fact, it is not too difficult to compute the distance between the points z and w:

$$d(z,w) = \ln\left(\frac{|z-\bar{w}| + |z-w|}{|z-\bar{w}| - |z-w|}\right). \tag{5.2}$$

See the exercises below.

In practice another form of the distance formula is more useful. Set

$$u(z,w) = \frac{|z-w|^2}{4\Im(z)\Im(w)} \tag{5.3}$$

Then

$$\cosh d(z,w) = 1 + 2u(z,w).$$

The function u is called the standard point-pair invariant.

It is interesting to notice that hyperbolic circles are Euclidean circles at the same time. Set the center to be w and the radius $r = d(z,w)$. The formula (5.2) shows that z satisfies the equation

$$\left|\frac{z-w}{z-\bar{w}}\right| = k, \quad e^r = \frac{1+k}{1-k}.$$

With $T(z) = (z-w)/(z-\bar{w})$, $\zeta = T(z)$, we have $|\zeta| = k$ and this is a circle. Then the locus of z is T^{-1} of this circle. Since T^{-1} is a l.f.t., this is also a circle.

Remark. Since l.t.f. maps circles to circles (or lines) but do not preserve the centers, a hyperbolic circle is a Euclidean circle with different center.

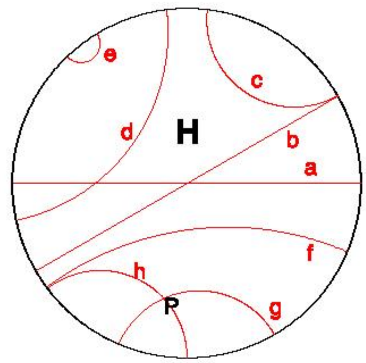


Figure 4. The hyperbolic disc and its geodesics

Remark. There is another model of the hyperbolic space, which is the Poincaré disc $\mathbb{D}$. This is the unit disc with the metric

$$ds^2 = 4\frac{dx^2 + dy^2}{(1 - (x^2 + y^2))^2} = 4\frac{|dz|^2}{(1 - |z|^2)^2}. \tag{5.4}$$

As a domain in the complex plane it is conformal with the upper-half space $\mathbb{H}$ using

$$f : \mathbb{D} \to \mathbb{H}, \quad z \to -i\frac{z+1}{z-1}.$$

The normalization guarantees that the origin of $\mathbb{D}$ corresponds to the point $i \in \mathbb{H}$. Here the geodesics are the images of the ones in $\mathbb{H}$ under f, which is a l.f.t. Consequently, they are also arcs of circles, or line segments. Since the geodesics of $\mathbb{H}$ meet the real axis at right angles, conformality implies that the geodesics in $\mathbb{D}$ are perpendicular to the circle $|z| = 1$. So they are circular arcs with this property and diameters of the circle. We see that Euclid's fifth postulate is not satisfied, for instance, in the figure there are many geodesics from P parallel to (not intersecting) b, e.g. h and g.

Let us show how formula (5.4) is obtained. Setting $w = f(z)$ and $w = x + iy$, we have

$$\frac{dw}{dz} = -\frac{2}{(z-i)^2}, \quad w = \frac{-i|z|^2 + 2\Re(z) + i}{|z-i|^2}$$

which implies that

$$\frac{|dw|^2}{\Im(w)^2} = \frac{4|dz|^2 / |z-i|^4}{(1-|z|^2)^2 / |z-i|^4} = \frac{4|dz|^2}{(1-|z|^2)^2}.$$

Eq. (5.4) is useful when using polar coordinates.

How does one measure area in the hyperbolic plane? The formula for the metric tells us that in the x direction we measure lengths infinitesimally as dx/y and in the y direction as dy/y. The area should be the product of the lengths so we arrive at the hyperbolic measure

$$d\mu(z) = \frac{dx\,dy}{y^2}. \tag{5.5}$$

We are interested in calculating areas of hyperbolic circles and triangles (and more generally polygons).

Theorem 5.2 (Gauss defect or Gauss-Bonnet formula)**.** *The area of a hyperbolic triangle with vertices in* $\mathbb{H} \cup \{\infty\}$ *is*

$$\pi - (\alpha + \beta + \gamma),$$

where α, β, γ *are the interior angles at the vertices. If a vertex is at* $\mathbb{R} \cup \{\infty\}$ *(we say that the vertex is at the boundary of* $\mathbb{H}$*), then its interior angle is* 0*.*

Proof. Let us assume first that there is a vertex at the boundary of $\mathbb{H}$. By using $1/(z-a)$ we can further assume that it is at infinity. Then two of the sides are segments parallel to the imaginary axis. The third side is an arc. By using a translation we can assume the arc is given by $Re^{i\theta}$, $\theta \in [a, b]$. Then the area is

$$A = \int_{R\cos a}^{R\cos b} \int_{\sqrt{R^2-x^2}}^{\infty} \frac{dx\,dy}{y^2} = \int_{R\cos a}^{R\cos b} \frac{1}{\sqrt{R^2-x^2}}\,dx = [\arcsin\theta]_{\cos a}^{\cos b} = (\pi/2 - a)$$
$$-(\pi/2 - b) = (b - a) = \beta + \pi - \alpha,$$

since on the interior angles, say β matches with b and the other α is supplementary to a. This is $\pi - (\alpha + \beta + 0)$ and the formula is correct.

If no vertex is on the boundary of $\mathbb{H}$, we use a l.f.t. to move one of the sides to become parallel with the imaginary axis. Call α, β, γ the interior angles. If the side is AB is vertical, consider the two triangles ACD and BCD, where D is at infinity. Using the result above

$$\mathrm{Area}(ABC) = \mathrm{Area}(ACD) - \mathrm{Area}(BCD) = \pi - (\alpha + \widehat{ACD}) + \pi - (\widehat{DBC} + \widehat{DCB})$$
$$= \pi - (\alpha + \beta + \gamma)$$

since $\widehat{ACD} = \gamma + \widehat{BCD}$ and $\widehat{DBC} = \pi - \beta$. This proves the result. $\square$

Remark. In particular it follows that the sum of the interior angles is strictly less than π, unlike Euclidean geometry.

Remark. There are many reasons why we care about hyperbolic triangles. One is that fundamental domains for many arithmetically defined discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$ are unions of simple hyperbolic triangles.

We compute the area of a hyperbolic ball of radius r. By using the model of hyperbolic space in the unit disc we can assume that the center is at $(0,0)$ and the Euclidean radius is R (a hyperbolic circle is also a Euclidean circle). Putting the center at $(0,0)$ makes both centers agree (with (ρ,θ) polar coordinates in the Euclidean plane)

$$d(0,R)=\int_0^R \frac{2}{1-\rho^2}d\rho=\ln\frac{1+R}{1-R}=r,$$

which shows how R and r are related:

$$e^r=\frac{1+R}{1-R}\Longrightarrow R=\frac{e^r-1}{e^r+1}=\tanh(r/2).$$

Now the area of the hyperbolic disc is

$$A(r)=\int_0^{2\pi}\int_0^R \frac{4}{(1-\rho^2)^2}\rho\, d\rho d\theta=4\pi\left[\frac{1}{1-\rho^2}\right]_0^R=4\pi\left(\frac{1}{1-R^2}-1\right)$$

$$=4\pi\left(\cosh^2(r/2)-1\right)=4\pi\sinh^2(r/2).$$

It is extremely instructive to compute the length of the hyperbolic circle of radius r. Using the standard formula from calculus $dx^2+dy^2=d\rho^2+\rho^2d\theta^2$ we get on a circle of radius R

$$L(r)=\int_0^{2\pi}\frac{2R}{1-R^2}d\theta=4\pi\frac{R}{1-R^2}=4\pi\tanh(r/2)\cosh^2(r/2)$$

$$=4\pi\sinh(r/2)\cosh(r/2)=2\pi\sinh r.$$

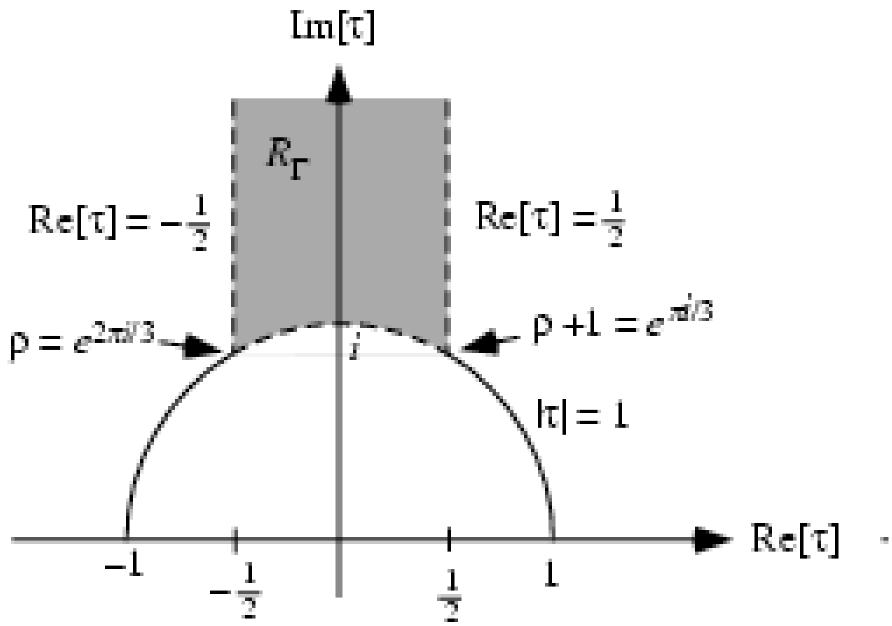
Notice that $L(r)\sim\pi e^r$ and $A(r)\sim\pi e^r$, as $r\to\infty$, so the area and the length grow at the same rate!

