LANGENHOP LECTURE

AN ELEMENTARY INTRODUCTION TO LANGLANDS PROGRAM

Freydoon Shahidi

Purdue University

May 16, 2016

Southern Illinois University

Abstract

We use a simple counting function to introduce two different aspects of Langlands program through some basic special cases: Spectral theory of Maass forms and Artin reciprocity law. The talk is aimed at a general audience with some very basic mathematical familiarity. But no specialized knowledge of number theory is assumed.

Let us start with the counting function

$r(m)$ = Number of points (a, b) with integer coordinates such that $a^2 + b^2 = m$, where m is a natural (whole) number.

Or said in mathematical notation

$$r(m) = \#\{(a, b) \in \mathbb{Z}^2 | a^2 + b^2 = m\}, \qquad (1)$$

where $\mathbb{Z}$ denotes the set (ring) of integers.

We now mention two well–known problems.

Problem 1. Determine all whole numbers m such that $r(m) \neq 0$.

Problem 2. Let x be a positive real number. Estimate

$$P(x) := \sum_{m \leq x} r(m) = \text{ The number of integral points inside the circle of radius } \sqrt{x} \text{ centered at the origin.} \quad (2)$$

Here are the answers:

Answer to Problem 1 (Fermat, Euler, Gauss). Write

$$m = m_1 n^2,$$

with m_1 and n integers, where m_1 is a product of distinct primes (square free). The number $r(m) \neq 0$ if and only if every prime divisor of m_1 is either 2 or $p \equiv 1(4)$, said p is congruent to 1 modulo 4, i.e., 1 is the remainder of division of p by 4.

$$(5 = 1^2 + 2^2, 17 = 1^2 + 4^2, 29 = 2^2 + 5^2,$$
$$37 = 1^2 + 6^2, 41 = 4^2 + 5^2).$$

This introduces us to the notion of "residue modulo an integer r". This means all possible remainders that one gets upon dividing integers by r. Clearly that is the set

$$\{0, 1, 2, \ldots, r-1\} := \mathbb{Z}/r\mathbb{Z} \tag{3}$$

We note that this set is closed under addition and multiplication if we continue to take residues (remainders).

This is an example of the structure of a "ring". It satisfies similar structural properties as the **ring of integers** $\mathbb{Z}$. For example if x is in

$$\{0, 1, \ldots, r\} = \mathbb{Z}/r\mathbb{Z},$$

$-x = r - x$ since

$$-x + (r - x) = r \equiv 0 (\text{mod } r).$$

Note that integers don't have reciprocals (inverses) unless they are 1 or -1. On the hand every integer $0 < x < r$ which is relatively prime to r, i.e., doesn't have any common factor with r has an inverse mod r.

Example. In $\mathbb{Z}/6\mathbb{Z} = \{0, 1, 2, 3, 4, 5\}$, $5^{-1} = 5$, since $5 \cdot 5 = 25 \equiv 1(6)$. Thus 1 and 5 are the only **invertible** elements of $\mathbb{Z}/6\mathbb{Z}$.

Conclusion. If $r = p$ is a prime, then every non–zero member of $\mathbb{Z}/p\mathbb{Z}$ is invertible. Such a ring is called a **field** and $\mathbb{Z}/p\mathbb{Z}$ is an example of what we call a **finite field**. We will return to this momentarily.

Answer to Problem 2 (Gauss Circle Problem)

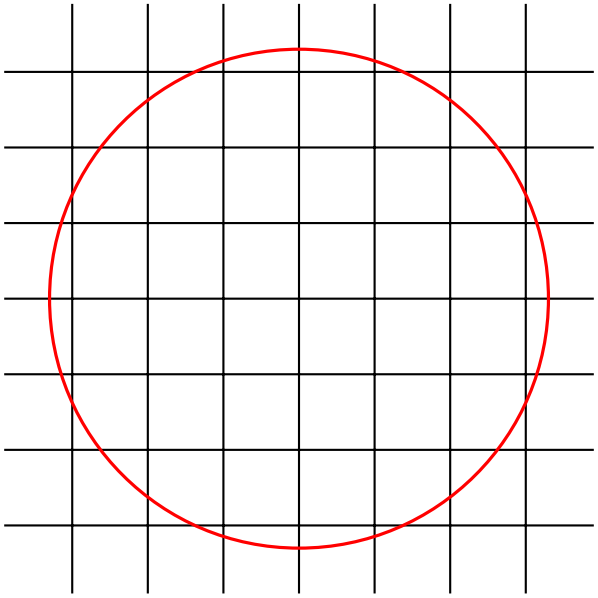
$$P(x) \sim \pi x + O(\sqrt{x}), \tag{4}$$

($\pi = 3.1415\ldots$) where $O(\sqrt{x})$ means

$$O(\sqrt{x})/\sqrt{x}$$

is bounded. In other words πx indicates the **leading** or **dominant** behaviour of $P(x)$ as x gets larger and larger. The remainder or error term which gives the behaviour of $P(x) - \pi x$ is expected to be $O(x^{\frac{1}{4}+\varepsilon})$. But this is not yet proved and the best error term so far is $O(x^{\frac{23}{73}+\varepsilon})$. Here ε is any small positive number.