Remark. We also notice that although in Euclidean geometry the circle $|z|=1$ is at finite distance from the origin, it is at infinite distance in hyperbolic geometry: $r=\ln(1+R)/(1-R)\to\infty$ as $R\to 1$. This way the hyperbolic disc $\mathbb{D}$ (and $\mathbb{H}$) becomes a complete Riemannian manifold. It can be proved that its curvature is constant $=-1$. The Gauss defect formula can alternatively be proved from the value of the curvature and the general Gauss-Bonnet theorem.

6. Holomorphic modular forms for $\mathrm{SL}_2(\mathbb{Z})$

In this section we only consider $\Gamma=\mathrm{SL}_2(\mathbb{Z})$. The orbit of a point $z\in\mathbb{H}$ is defined to be $\mathrm{orb}(z)=\{\gamma z,\gamma\in\Gamma\}$.

Definition. A fundamental domain D of Γ is a subset of $\mathbb{H}$ such that every orbit of Γ in $\mathbb{H}$ has one element in D and two points in D are in the same orbit iff they are on ∂D.



Theorem 6.1. *(a) The standard fundamental domain for $SL_2(\mathbb{Z})$: Let*

$$D = \{z \in \mathbb{H}, -\frac{1}{2} \leq \Re(z) \leq \frac{1}{2}, |z| \geq 1\}.$$

Then D is a fundamental domain of Γ.

(b) Let $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ be the matrices acting as translation $T(z) = z + 1$ and inversion $S(z) = -1/z$. Then Γ is generated by them, $\Gamma = \langle T, S\rangle$. Moreover, $S^2 = -I$ and $(ST)^3 = -I$. The stabiliser of i is $\langle S\rangle$ and the stabiliser of $\rho = e^{2\pi i/3}$ is $\langle ST\rangle$.

Hint of proof: Given a point $z \in \mathbb{H}$, we can apply enough powers of T to move it within the strip $\{w, -\frac{1}{2} \leq \Re(w) \leq \frac{1}{2}\}$. If $m = \operatorname{dist}(\Re(z), \mathbb{Z})$, then $\Re(T^{-m}z) \in [-1/2, 1/2]$, since every number is within 1/2 from an integer. If $z_2 = T^{-m}z \in D$, we are done. Otherwise, $|z_2| < 1$. Then this implies that $|S(z_2)| = |-1/z_2| > 1$. We repeat the process to get z_3 with $\Re(z_3) \in [-1/2, 1/2]$. The process cannot be continued indefinitely, as $\mathrm{SL}_2(\mathbb{Z})$ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$.

We consider the set $C_B = \{z \in \mathbb{H}, |\Re(z)| \leq 1/2, \Im(z) > B\}$. This is called a cuspidal sector. The map $q = e^{2\pi iz}$ maps it to the punctured disc of radius $e^{-2\pi B}$, called D_B^*, in a biholomoprhic way, up to the identification of $\Re(z) = -1/2$ with $\Re(z) = 1/2$. A function which is periodic with period 1 on C_B induced a function $\tilde{f}$ on D_B^*. If $\tilde{f}$ is meromorphic at 0, i.e. has at most a pole and not an essential singularity, then we say that f is meromorphic at ∞. This is equivalent to $\exists N \in$

$\mathbb{N}, \exists K$, $|\tilde{f}(q)q^N| \leq K$. Then $\tilde{f}(q) = \sum_{n=-N}^{\infty} c_n q^n$ is the Laurent expansion of $\tilde{f}$ and

$$f(z) = \sum_{n=-N}^{\infty} c_n e^{2\pi i n z}$$

is the Fourier expansion of f at the cusp $i\infty$. The integer $-N$ is the order of f at ∞. If $N = 0$, we say that f is holomorphic at ∞ and if $N \geq 1$ that f is cuspidal at ∞.

Naturally in our mind we have automorphic functions, i.e. functions $f : \mathbb{H} \to \mathbb{C}$ of weight k for Γ, i.e.

$$f(\gamma z) = (cz + d)^k f(z). \tag{6.1}$$

These are periodic, since $f(z + 1) = f(z)$, i.e. the lower row of T is $(0, 1$. So the above commends on q expansion at ∞ apply.

Definition. A modular form for Γ is a function $f : \mathbb{H} \to \mathbb{C}$ that is holomorphic in $\mathbb{H}$ and at ∞ and transforms according to (6.1). If f is cuspidal, then we call it holomorphic cusp form.

We easily see that for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, k has to be even. This follows from the remark that $-I \in \mathrm{SL}_2(\mathbb{Z})$, but introduces the same Möbius transformation as I, so that $f(-Iz) = f(z) = (-1)^k f(z)$.

Remark. The cuspidality condition may see not well motivated. The following is an important point of view in the theory of elliptic curves: If $f(z)$ is a cusp form of weight 2, then $f(z)dz$ is a holomorphic differential on $\mathbb{H}$ and at ∞ and invariant under Γ. First we check that $f(z)dz$ is invariant:

$$f(\gamma z)d(\gamma z) = (cz + d)^2 f(z)\gamma'(z)dz = (cz + d)^2 f(z)\frac{1}{cz + d)^2}dz = f(z)dz,$$

using (5.1). The holomorphicity in $\mathbb{H}$ is obvious, it is only ∞ that can be an issue. We have $dq = 2\pi i e^{2\pi i z}dz = 2\pi i q dz$, so that

$$f(z)dz = \tilde{f}(q)\frac{dq}{2\pi i q} = \frac{1}{2\pi i}\sum_{n\geq 1} c_n q^{n-1}.$$

Lemma 6.1. *If f is a cusp form, then $f(z) = O(e^{-2\pi y})$, as $y \to \infty$.*

Proof. Having a convergent Taylor series at $q = 0$ with $c_0 = 0$ means $\tilde{f}(q) = O(q)$, which translates to the result, as $|q| = e^{-2\pi y}$.

□

Lemma 6.2. *(Hecke bound on the Fourier coefficients of cusp forms) If f is a cusp form of weight k with Fourier expansion*

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

then

$$|a_n| = O(n^{k/2}).$$

Proof. We look at $g(z) = y^{k/2}|f(z)|$. Using (4.2) we get

$$\Im(\gamma z)^{k/2}|f(\gamma z)| = \left(\frac{y}{|cz+d|^2}\right)^{k/2} |(cz+d)^k f(z)| = y^{k/2}|f(z)|.$$

As $f(z)$ is decaying exponentially in y, as $y \to \infty$, this decay overpowers the polynomial increase of $y^{k/2}$ and $g(z)$ decays at ∞. In particular it is bounded on the fundamental domain D. Since it is automorphic of weight 0, it is bounded on all of $\mathbb{H}$. Hecke's bound will use $y \to 0$ in the proof. We recover the Fourier coefficients of f by Fourier analysis on $[0, 1]$:

$$e^{-2\pi ny} a_n = \int_0^1 f(x+iy)e^{-2\pi inx}\, dx = O(y^{-k/2}).$$

(One needs to know that the Fourier expansion is valid on all of $\mathbb{H}$). Plugging $y = 1/n$ produces the result, as it makes $e^{-2\pi ny} = e^{-2\pi}$ fixed.

□

7. Epstein zeta function and Eisenstein series

The natural generating function for counting the lattice points for a lattice L is the function

$$D(z, s) = \sum_{(m,n)\neq(0,0)} \frac{1}{|m+nz|^s}.$$

However, this converges for $\Re(s) > 2$, and analytic number theorists prefer to work with Dirichlet series converging for $\Re(s) > 1$, in accord with the Riemann zeta function

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

So we modify it first to

$$\sum_{(m,n)\neq(0,0)} \frac{1}{|m+nz|^{2s}}$$

and then to the Epstein zeta function

$$B(z, s) = \sum_{(m,n)\neq(0,0)} \frac{\Im(z)^s}{|m+nz|^{2s}}.$$

The introduction of $\Im(z)^s$ allows to write the term as $\Im(\gamma z)^s$, where the second row of γ is (n, m) (at least this is possible when $(n, m) = 1$).

For the Gauss circle problem we need to plug in $z = i$. Another function that plays a role here is the Dirichlet L-series associated with the (nontrivial) character (mod 4), $\chi(\cdot)$, defined as

$$\chi(n) = \begin{cases} 1, & n \equiv 1 \pmod 4, \\ -1, & n \equiv 3 \pmod 4, \\ 0, & n \equiv 0 \pmod 2. \end{cases}$$

Then

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

Unlike the Riemann zeta function this converges for $\Re(s) > 0$, although not absolutely in the strip $0 < \Re(s) \leq 1$. And it can be evaluated at 1:

$$L(1,\chi) = 1 - \frac{1}{3} + \frac{1}{5} - \cdots = \arctan(1) = \frac{\pi}{4}.$$

Since $B(i,s) = 4\zeta(s)L(s,\chi)$ (see exercises below), this is the generating series for the Gauss circle problem introduced above. Since $\zeta(s)$ has an analytic continuation to $\Re(s) > 0$ with single pole at $s = 1$ with residue 1, then $B(i,s)$ satisfies the conditions of the theorem with $a = \pi$. The result is the main term in the Gauss circle problem.

Here is a slightly more advanced point of view: Let $\mathbb{R}^2$ act by translation on itself: $(x,y)\cdot(z,w) = (x+z, y+w)$, where $\cdot$ denotes the action. This is clearly a group action, since $(0,0)$ does not move the point and $g_1 \cdot (g_2 \cdot p) = (g_1 g_2) \cdot p$ (associativity of addition!). If we restrict the action to the subgroup $\mathbb{Z}^2$, then the integer lattice is the orbit of the origin $(0,0)$. So the Gauss circle problem asks to count the number of points in this orbit at distance $\leq \sqrt{x}$ from a fixed point, here the origin.

8. From counting lattice points to the hyperbolic lattice point problem

In number theory one also imposes extra conditions at our counting problems. For instance, we can ask for the number of lattice points with relative prime coordinates (these points are visible from the origin, i.e. the segment from $(0,0)$ to (m,n) does not contain other lattice points.) These lead to consider

$$E(z,s) = \sum_{(c,d)=1} \frac{\Im(z)^s}{|cz+d|^{2s}}$$

for $z = i$, which gives

$$E(i,s) = \sum_{(c,d)=1} \frac{1}{(c^2+d^2)^s}.$$

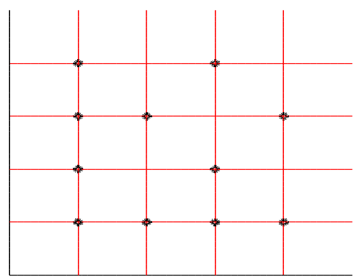


FIGURE 5. Visible points from $(0,0)$ in the first quadrant

The relation between $B(z,s)$ and $E(z,s)$ is very simple. Setting $d = (m,n)$, $m = dm'$, $n = dn'$, $(m',n') = 1$, we have

$$B(z,s) = \sum_{d=1}^{\infty} \frac{1}{d^{2s}} \sum_{(m',n')=1} \frac{\Im(z)^s}{|m'z+n'|^{2s}} = \zeta(2s)E(z,s).$$

So the residue of $E(i,s)$ at $s=1$ is $\pi/\zeta(2) = 6/\pi$. This gives

$$\#\{(m,n),(m,n)=1, |(m,n)| \le \sqrt{x}\} \sim \frac{6}{\pi}x, \quad x \to \infty.$$

One can consider lattice-counting problems in higher-dimensional Euclidean space. As far as the main term of the counting function is concerned, Gauss' argument works the same way, e.g.

$$\#\left\{(a,b,c,d) \in \mathbb{Z}^4, a^2+b^2+c^2+d^2 \le x\right\} \sim c_4 x^2$$

where c_4 is the volume of the unit ball in $\mathbb{R}^4$. What if we impose the restriction that $ad - bc = 1$? i.e. we consider the 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

from $\mathrm{SL}_2(\mathbb{R})$. How many such matrices with integer entries have norm $a^2+b^2+c^2+d^2 \le \sqrt{x}$? This is really a question about the growth of the group $\mathrm{SL}_2(\mathbb{Z})$. First of all we realize that actually $a^2+b^2+c^2+d^2=4u(\gamma i,i)+2$. This is because

$$u(\gamma i,i)=\frac{|\gamma i-i|^2}{4\Im(\gamma i)}=\frac{|ai+b+c-di|^2}{4}=\frac{a^2+b^2+c^2+d^2-2(ad-bc)}{4}.$$

We also know that $\cosh d(\gamma i,i)=2u(\gamma i,i)+1$. So the condition $a^2+b^2+c^2+d^2\leq X$ can be understood as $d(\gamma i,i)\leq\cosh^{-1}(X/2)$. So we are asking to count the number of points in the orbit of i that are within distance $\cosh^{-1}(X/2)$ from the point i. More generally: We fix two points z and w in $\mathbb{H}$ and consider the orbit Γz of z. Here Γ can be $\mathrm{SL}_2(\mathbb{Z})$, or other similar group. We are interested to count the points in this orbit with a certain distance from w. Set

$$P(X)=\#\left\{\gamma\in\Gamma, 4u(\gamma z,w)+2\leq X\right\}.$$

We would like to estimate $P(X)$ as $X\to\infty$. This is the hyperbolic lattice counting problem. For $\mathrm{SL}_2(\mathbb{Z})$ here is the result

$$P(X)=6X+O(X^{2/3}).$$

In view of the fact that the length of the hyperbolic circle and the area of the hyperbolic disc it encloses are comparable for large r, the argument of Gauss for the standard lattice-counting problem cannot possible work. The spectral method does (as well as methods from dynamical systems, introduced by Margulis et al.) The spectral method does not only provide the main term in the asymptotics but gives information on the error term. Here is the general result:

$$P(X)=\sum_{s_j\in(1/2,1]}2\pi^{1/2}\frac{\Gamma(s_j-1/2)}{\Gamma(s_j+1)}u_j(z)u_j(w)X^{s_j}+O(X^{2/3}).$$

We need to introduce (and study) the quantities of the right-hand side. The Gamma function is $\Gamma(s)=\int_0^\infty e^{-t}t^s dt/t$ for $\Re(s)>0$ and generalizes the factorial as $\Gamma(n)=(n-1)!$, $n\in\mathbb{N}$. More, importantly, s_j are 'spectral parameters': the numbers $\lambda_j=s_j(1-s_j)$ are eigenvalues of the Laplace operator on $L^2(\Gamma\backslash\mathbb{H})$ and $u_j(z)$ are the corresponding eigenfunctions. As a prelude of things to come, we mention that the Laplace operator can have infinitely many L^2 eigenvalues but only those $<1/4$ contribute to the above sum. It could be that actually the error term is larger than some terms of the sum: If $s_j<2/3$ the corresponding term is smaller than the error term. This corresponds to eigenvalues $s_j(1-s_j)>(2/3)(1/3)=2/9$. It can be proved that $\mathrm{SL}_2(\mathbb{Z})$ has no eigenvalues in the interval $(0,1/4]$ corresponding to $s_j\in[1/2,1)$. In general, eigenvalues $\lambda_j<1/4$ are called small eigenvalues. The Selberg eigenvalue conjecture is that

$$\lambda_1\geq 1/4$$

for groups of 'arithmetic nature', called congruence subgroups. For $\mathrm{SL}_2(\mathbb{Z})$ the contribution to the sum of the eigenvalue $\lambda_0=0$, i.e. $s_0=1$ is calculated as follows. We have $\Gamma(1/2)=\sqrt{\pi}$. The eigenfunction corresponding to λ_0 is the constant

$$u_0(z)=\frac{1}{\sqrt{\mathrm{Area}(\Gamma\backslash\mathbb{H})}}=\frac{1}{\sqrt{\pi/3}}.$$

The area of $\Gamma\backslash\mathbb{H}$ is in this case the area of a hyperbolic triangle with vertices at ∞, $e^{i\pi/3}$, $e^{2\pi i/3}$ and interior angles 0, $\pi/3$, $\pi/3$.

9. Properties of the Eisenstein series and the Laplace operator

What about computing the number of lattice points inside a disc of radius R but for the skewed lattice generated by 1 and z? The technique with the factorization of the Epstein zeta function $B(i,s)$ is not available. According to Th. 1.3 it is expedient to know the analytic/meromorphic continuation of $B(z,s)$ for values of s with $\Re(s)\le 1$. Here are two important points:

- $B(z,s)$ is a function periodic in the $x=\Re(z)$ variable with period 1. This is obvious, since changing z to $z+1$, simply changes the basis of the lattice (recall 1 is a basis element).
- $B(z,s)$ is certainly not holomorphic, or harmonic is z. Let us apply $\Delta=\partial_x^2+\partial_y^2$ to $y^s=\Im(z)^s$, which is the function we are somehow automorphizing. We get
$$\Delta y^s=s(s-1)y^{s-2},$$
and this is certainly not zero. However, is we multiply it with y^2 we get almost what we started with, up to the factor $s(s-1)$. In fact, this means that y^s satisfies the eigenvalue equation
$$y^2\left(\partial_x^2+\partial_y^2\right)f(z)=-s(1-s)f(z).$$

The question is what kind of operator is the one appearing on the left. It ends up that this is the hyperbolic Laplacian

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

The spectral theory of automorphic forms is essentially the study of the hyperbolic laplacian and its eigenvalues/eigenfunction, not on the whole hyperbolic plane (although this is a first step) but on domains in it representing the action of discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$ (fundamental domains) with appropriate boundary conditions. We start with explaining the factor y^2 in front of the Euclidean Laplacian $\Delta_{eucl}=\partial_x^2+\partial_y^2$. The general form for the Laplacian in the metric g_{ij} is

$$\Delta=\frac{1}{\sqrt{g}}\sum_{ij}\frac{\partial}{\partial x_i}\sqrt{g}g^{ij}\frac{\partial}{\partial x_j}, \tag{9.1}$$

where $g=\sqrt{\det(g_{ij})}$ and g^{ij} is the inverse matrix to g_{ij}. This definition for the general Laplace operaror agrees with the alternative $\Delta f=\mathrm{div}(\mathrm{grad}(f))$, where $\mathrm{grad}(f)$ is the gradient vector field and div represents its divergence. Rather than introduce these notions for general Riemannian manifolds, we stick to the definition (9.1). Here $g_{11} = g_{22} = y^{-2}$ and $g_{12} = g_{21} = 0$, $g^{11} = g^{22} = y^2$, $g^{12} = g^{21} = 0$, and $g = y^{-2}$. The following heuristic (or aposteriori argument) shows that this makes sense. Let f and g be two smooth functions, say, vanishing on the boundary of the region V. We would like the Laplace operator to be symmetric for such functions. This means

$$\int_V f(\Delta g) d\mu(z) = \int_V g(\Delta f) d\mu(z).$$

If we use $\Delta f = y^2 \Delta_{eucl}$ and (5.5) we get

$$\int_V f(\Delta g) d\mu(z) - \int_V g(\Delta f) d\mu(z) = \int_V f\Delta_{eucl} g - g\Delta_{eucl} f \, dx \, dy = \int_{\partial V} f\frac{\partial g}{\partial n} - g\frac{\partial f}{\partial n} dl$$

by the standard Green's formula in $\mathbb{R}^2$. Here $\partial/\partial n$ is the normal derivative. With the assumption of vanishing on ∂V, the right-hand side is 0 and the hyperbolic Laplacian is symmetric.

A standard technique to understand partial differential operators is to separate variables and also to consider solutions that have a special symmetry. For instance, we already noticed that y^s satisfies the eigenvalue equation

$$\Delta f(z) + s(1-s)f(z) = 0. \tag{9.2}$$

There should be a second linearly independent solution, and this is y^{1-s} as easily seen. Clearly they are linearly independent, unless $s = 1 - s \Longrightarrow s = 1/2$. In such a case we verify that $y^{1/2} \ln y$ is the second solution. These solutions are independent of the x variable. They show in the zeroth Fourier coefficient of the Eisenstein series.