It is rather simple to show (4):
$P(x)$ is approximately the number of unit squares whose corners have integral coordinates and lie inside the circle of radius $\sqrt{x}$ centered at the origin:



Gauss Circle Problem

A good estimate of this number is the area of the circle and thus

$$P(x) \sim \pi(\sqrt{x})^2 = \text{ area of the circle.}$$

The error term is then no larger than the perimeter of the circle which is $O(\sqrt{x})$.

Problem 2 is a question about counting points in a "grid" or "lattice" inside a domain, here a circle is a "Euclidean space". It can be asked about other types of spaces, for example, "hyperbolic spaces" like the surface of hyperboloid, and a circle there: the set of points of equal distance from a fixed point — but here the distance is a "**hyperbolic distance**".

The difference between a hyperbolic plane and a Euclidean one is that the hyperbolic one is "curved" by a negative constant curvature, while the Euclidean one is flat.

How do we produce hyperbolic surfaces?

One then takes a Euclidean space and deforms it under the action of a discrete matrix group:

Euclidean space = upper half plane of complex numbers :

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

Discrete group (closed under matrix multiplication and inversion) $= SL_2(\mathbb{Z})$ or its "congruence" subgroups Γ

$$SL_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| ad - bc = 1, a, b, c, d \in \mathbb{Z} \right\}$$

We will then identify two points z, z' in H, if

$$z' = \frac{az + b}{cz + d} := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z$$

for some $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.

The set of orbits obtained under these identifications will be denoted by $\Gamma\backslash H =$ our hyperbolic space. We note that the "simply connected" space H is the "universal cover" of $\Gamma\backslash H$.

Hyperbolic spaces have their own distance functions (you take a geodesic and measure it)

$$d(z,w) = \log \frac{|z-\overline{w}| + |z-w|}{|z-\overline{w}| - |z-w|}, \ (z,w \in H)$$

whose hyperbolic cosine, cosh equals

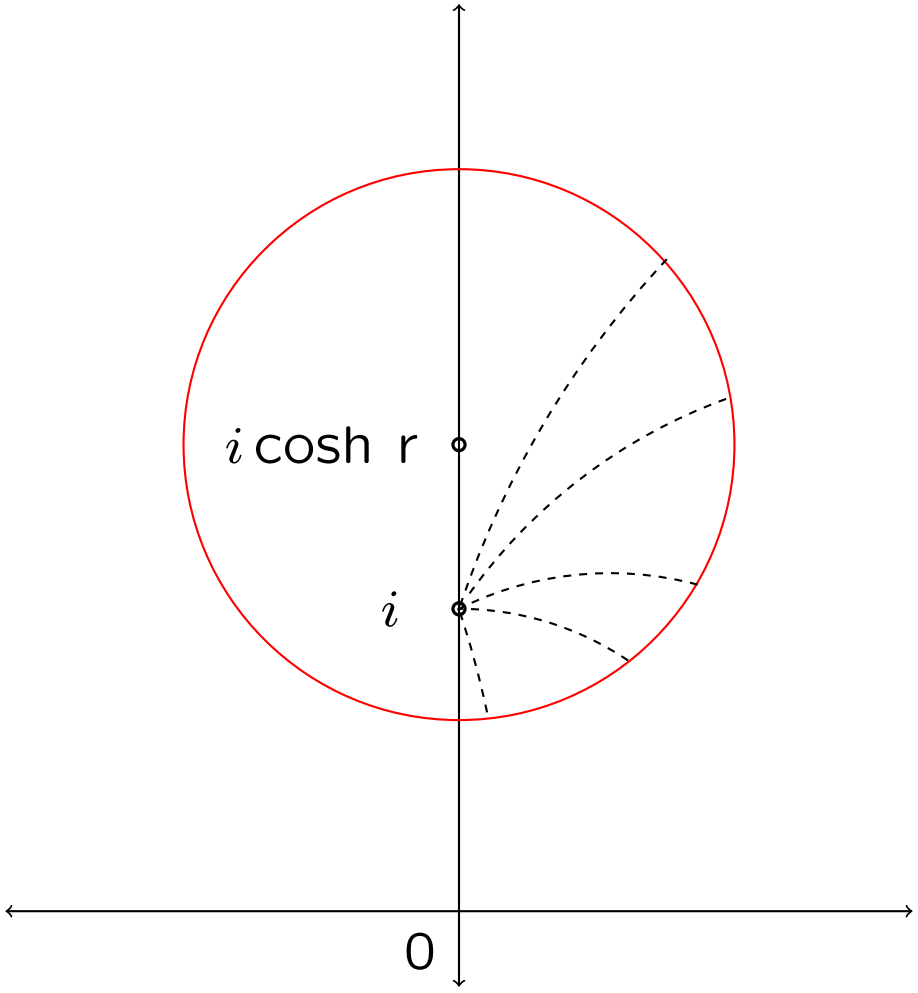
$$\cosh(d(z,w)) = \frac{1}{2}(e^{d(z,w)} + e^{-d(z,w)}) = 1 + 2u(z,w),$$

where

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

Hyperbolic distance is computed by means of the measure $|dz|/y$, where $|dz|$ is the euclidean measure, which is inversely proportional to the imaginary coordinate y as opposed to the euclidean one.