On the other hand it is interesting (in fact necessary) to know the eigenfunctions of the Laplace operaror that depend only on the hyperbolic distance and not on the polar angle. For this we write the Laplace operator as

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{\cosh r}{\sinh r}\frac{\partial}{\partial r} + \frac{1}{\sinh^2 r}\frac{\partial^2}{\partial \theta^2}$$

and using u as

$$\Delta = u(u+1)\frac{\partial^2}{\partial u^2} + (2u+1)\frac{\partial}{\partial u} + \frac{1}{4u(u+1)}\frac{\partial^2}{\partial \theta^2}.$$

For the proofs look at the exercises. We assume that $f(z) = f(r) = F(u) = F_s(u)$ and it satisfies (9.2). Then

$$u(u+1)F''(u) + (2u+1)F'(u) + s(1-s)F(u) = 0$$

A solution of this is the hypergeometric function

$$F_s(u) = F(s, 1-s, 1, -u).$$

Students are usually not familiar with the hypergeometric functions. Unfortunately special functions, like hypergeometric functions, Legendre functions, and Bessel functions show up regularly in the spectral theory of automorphic forms (depending on what expansion one works with) and are quite intimidating at the beginning. Here is a quick note on hypergeometric functions. The Gauss hypergeometric function $F(a,b,c,z)$ is defined as

$$F(a,b,c,z)=\sum_{n=0}^{\infty}\frac{a(a+1)\cdots(a+n-1)b(b+1)\cdots(b+n-1)}{n!c(c+1)\cdots(c+n-1)}z^n,$$

where $|z|<1$, $c\in\mathbb{C}\setminus\{0,-1,-2,\ldots\}$. We need to know the differential equation satisfied by it and it is

$$z(1-z)F''(z)-((a+b+1)z-c)F'(z)-abF(z)=0$$

and the integral representation

$$F(a,b,c,z)=\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt,$$

for $\Re(c)>\Re(b)>0$. We see by substituting $z=-u$ that $F(s,1-s,1,-u)$ satisfies the eigenvalue equation in u and is independent of θ. Its value at $u=0$ is 1. There should be a second linear independent solution. This is the Green's function for functions invariant under rotations. It is defined as

$$G_s(u)=\frac{1}{4\pi}\int_0^1 (t(1-t))^{s-1}(t+u)^{-s}\,dt.$$

We need to show that Δ in the u coordinates has eigenfunction

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

with eigenvalue $s(1-s)$. If we prove that

$$(\Delta+s(1-s))(t(1-t))^{s-1}(t+u)^{-s}=s\frac{d}{dt}\{(t(1-t))^{s}(t+u)^{-s-1}$$

then differentiation under the integral sign will give 0 as

$$\int_0^1 \frac{d}{dt}\{(t(1-t))^{s}(t+u)^{-s-1}\}\,dt=(t(1-t))^{s}(t+u)^{-s}|_0^1=0$$

as t^s vanishes at 0 and $(1-t)^s$ vanishes at 1 for $\Re(s)>0$. The calculation is not inspiring, using (10.2)

$$\begin{aligned}(\Delta+s(1-s))(t+u)^{-s} \;&=\; u(u+1)s(s+1)(t+u)^{-s-2}+(2u+1)(-s)(t+u)^{-s-1}\\ &\quad +s(1-s)(t+u)^{-s}.\end{aligned}$$

Afterwards we multiply with $(t(1-t))^s$ which is independent of u, while

$$s\frac{d}{dt}\{(t(1-t))^{s}(t+u)^{-s-1}\}=s^2(t^{s-1}(1-t)^s-t^s(1-t)^{s-1}(t+u)^{-s-1}+s(-s-1)t^s(1-t)^s(t+u)^{-s-2}$$

The factor $t^{s-1}(1-t)^{s-1}(t+u)^{-s-2}$ appears throughout so we need to prove the simpler looking:

$$u(u+1)s(s+1)+(2u+1)(-s)(t+u)+s(1-s)(t+u)^2 = s^2(1-2t)-s(s+1)t(1-t).$$

Both sides are quadratic polynomials in u, so we compare the coefficients of 1, u and u^2 on both sides:

$$u^2: \quad s(s+1)-2s+s(1-s)=0$$

from the left-hand side with no quadratic term on the right-hand side.

$$u: \quad s(s+1)-2st-s+s(1-s)2t = s^2(1-2t)$$

$$1: \quad -st+s(1-s)t^2 = ts(-1+t-ts)$$

from the left-hand side, while from the right-hand side we get

$$s^2(1-2t)t-s(s+1)t(1-t) = ts[s(1-2t)-(s+1)(1-t)] = ts(-2ts-1+st+t)$$

10. Exercises

(1) In this problem $\int_{c-i\infty}^{c+i\infty}$ denotes a contour integral along the vertical line $\Re(s) = c$ traversed upwards.

(a) Prove that, for $c>0$, $\displaystyle\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^s}{s}ds = \begin{cases} 1, & x>1, \\ 1/2, & x=1, \\ 0, & 0<x<1. \end{cases}$ (Perron formula)

Look at Figure 6 for the contour to consider.

For $x>1$, inside the contour there is a pole at $s=0$, which is simple:

$$\operatorname{Res}(f,0) = \lim_{s\to 0} s\frac{x^s}{s} = x^0 = 1.$$

We need to control the integral on γ_R. As above, on the left semicircle $s(t) = c+Re^{it}$, $\pi/2 \le t \le 3\pi/2$.

$$|x^s| = |e^{s\log x}| = e^{\log x(c+R\cos t)}$$

We also remark that the inequality $\sin y \ge 2y/\pi$ holds for $0 \le y \le \pi/2$. This follows from the concavity of $\sin y$ on $[0,\pi/2]$. The secant line $2y/\pi$ from $(0,0)$ to $(\pi/2,1)$ is below the graph. Now

$$\left|\int_{\gamma_R}\frac{x^s}{s}ds\right| = \left|\int_{\pi/2}^{3\pi/2}\frac{x^c x^{R\cos t}}{c+Re^{it}}iRe^{it}dt\right| = \left|\int_0^{\pi}\frac{x^c x^{-R\sin y}}{c+Re^{i(y+\pi/2)}}Re^{iy}dy\right| \le \int_0^{\pi}\frac{x^c x^{-R\sin y}}{R-c}Rdy$$

with the substitution $t = y+\pi/2$. The last integral can be split into two equal integrals over $[0,\pi/2]$ and $[\pi/2,\pi]$, since $\sin y$ takes the same values in both. We get

$$\left|\int_{\gamma_R}\frac{x^s}{s}ds\right| \le 2\int_0^{\pi/2}\frac{Rx^c x^{-R\sin y}}{R-c}dy \le 2\int_0^{\pi/2}\frac{Rx^c x^{-R2y/\pi}}{R-c}dy = \frac{Rx^c}{R-c}\left[\frac{x^{-2Ry/\pi}}{-R(\log x)2/\pi}\right]_0^{\pi/2}$$

$$= \frac{\pi x^c}{(R-c)(\log x)2}\left(-x^{-R}+1\right) \to 0, \quad R \to \infty,$$

as $x > 1$.

For $0 < x < 1$ the parametrization of the right semicircle is $s(t) = c + Re^{it}$, $-\pi/2 \le t \le \pi/2$. We substitute $t = y - \pi/2$:

$$\left|\int_{\gamma_R} \frac{x^s}{s} ds\right| = \left|\int_{-\pi/2}^{\pi/2} \frac{x^c x^{R\cos t}}{c+Re^{it}} iRe^{it} dt\right| \le \int_{-\pi/2}^{\pi/2} R\frac{x^c x^{R\cos t}}{R-c} dt = 2\int_0^{\pi/2} \frac{Rx^c x^{R\sin y}}{R-c} dy$$

Now $x < 1$, so $\log x < 0$ and

$$x^{R\sin y} = e^{\log x R \sin y} \le e^{\log x R 2y/\pi} = x^{2Ry/\pi}.$$

$$\left|\int_{\gamma_R} \frac{x^s}{s} ds\right| \le 2\int_0^{\pi/2} \frac{Rx^c x^{R2y/\pi}}{R-c} dy = \frac{Rx^c}{R-c}\left[\frac{x^{2Ry/\pi}}{R(\log x)2/\pi}\right]_0^{\pi/2}$$

$$= \frac{\pi x^c}{(R-c)(\log x)2}\left(x^R - 1\right) \to 0, \quad R \to \infty,$$

as $x < 1$.

For $x = 1$ we compute the integral directly:

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{1}{s} ds = \frac{1}{2\pi i}\int_{\infty}^{\infty} \frac{idt}{c+it} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{c}{c^2+t^2} - \frac{it}{c^2+t^2} dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{c}{c^2+t^2} dt$$

as the function $t/(c^2+t^2)$ is odd. This integral is elementary: substitute $t = cu$ to get

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{1}{s} ds = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{c \cdot cdt}{c^2+c^2u^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{du}{1+u^2} = \left[\frac{\arctan(u)}{2\pi}\right]_{-\infty}^{\infty} = \frac{1}{2\pi}\pi = \frac{1}{2}.$$

(b) Let the function $f(s)$ be defined by the absolutely convergent series

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}, \quad \Re(s) > a \ge 0.$$

Show that for $x \notin \mathbb{Z}$

$$\sum_{n \le x} a_n = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(s)\frac{x^s}{s} ds, \quad c > a.$$

Since $|n^s| = n^{\Re(s)}$ and the series converges absolutely for $\Re(s) > a$, we have that the series

$$\sum_{n=1}^{\infty} \frac{|a_n|}{n^{\Re(s)}} < \infty, \quad \Re(s) > a.$$

This implies that, if we fix $\Re(s) = c > a$, then the convergence of the series in the s variable is uniform (Weierstraß test). Moreover, on the vertical line $\Re(s) = c$ we have

$$\left|\frac{x^s}{s}\right| = \frac{x^c}{|s|},$$

which is bounded so we get uniform convergence of the series even when multiplied by x^s/s. This means we can interchange summation and integration to get

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(s)\frac{x^s}{s}ds = \sum_{n=1}^{\infty} a_n \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{(x/n)^s}{s}ds = \sum_{n=1}^{\infty} a_n \left\{ \begin{array}{ll} 0, & x < n, \\ 1, & x > n \end{array} \right\} = \sum_{n<x} a_n.$$

(2) Let a_n be multiplicative. Show that its Dirichlet series has a product expansion

$$D(s) = \prod_p \sum_{j=0}^{\infty} a_{p^j} p^{-js}.$$

This is a generalisation of the Euler product for the Riemann zeta function. We use the unique factorisation of the integers into primes and expand the infinite product on the right-hand side. We use the multiplicativity of a_n in the form $a_{p^m}a_{q^j} = a_{p^m q^j}$ and its generalisation to any finite number of factors.

(3) Show that

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

We have

$$\zeta(s-k)\zeta(s) = \sum_{l=1}^{\infty} \frac{1}{l^{s-k}} \sum_{m=1}^{\infty} \frac{1}{m^s} = \sum_{l,m} \frac{l^k}{(lm)^s} = \sum_{n=1}^{\infty} \sum_{lm=n} \frac{l^k}{n^s} = \sum_{n=1}^{\infty} \frac{\sum_{l|n} l^k}{n^s} = \sum_{n=1}^{\infty} \frac{\sigma_k(n)}{n^s}.$$

(4) Prove that $B(s)$ converges absolutely for $\Re(s) > 1$.

(5) Prove that $r(n) = 4\sum_{d|n}\chi(d)$, where χ is the quadratic character $\pmod 4$. Show that this implies that

$$B(i, s) = 4\zeta(s)L(s, \chi).$$

Let p and q be primes. It is well-known that a prime is a sum of two squares iff it is congruent to $1 \pmod 4$. Use the fact that $\mathbb{Z}[i]$ is a unique factorization domain with primes

- $1 + i$, $1 - i$,
- q, if $q \equiv 3 \pmod 4$,
- $a + ib$, $a - ib$ with $a^2 + b^2 = p$, $p \equiv 1 \mod 4$.

Proof: The units in $\mathbb{Z}[i]$ are $\pm i$, ± 1. When we write

$$n = A^2 + B^2 = (A + iB)(A - iB) = 2^{\alpha} \prod p^r \prod q^s$$

with $p \equiv 1 \pmod 4$ and $q \equiv 3 \mod 4$ we have

$$n = (1+i)^\alpha (1-i)^\alpha \prod (a+ib)^r (a-ib)^r \prod q^s.$$

This gives that

$$A + iB = i^t (1+i)^{a_1} (1-i)^{a_2} \prod (a+ib)^{r_1} (a-ib)^{r_2} \prod q^{s_1}$$

$$A - iB = i^{-t} (1-i)^{a_1} (1+i)^{a_2} \prod (a-ib)^{r_1} (a+ib)^{r_2} \prod q^{s_2}$$

so that $a_1 + a_2 = a$, $r_1 + r_2 = r$, $s_1 + s_2 = s$ and $s_1 = s_2$. So s has to be even and we must split the powers of q equally between $A + iB$ and $A - iB$. If s is odd for a prime $q \equiv \pmod 4$, then $r(n) = 0$. Otherwise, we will show that $r(n) = 4 \prod (r+1)$. We have $r+1$ choices for r_1, which, therefore, force the value of r_2. We have $\alpha + 1$ choices for a_1 and these fix a_2. We have 4 choices for t. However, since $(1-i)/(1+i) = -i$,

$$(1+i)^{a_1} (1-i)^{a_2} = (1+i)^{a_1+a_2} (-i)^{a_2} = (1+i)^a (-i)^{a_2}.$$

This shows that, varying a_1 produces only modifications in the exponent of the unit i. So the choices for a_1 are irrelevant. We are left with independent choices only for r_1 and t. These are $4 \prod (r+1)$.

Notice that, if we define $a(n) = \sum_{d|n} \chi(d)$, we need to show that $r(n) = 4a(n)$. We have that $a(n)$ is a multiplicative function, so that

$$a(2^\alpha) = 1, \quad a(p^r) = \sum_{i=0}^{r} \chi(p^i) = r + 1,$$

$$a(q^s) = \sum_{j=0}^{s} \chi(q^j) = 1 - 1 + \cdots \pm 1 = \frac{1 + (-1)^s}{2},$$

$$a(n) = \prod (r+1) \prod \frac{1 + (-1)^s}{2}.$$

This means that $a(n)$ is 0 if there is a factor q^s with s odd and $q \equiv 3 \pmod 4$. In such a case we also know that $r(n) = 0$ from the discussion above. Otherwise, the two formulas agree.

To show that

$$B(i, s) = \sum_{n=1}^{\infty} \frac{r(n)}{n^s} = 4\zeta(s) L(s, \chi)$$

we notice that the coefficient of n^{-s} from the right-hand side is

$$4 \sum_{d|n} 1 \cdot \chi(d).$$

(6) Show that $\zeta(2) = \pi^2/6$.

We use Parceval's identity on $\mathbb{R}/\mathbb{Z}$, said differently, for any periodic function $f : \mathbb{R} \to \mathbb{R}$ with period 1 and Fourier coefficients

$$\hat{f}(n) = \int_0^1 f(x) e^{-2\pi i n x}\, dx$$

we have

$$\sum_{n\in\mathbb{Z}} \left|\hat{f}(n)\right|^2 = \|f\|^2 .$$

We use $f(x) = x - [x]$, the fractional part of x. We have

$$\|f\|^2 = \int_0^1 x^2\, dx = \frac{1}{3}, \quad \hat{f}(0) = \int_0^1 x\, dx = \frac{1}{2}.$$

On the other hand, for $n \neq 0$ we use integration by parts to get

$$\hat{f}(n) = \int x e^{-2\pi i n x}\, dx = \left[\frac{x e^{-2\pi i n x}}{-2\pi i n}\right]_0^1 + \int_0^1 \frac{e^{-2\pi i n x}}{2\pi i n}\, dx = -\frac{1}{2\pi i n}.$$

These give

$$\frac{1}{4} + \sum_{n\neq 0} \frac{1}{4\pi^2 n^2} = \frac{1}{3} \Longrightarrow 2\sum_{n=1}^{\infty} \frac{1}{4\pi^2 n^2} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \Longrightarrow \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

(7) Use the summation by parts formula in the form

$$\sum_{n\leq x} a_n f(n) = A(x) f(x) - \int_1^x A(u) f'(u)\, du,$$

where

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

and $f(x)$ is a continuously differentiable function on $[1, x]$ to show that $\zeta(s)$, which initially converges for $\Re(s) > 1$ has a meromorphic continuation to the region $\Re(s) > 0$ with single simple pole at $s = 1$ with residue 1.

Proof: Let $a_n = 1$, so that $A(x) = [x]$, and $f(u) = u^{-s}$, so that $f'(u) = -su^{-s-1}$. Then we get

$$\sum_{n\leq x} \frac{1}{n^s} = [x]x^{-s} - \int_1^x \frac{[u](-s)}{u^{s+1}}\, du = [x]x^{-s} + s\int_1^x \frac{u - \{u\}}{u^{s+1}} du$$

$$= [x]x^{-s} + s\left[\frac{u^{-s+1}}{-s+1}\right]_1^x - s\int_1^x \frac{\{u\}}{u^{s+1}} du$$

$$= [x]x^{-s} - s\frac{x^{-s+1}}{s-1} + \frac{s}{s-1} - s\int_1^x \frac{\{u\}}{u^{s+1}} du.$$

Letting $x \to \infty$, we have for $\Re(s) > 1$

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{\{u\}}{u^{s+1}}du.$$

Here is where we can see the meromorphic continuation. The given integral converges not only for $\Re(s) > 1$ but for $\Re(s) > 0$, since the numerator is bounded between 0 and 1 and (with $\sigma = \Re(s)$)

$$\int_1^\infty \frac{1}{u^{\sigma+1}}du = \left[\frac{u^{-\sigma}}{-\sigma}\right]_1^\infty = \frac{1}{\sigma}.$$

The convergence of the improper integral implies that it defines a holomorphic function of s in $\Re(s) > 0$. The second term $s/(s-1) = 1 + 1/(s-1)$ has a pole at $s = 1$ with residue 1.

(8) Poisson summation formula. Let f be in $L^1(\mathbb{R})$ with $\hat{f} \in L^1(\mathbb{R})$ and both of bounded variation. function, we define its Fourier transform by

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

Then

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \hat{f}(m).$$

We construct a periodic function

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

which has period 1 and convergence under the assumptions on f. We calculate the Fourier coefficients for $g(x)$:

$$\hat{g}(m) = \int_0^1 g(x)e^{-2\pi imx}\,dx = \sum_{n\in\mathbb{Z}}\int_0^1 f(x+n)e^{-2\pi imx}\,dx$$

$$= \sum_{n\in\mathbb{Z}}\int_n^{n+1} f(y)e^{-2\pi imy}\,dy = \int_{-\infty}^{\infty} f(y)e^{-2\pi imy}\,dy = \hat{f}(m).$$

By the Fourier inversion formula for g:

$$g(x) = \sum_{m\in\mathbb{Z}} \hat{g}(m)e^{2\pi imx}.$$

We plug $x = 0$ to get the Poisson summation formula.

(9) Let $T(z) = \dfrac{az+b}{cz+d}$. Assume that it maps the real line to the real line. Show that we can choose a, b, c, d to be real numbers.

Proof: Set $\mu = T(0) = b/d$, $\nu = T(\infty) = a/c$ and $s = T(1) = (a+b)/(c+d)$, which are real by assumption. This gives

$$a + b = s(c + d) = \nu c + \mu d \Longrightarrow (s - \nu)c = (\mu - s)d.$$

If $(s-\nu)(\mu - s) \neq 0$, then $c = d(\mu - s)/(s - \nu) = \rho d$, with ρ real. Also $a = \nu c = \nu\rho d$, $b = \mu d$. These imply

$$Tz = \frac{az + b}{cz + d} = \frac{\nu\rho dz + \mu d}{\rho dz + d} = \frac{\nu\rho z + \mu}{\rho z + 1},$$

which has real coefficients. If $s = \nu = \mu$, $b = sd$, $a = sc$ and

$$Tz = \frac{az + b}{cz + d} = \frac{scz + sd}{cz + d} = s \in \mathbb{R}.$$

The case $d = 0$ is even easier: $Tz = (a/d)z + (b/d)$ with $b/d \in \mathbb{R}$ ($T(0)$ is real), and $T(1) = a/d + b/d \in \mathbb{R}$. So $a/d \in \mathbb{R}$.

(10) Show that

$$d(z, w) = \ln\left(\frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}\right)$$

for the hyperbolic distance between z and w.

Proof: If $z = ia$ and $w = ia$, with $a > b > 0$ the formula reduces to

$$\ln\frac{a + b + a - b}{a + b - a + b} = \ln\frac{a}{b}$$

which agrees with the proof of Theorem 5.1.

If z and w have the same real part, the right-hand side of the formula is the same as for $\Im(z)$ and $\Im(w)$, so it is again obvious.

If $\Re(z) \neq \Re(w)$, then we use the l.f.t. g introduced in Th. 5.1. We have

$$d(z, w) = d(g(z), g(w)) = \ln\frac{|gz - \overline{gw}| + |gz - gw|}{|gz - \overline{gw}| - |gz - gw|}.$$

So it suffices to prove that the fraction above is equal to

$$\frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}.$$

An easy calculation shows that ($\alpha, \beta \in \mathbb{R}$)

$$g(z) - \overline{gw} = \frac{z - \beta}{z - \alpha} - \frac{\bar{w} - \beta}{\bar{w} - \alpha} = \frac{(\beta - \alpha)(z - \bar{w})}{(z - \alpha)(\bar{w} - \alpha)}.$$

A similar calculation holds with w instead of $\bar{w}$. Take absolute values, cancel $\beta - \alpha$ and we get the result.

(11) Show that

$$\cosh d(z,w) = 1 + 2u(z,w).$$

This follows the pattern of the previous exercise. If $z = ia$, $w = ib$

$$u(z,w) = \frac{(b-a)^2}{4ab} \Longrightarrow 2u(z,w) + 1 = \frac{b^2+a^2}{2ab} = \frac{b/a + a/b}{2}$$
$$= \frac{e^{\ln(b/a)} + e^{-\ln(b/a)}}{2} = \cosh \ln(b/a) = \cosh d(z,w).$$

For the general case use g. It suffices to prove that

$$u(gz, gw) = u(z,w).$$

Since $\Im(gz) = (\beta - \alpha)\Im(z)/\left|z-\alpha\right|^2$ and similarly for $\Im(gw)$ we get

$$u(gz,gw) = \frac{\frac{(\alpha-\beta)^2|z-w|^2}{|z-\alpha|^2|w-\alpha|^2}}{4\frac{(\beta-\alpha)\Im(z)}{|z-\alpha|^2}\frac{(\beta-\alpha)\Im(w)}{|w-\alpha|^2}} = \frac{\left|z-w\right|^2}{4\Im(z)\Im(w)} = u(z,w).$$

(12) (The Laplacian in polar coordinates) Show that

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{\cosh r}{\sinh r}\frac{\partial}{\partial r} + \frac{1}{\sinh^2 r}\frac{\partial^2}{\partial \theta^2}$$

The hyperbolic metric in polar coordinates is:

$$ds^2 = dr^2 + (\sinh r)^2 d\theta^2$$

This can be seen by the hyperbolic metric in the disc

$$ds^2 = \frac{4(dx^2+dy^2)}{(1-|z|^2)^2}$$

We compute the relation between $\rho = |z|$ and r. By putting the origin of the polar coordinates at 0 in the Poincaré disc we get

$$r = \int_0^{|z|} \frac{2}{1-\rho^2}d\rho = \int_0^{\rho} \frac{1}{1+\rho} + \frac{1}{1-\rho}d\rho = \ln\frac{1+\rho}{1-\rho}$$

We solve for $\rho = \tanh r/2$. This gives

$$1 - \tanh^2(r/2) = 1/\cosh^2(r/2), \quad d\rho = \frac{1}{2\cosh^2(r/2)}.$$

Now as is well-known in Euclidean polar coordinates

$$dx^2 + dy^2 = d\rho^2 + \rho^2 d\theta^2$$

This gives

$$ds^2 = \frac{4(d\rho^2 + \rho^2 d\theta^2)}{(1-\rho^2)^2} = 4\frac{(1/4)dr^2/\cosh^4(r/2) + \tanh^2(r/2)d\theta^2}{\cosh^{-4}(r/2)}$$
$$= dr^2 + 4\cosh^2(r/2)\sinh^2(r/2)d\theta^2 = dr^2 + \sinh^2 r d\theta^2.$$

Now we compute the Laplacian in these coordinates: The metric matrix is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \sinh^2 r \end{pmatrix}, \quad g_{ij}^{-1} = g^{ij} = \begin{pmatrix} 1 & 0 \\ 0 & \sinh^{-2} r \end{pmatrix}$$

so that $g = \det(g_{ij}) = \sinh^2 r$. The general form for the Laplacian in the metric is

$$\Delta = \frac{1}{\sqrt{g}} \sum_{ij} \frac{\partial}{\partial x_i} \sqrt{g} g^{ij} \frac{\partial}{\partial x_j}$$

which gives for the hyperbolic Laplacian

(10.1)

$$\Delta = \frac{1}{\sinh r}\left(\frac{\partial}{\partial r}(\sinh r \frac{\partial}{\partial r}) + \frac{\partial}{\partial \theta}(\sinh^{-1} r \frac{\partial}{\partial \theta})\right) = \frac{\partial^2}{\partial r^2} + \frac{\cosh r}{\sinh r}\frac{\partial}{\partial r} + \frac{1}{\sinh^2 r}\frac{\partial^2}{\partial \theta^2}$$

(13) Find the expression of the hyperbolic laplacian in u and θ coordinates.

We have for the u variable $\cosh r = 1 + 2u$. We have $\sinh^2 r = \cosh^2 r - 1 = (2u+1)^2 - 1 = 4u(u+1)$. Moreover, by differentiation w e get

$$(\sinh r) r_u = 2 \Longrightarrow r_u = \frac{2}{\sinh r}$$

and

$$\cosh r (r_u)^2 + (\sinh r)(r)_{uu} = 0 \Longrightarrow r_{uu} = -\frac{4\cosh r}{\sinh^3 r}$$

We have for the differentiation operators, using the chain rule (note that since r determines u and vice-versa (irrespective of θ, we do not need to worry about the θ variables).

$$\frac{\partial}{\partial u} = \frac{\partial r}{\partial u}\frac{\partial}{\partial r}, \quad \frac{\partial^2}{\partial u^2} = (r_u)^2 \frac{\partial^2}{\partial r^2} + r_{uu}\frac{\partial}{\partial r}.$$

We plug them into (10.1)

$$\Delta = (r_u)^{-2}\left(\frac{\partial^2}{\partial u^2} - r_{uu} r_u^{-1} \frac{\partial}{\partial u}\right) + \frac{\cosh r}{\sinh r} r_u^{-1} \frac{\partial}{\partial u} + \frac{1}{\sinh^2 r}\frac{\partial^2}{\partial \theta^2}$$

Now $r_u^{-2} = (1/4)\sinh^2 r = u(u+1)$. For the coefficient of $\partial/\partial u$ we get:

$$-r_u^{-3} r_{uu} + \frac{\cosh r}{\sinh r} r_u^{-1} = -\frac{\sinh^3 r}{8}\frac{-4\cosh r}{\sinh^3 r} + \frac{\cosh r}{\sinh r}\frac{\sinh r}{2} = \cosh r = 2u + 1.$$

This gives for the Laplacian in the (u, θ) coordinates:

(10.2)
$$\Delta = u(u+1)\frac{\partial^2}{\partial u^2} + (2u+1)\frac{\partial}{\partial u} + \frac{1}{4u(u+1)}\frac{\partial^2}{\partial \theta^2}.$$

Remark. Notice the formulas 1.16, 1.20, 1.21 in [16]. They differ from the above in the sense that $2\theta = \phi$. This is due to the fact that he uses the matrix

$$T = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

to act at the point i as a rotation. But in the action of $\mathrm{SL}(2,\mathbb{R})$ this acts as rotation of angle 2ϕ: this is seen by computing its derivative at i:

$$T'(z) = \frac{1}{(-\sin\phi z + \cos\phi)^2} \Longrightarrow T'(i) = \frac{1}{(e^{-i\phi})^2} = e^{2i\phi}$$

(14) Show by direct calculation that the hyperbolic volume element

$$d\mu(z) = \frac{dxdy}{y^2}$$

is invariant under the action of $\mathrm{SL}_2(\mathbb{R})$.

Using the language of differential forms we have

$$\frac{dx\,dy}{y^2} = \frac{dx \wedge dy}{y^2} = \frac{i}{2}\frac{dz \wedge d\bar{z}}{(\Im z)^2}.$$

Since $d(\gamma z) = \gamma'(z)dz$ we get

$$\frac{i}{2}\frac{d(\gamma z)\wedge d\overline{(\gamma z)}}{(\Im(\gamma z))^2} = \frac{i}{2}\frac{\gamma'(z)dz \wedge \overline{\gamma'(z)}d\bar{z}}{y^2/\left|cz+d\right|^4} = \frac{i}{2}\frac{\left|\gamma'(z)\right|^2 dz\wedge d\bar{z}}{y^2/\left|cz+d\right|^4} = \frac{i}{2}\frac{dz\wedge d\bar{z}}{y^2},$$

using (4.2), and (5.1).

(15) (i) Show that the Euclidean Laplace operator can be calculated as

$$\Delta_{eucl} = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}} = 4\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z}.$$

(ii) Suppose that U and V are open sets in the complex plane. Prove that if $f : U \to V$ and $g : V \to \mathbb{C}$ are two functions that are differentiable in the real sense (in x and y) and $h = g \circ f$, then the complex version of the chain rule is

$$\frac{\partial h}{\partial z} = \frac{\partial g}{\partial z}\frac{\partial f}{\partial z} + \frac{\partial g}{\partial \bar{z}}\frac{\partial \bar{f}}{\partial z}, \quad \frac{\partial h}{\partial \bar{z}} = \frac{\partial g}{\partial z}\frac{\partial f}{\partial \bar{z}} + \frac{\partial g}{\partial \bar{z}}\frac{\partial \bar{f}}{\partial \bar{z}}.$$

(iii) Show that Δ commutes with any element of $\mathrm{SL}_2(\mathbb{R})$, i.e

$$\Delta(f(\gamma z)) = (\Delta f)(\gamma z).$$

Proof:

(i) We have for a function f with continuous second partial derivatives

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right), \quad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right)$$

which gives

$$4\frac{\partial^2 f}{\partial z \partial \bar{z}} = \left(\frac{\partial}{\partial x} - i\frac{\partial f}{\partial y}\right)(f_x + if_y) = \left(\frac{\partial}{\partial x} - i\frac{\partial f}{\partial y}\right) f_x + i\left(\frac{\partial}{\partial x} - i\frac{\partial f}{\partial y}\right) f_y$$

$$= f_{xx} - if_{yx} + i(f_{xy} - if_{yy}) = f_{xx} - if_{yx} + if_{xy} + f_{yy} = f_{xx} + f_{yy} = \Delta_{eucl} f,$$

since the mixed partial f_{xy}, f_{yx} are equal for functions with continuous second partial derivatives. Reversing the order of the calculation gives $4\partial_{\bar{z}}\partial_z f = \Delta_{eucl} f$.

(ii) Set $f(x,y) = u(x,y) + iv(x,y)$, i.e., $(x,y) \to (u(x,y), v(x,y)) \to h(u,v)$. By the standard chain rule for functions of two variables we have

$$\begin{aligned} \frac{\partial h}{\partial x} &= \frac{\partial g}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial y}\frac{\partial v}{\partial x} \\ \frac{\partial h}{\partial y} &= \frac{\partial g}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial g}{\partial y}\frac{\partial v}{\partial y}. \end{aligned} \tag{10.3}$$

Moreover,

$$\begin{aligned} \frac{\partial f}{\partial z} &= \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) = \frac{1}{2}(u_x + iv_x - i(u_y + iv_y)) \\ \frac{\partial \bar{f}}{\partial z} &= \frac{1}{2}\left(\frac{\partial \bar{f}}{\partial x} - i\frac{\partial \bar{f}}{\partial y}\right) = \frac{1}{2}(u_x - iv_x - i(u_y - iv_y)). \end{aligned} \tag{10.4}$$

By the definition of $\partial/\partial z$ and $\partial/\partial \bar{z}$ we have

$$\begin{aligned} \frac{\partial g}{\partial z} &= \frac{1}{2}\left(\frac{\partial g}{\partial x} - i\frac{\partial g}{\partial y}\right) \\ \frac{\partial g}{\partial \bar{z}} &= \frac{1}{2}\left(\frac{\partial g}{\partial x} + i\frac{\partial g}{\partial y}\right). \end{aligned}$$

We add and subtract the last two equations to get

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial z} + \frac{\partial g}{\partial \bar{z}}, \quad \frac{\partial g}{\partial y} = \frac{1}{i}\left(\frac{\partial g}{\partial \bar{z}} - \frac{\partial g}{\partial z}\right).$$

We substitute the last equations to (10.3), multiply the second equation in (10.3) by i, subtract them to get

$$\begin{aligned} \frac{\partial h}{\partial z} &= \frac{\partial g}{\partial x}\left(\frac{1}{2}\frac{\partial u}{\partial x} - \frac{1}{2}i\frac{\partial u}{\partial y}\right) + \frac{\partial g}{\partial y}\left(\frac{1}{2}\frac{\partial v}{\partial x} - \frac{1}{2}i\frac{\partial v}{\partial y}\right) \\ &= (g_z + g_{\bar{z}})\frac{1}{2}(u_x - iu_y) + \frac{1}{i}(g_{\bar{z}} - g_z)\frac{1}{2}(v_x - iv_y) \\ &= g_z\frac{1}{2}(u_x - iu_y + iv_x + v_y) + g_{\bar{z}}\frac{1}{2}(u_x - iu_y - iv_x - v_y) \\ &= g_z f_z + g_{\bar{z}}\bar{f}_z \end{aligned} \tag{10.5}$$

using equations (10.4). Similarly we get

$$\begin{aligned}
(10.6) \qquad \frac{\partial h}{\partial \bar{z}} &= \frac{\partial g}{\partial x}\left(\frac{1}{2}\frac{\partial u}{\partial x}+\frac{1}{2}i\frac{\partial u}{\partial y}\right)+\frac{\partial g}{\partial y}\left(\frac{1}{2}\frac{\partial v}{\partial x}+\frac{1}{2}i\frac{\partial v}{\partial y}\right)\\
&= (g_z+g_{\bar z})\frac{1}{2}(u_x+iu_y)+\frac{1}{i}(g_{\bar z}-g_z)\frac{1}{2}(v_x+iv_y)\\
&= g_z\frac{1}{2}(u_x+iu_y+iv_x-v_y)+g_{\bar z}\frac{1}{2}(u_x+iu_y-iv_x+v_y)\\
&= g_z f_{\bar z}+g_{\bar z}\bar{f}_{\bar z},
\end{aligned}$$

since

$$\begin{aligned}
\frac{\partial f}{\partial \bar z} &= \frac{1}{2}\left(\frac{\partial f}{\partial x}+i\frac{\partial f}{\partial y}\right)=\frac{1}{2}(u_x+iv_x+i(u_y+iv_y))\\
\frac{\partial \bar f}{\partial \bar z} &= \frac{1}{2}\left(\frac{\partial \bar f}{\partial x}+i\frac{\partial \bar f}{\partial y}\right)=\frac{1}{2}(u_x-iv_x+i(u_y-iv_y)).
\end{aligned}$$

(iii) We remark that, since $\gamma(z)$ is holomorphic we have

$$\frac{\partial \gamma}{\partial z}=\gamma'(z), \quad \frac{\partial \gamma}{\partial \bar z}=0.$$

Also $\partial\bar\gamma/\partial\bar z=\overline{\gamma'(z)}$. This gives by applying (ii)

$$\frac{\partial}{\partial \bar z}f\circ\gamma=\left(\frac{\partial f}{\partial \bar z}\right)\circ\gamma\cdot\overline{\gamma'}.$$

Now we calculate

$$\frac{\partial}{\partial z}\left(\left(\frac{\partial f}{\partial \bar z}\right)\circ\gamma\cdot\overline{\gamma'}\right)=\left(\frac{\partial^2 f}{\partial z\partial \bar z}\right)\circ\gamma\cdot\gamma'\cdot\overline{\gamma'}=\left(\frac{\partial^2 f}{\partial z\partial \bar z}\circ\gamma\right)\left|\gamma'\right|^2,$$

since an application of the product rule gives $\partial\overline{\gamma'}/\partial z=0$ ($\overline{\gamma'}$ is anti-holomorphic, so is killed by $\partial/\partial z$. So we have proved

$$\Delta_{eucl}(f\circ\gamma)=(\Delta_{eucl}f)\circ\gamma\cdot\left|\gamma'\right|^2,$$

which gives

$$\Delta(f\circ\gamma)=\left(y^2\Delta_{eucl}\right)(f\circ\gamma)=y^2(\Delta_{eucl}f)\circ\gamma\cdot\left|\gamma'\right|^2=\frac{y^2}{|cz+d|^4}(\Delta_{eucl}f)\circ\gamma$$

$$=(\Im(\gamma z)^2)(\Delta_{eucl}f)\circ\gamma=(y^2\Delta_{eucl}f)\circ\gamma=(\Delta f)\circ\gamma.$$

References

[1] Ahlfors, Lars V. Complex analysis. An introduction to the theory of analytic functions of one complex variable. Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, 1978. xi+331 pp. ISBN: 0-07-000657-1

[2] Davenport, Harold Multiplicative number theory. Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, 74. Springer-Verlag, New York, 2000. xiv+177 pp. ISBN: 0-387-95097-4

[3] Kowalski, Emmanuel Un cours de théorie analytique des nombres. (French) [A course on analytic number theory] Cours Spécialisés [Specialized Courses], 13. Société Mathématique de France, Paris, 2004. x+232 pp. ISBN: 2-85629-161-9

[4] Patterson, S. J. An introduction to the theory of the Riemann Zeta-function, Cambridge studies in advanced mathematics 14, 1988.

[5] Murty, M. Ram Problems in analytic number theory. Second edition. Graduate Texts in Mathematics, 206. Readings in Mathematics. Springer, New York, 2008. xxii+502 pp. ISBN: 978-0-387-72349-5 11-01

[6] Stein, Elias M.; Shakarchi, Rami Complex analysis. Princeton Lectures in Analysis, II. Princeton University Press, Princeton, NJ, 2003. xviii+379 pp. ISBN: 0-691-11385-8

[7] Titchmarsh, E. C. The theory of the Riemann zeta-function. Second edition. Edited and with a preface by D. R. Heath-Brown. The Clarendon Press, Oxford University Press, New York, 1986. x+412 pp. ISBN: 0-19-853369-1

[8] Titchmarsh, E. C. The theory of functions, Oxford University Press, second edition.

[9] Stromberg, Karl R. Introduction to classical real analysis. Wadsworth International Mathematics Series. Wadsworth International, Belmont, Calif., 1981. ix+575 pp. ISBN: 0-534-98012-0

General references for automorphic forms

[10] Bump, Daniel Automorphic forms and representations. Cambridge Studies in Advanced Mathematics, 55. Cambridge University Press, Cambridge, 1997. xiv+574 pp. ISBN: 0-521-55098-X

[11] Borel, Armand Automorphic forms on $\mathrm{SL}_2(R)$. Cambridge Tracts in Mathematics, 130. Cambridge University Press, Cambridge, 1997. x+192 pp. ISBN: 0-521-58049-8

[12] Iwaniec, Henryk Topics in classical automorphic forms. Graduate Studies in Mathematics, 17. American Mathematical Society, Providence, RI, 1997. xii+259 pp.

Books on Spectral theory of automorphic forms

[13] Bruggeman, Roelof W. Families of automorphic forms. Monographs in Mathematics, 88. Birkhäuser Boston, Inc., Boston, MA, 1994. x+317 pp. ISBN: 3-7643-5046-6

[14] Hejhal, Dennis A. The Selberg trace formula for $\mathrm{PSL}(2, R)$. Vol. I. Lecture Notes in Mathematics, Vol. 548. Springer-Verlag, Berlin-New York, 1976. vi+516 pp.

[15] Hejhal, Dennis A. The Selberg trace formula for $\mathrm{PSL}(2,\ R)$. Vol. 2. Lecture Notes in Mathematics, 1001. Springer-Verlag, Berlin, 1983. viii+806 pp. ISBN: 3-540-12323-7

[16] Iwaniec, Henryk Spectral methods of automorphic forms. Second edition. Graduate Studies in Mathematics, 53. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002. xii+220 pp. ISBN: 0-8218-3160-7

[17] Venkov, Alexei B. Spectral theory of automorphic functions and its applications. Translated from the Russian by N. B. Lebedinskaya. Mathematics and its Applications (Soviet Series), 51. Kluwer Academic Publishers Group, Dordrecht, 1990. xiv+176 pp. ISBN: 0-7923-0487-X

[18] Venkov, A. B. Spectral theory of automorphic functions. A translation of Trudy Mat. Inst. Steklov. 153 (1981). Proc. Steklov Inst. Math. 1982, no. 4(153), ix+163 pp. (1983).

Books on discrete groups and hyperbolic geometry

[19] Katok, Svetlana Fuchsian groups. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1992. x+175 pp. ISBN: 0-226-42582-7; 0-226-42583-5

[20] Beardon, Alan F. The geometry of discrete groups. Graduate Texts in Mathematics, 91. Springer-Verlag, New York, 1983. xii+337 pp. ISBN: 0-387-90788-2

On the analytic continuation of $GL(2)$ Eisenstein series.

[21] Lang, Serge $\mathrm{SL}_2(R)$. Reprint of the 1975 edition. Graduate Texts in Mathematics, 105. Springer-Verlag, New York, 1985. xiv+428 pp. ISBN: 0-387-96198-4

[22] Faddeev, L. D. The eigenfunction expansion of Laplace's operator on the fundamental domain of a discrete group on the Lobačevskiĭplane. (Russian) Trudy Moskov. Mat. Obšč. 17 1967 323–350.

[23] Pavlov, B. S.; Faddeev, L. D. Scattering theory and automorphic functions. (Russian) Boundary value problems of mathematical physics and related questions in the theory of functions, 6. Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI) 27 (1972), 161–193.

[24] Lax, Peter D.; Phillips, Ralph S. Scattering theory for automorphic functions. Annals of Mathematics Studies, No. 87. Princeton Univ. Press, Princeton, N.J., 1976. x+300 pp. ISBN: 0-691-08179-4

[25] Colin de Verdière, Yves Une nouvelle démonstration du prolongement méromorphe des séries d'Eisenstein. (French) [A new proof of the meromorphic continuation of Eisenstein series] C. R. Acad. Sci. Paris Sér. I Math. 293 (1981), no. 7, 361–363.

[26] Colin de Verdière, Yves Pseudo-laplaciens. II. (French) [Pseudo-Laplacians. II] Ann. Inst. Fourier (Grenoble) 33 (1983), no. 2, 87–113.

On the prime geodesic theorem

[27] Sarnak, Peter Class numbers of indefinite binary quadratic forms. J. Number Theory 15 (1982), no. 2, 229–247.

[28] Iwaniec, Henryk Prime geodesic theorem. J. Reine Angew. Math. 349 (1984), 136–159. 11F11 (11F72 11L05)

[29] Luo, Wen Zhi; Sarnak, Peter Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}^2$. Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 207–237.

On small eigenvalues

[30] Huxley, M. N. Introduction to Kloostermania. Elementary and analytic theory of numbers (Warsaw, 1982), 217–306, Banach Center Publ., 17, PWN, Warsaw, 1985.

[31] Luo, W.; Rudnick, Z.; Sarnak, P. On Selberg's eigenvalue conjecture. Geom. Funct. Anal. 5 (1995), no. 2, 387–401.

[32] Kim, Henry H. Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2. With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak. J. Amer. Math. Soc. 16 (2003), no. 1, 139–183

On the hyperbolic lattice-counting problem

[33] Chamizo, Fernando, The large sieve in Riemann surfaces. Acta Arith. 77 (1996), no. 4, 303–313. Some applications of large sieve in Riemann surfaces. Acta Arith. 77 (1996), no. 4, 315–337.

[34] Huber, Heinz Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen. II. (German) Math. Ann. 142 1960/1961 385–398.

[35] Huber, Heinz Über eine neue Klasse automorpher Funktionen und ein Gitterpunktproblem in der hyperbolischen Ebene. I. (German) Comment. Math. Helv. 30 (1956), 20–62 (1955).

[36] Huber, Heinz Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen. (German) Math. Ann. 138 1959 1–26

[37] Patterson, S. J. A lattice-point problem in hyperbolic space. Mathematika 22 (1975), no. 1, 81–88.

[38] Phillips, Ralph; Rudnick, Zeév The circle problem in the hyperbolic plane. (English summary) J. Funct. Anal. 121 (1994), no. 1, 78–116.

On Weyl's Law

[39] Luo, Wenzhi Nonvanishing of L-values and the Weyl law. Ann. of Math. (2) 154 (2001), no. 2, 477–502.

[40] R. Phillips and P. Sarnak, On cusp forms for co-finite subgroups of $\mathrm{PSL}(2, R)$, Invent. math. 80 (1985), 339–364.

[41] Phillips, R.; Sarnak, P. Cusp forms for character varieties. Geom. Funct. Anal. 4 (1994), no. 1, 93–118.

[42] Phillips, R.; Sarnak, P. Perturbation theory for the Laplacian on automorphic functions. J. Amer. Math. Soc. 5 (1992), no. 1, 1–32.

[43] S. Wolpert, Disappearance of cusp forms in special families, Ann. of Math. 139 (1994), 239–291.

Göttingen lectures notes on the Selberg trace formula

[44] Selberg, Atle Collected papers. Vol. I. With a foreword by K. Chandrasekharan. Springer-Verlag, Berlin, 1989. vi+711 pp. ISBN: 3-540-18389-2

Review article

[45] Sarnak, Peter Spectra of hyperbolic surfaces. Bull. Amer. Math. Soc. (N.S.) 40 (2003), no. 4, 441–478 (electronic).

First article on Mass cusp forms

[46] Maass, Hans Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. (German) Math. Ann. 121, (1949). 141–183.

Calculations of fundamental domains, indices, genera for congruence subgroups

[47] Schoeneberg, Bruno Elliptic modular functions: an introduction. Translated from the German by J. R. Smart and E. A. Schwandt. Die Grundlehren der mathematischen Wissenschaften, Band 203. Springer-Verlag, New York-Heidelberg, 1974. viii+233 pp.

[48] Shimura, Goro Introduction to the arithmetic theory of automorphic functions. Reprint of the 1971 original. Publications of the Mathematical Society of Japan, 11. Kan Memorial Lectures, 1. Princeton University Press, Princeton, NJ, 1994. xiv+271 pp. ISBN: 0-691-08092-5.

On the Gauss circle problem and exponential sums

[49] S. W. Graham, G. Kolesnik, Van der Corput's method for exponential sums, London Mathematical Society Lecture Note Series 126 (Cambridge University Press, 1991).

[50] M. Huxley, Exponential sums and lattice points iii. Proceedings of the London Mathematical Society, 87(3): 591609, 2003

[51] H. Iwaniec and C. J. Mozzochi, 'On the divisor and circle problems', J. Number Theory 29 (1988) 60–93.

Department of Mathematics, University College London, Gower Street, London WC1E 6BT

E-mail address: petridis@member.ams.org, petridis@math.ucl.ac.uk

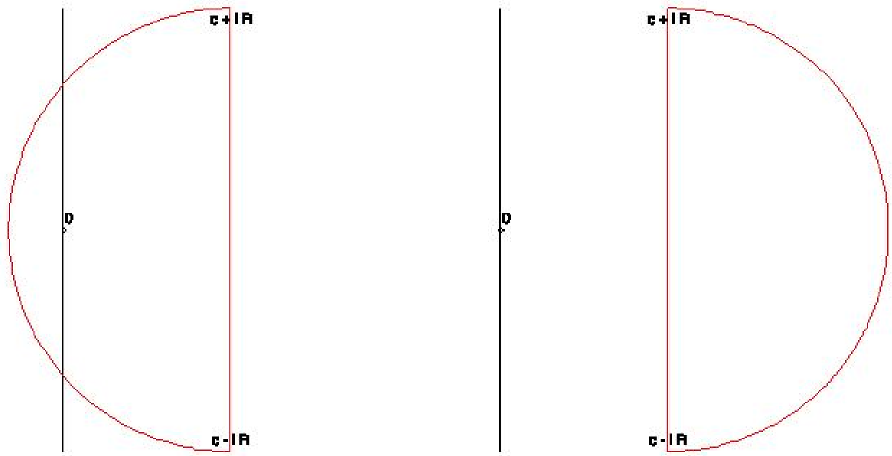


FIGURE 6. Contours for the Perron formula: on the left $x > 1$, on the right $x < 1$.