Here is a hyperbolic circle centered at $i = \sqrt{-1}$ and of hyperbolic radius r. The euclidean center of circle is at $i \cosh(r)$. One sees that when radius is depicted in euclidean plane, it seems to defer from one point to another. But their actual hyperbolic length are equal.



A hyperbolic circle

Recall that the euclidean lattice of points with integral coordinates is obtained by taking origin $(0,0) = \mathcal{O}$ and translating under $\mathbb{Z} \times \mathbb{Z}$ by

$$(a,b) \cdot \mathcal{O} = (a,b) = (0+a, 0+b).$$

A hyperbolic lattice is obtained the same way: Take $z \in H$ and let $\gamma \in \Gamma$ act on it as

$$\gamma \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

The lattice is then denoted by $\Gamma \cdot z$.

The problem is again to **estimate the number of (lattice) points in the lattice** $\Gamma \cdot z$ inside the hyperbolic circle of "radius" x, centered at a point $w \in H$, i.e., count

$$\begin{aligned} P_{z,w}(x): &= \#\{\gamma \in \Gamma \mid 2\cosh(d(\gamma \cdot z, w)) \leq x\} \\ &= \#\{\gamma \in \Gamma \mid 2 + 4u(\gamma \cdot z, w) \leq x\}. \end{aligned}$$

Examples and Corollaries: Take $z = w = \sqrt{-1} = i$ and $\Gamma = SL_2(\mathbb{Z})$. Then $P_{z,w}(x)$ gives

$$P_{i,i}(x) = \#\{(a,b,c,d) \in \mathbb{Z}^4 \mid ad - bc = 1 \text{ and } a^2 + b^2 + c^2 + d^2 \leq x\}$$

Gauss's circle area argument in the Euclidean plane no longer works since in hyperbolic geometry the isoperimetric inequality says

$$4\pi A + A^2 \leq L^2,$$

where A and L are the area and perimeter of the hyperbolic circle. Thus there are no dominant quantity and A and L are of same magnitude. The techniques involved are much more subtle.

One has to study a certain space of functions:

$$L^2(\Gamma\backslash H) = \Big\{ f\colon H \to \mathbb{C} \mid f(\gamma\cdot z) = f(z), \gamma \in \Gamma, \int_{\Gamma\backslash H} |f(z)|^2 \, \frac{dxdy}{y^2} < \infty \Big\},$$

where $z = x + y\sqrt{-1} \in H$, and consider the action of hyperbolic Laplacian

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

on $L^2(\Gamma\backslash H)$ and look at its eigenfunctions, i.e., all those $f \neq 0$ in $L^2(\Gamma\backslash H)$ for which there exists a complex number λ such that

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

Clearly f depends on λ. In this setting $\lambda \geq 0$.

If there are no eigenvalues λ for Δ in the range $0 < \lambda < \frac{2}{9}$, then using a trace formula one gets

$$P_{z,w}(x) = \frac{c\pi}{\mathrm{Vol}(\Gamma\backslash H)}\ x + O(x^{2/3}),$$

where $c = 1$ if $-I \notin \Gamma$ and $c = 2$, otherwise.

In the case of $\Gamma = SL_2(\mathbb{Z})$, $-I \in \Gamma$, one gets

$$\begin{aligned} P_{z,w}(x) &= \frac{2\pi}{\pi/3}\ x + O(x^{2/3}) \\ &= 6x + O(x^{2/3}) \end{aligned}$$

and one gets

$$\begin{aligned} \#\{(a,b,c,d) \in \mathbb{Z}^4 | ad - bc = 1, a^2 + b^2 + c^2 \\ + d^2 \leq x\} = 6x + O(x^{2/3}). \end{aligned}$$

The fact that there are no $0 < \lambda < 2/9$, was proved by Henry Kim and myself in 2002, as a first step towards striking improvements on Selberg's conjecture which demands no λ must exist in the range $0 < \lambda < 1/4$, all consequences of new cases of functoriality, a ground breaking conjecture due to Robert Langlands which has revolutionized a major part of number theory.

Another result that one can show by choosing Γ appropriately is

$$\sum_{m \leq x} r(m)r(m+1) = 8x + O(x^{2/3}).$$

Remark 1. The space of functions $L^2(\Gamma\backslash H)$ is an example of automorphic forms. They are the Maass forms or forms of "weight" zero on H, i.e., function on H, invariant under action by Γ. Their analogue when "weight" is 2 are those which are in one–one correspondence with "elliptic curves" by means of Shimura–Taniyama conjecture which was proved by Andrew Wiles for the class of elliptic curve needed in his proof of Fermat's Last Theorem. We recall that a modular form of weight k is a function on H satisfying

$$f(\gamma \cdot z) = (cz + d)^k f(z),$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.

Remark 2. The asymptotic formula for $P_{z,w}(x)$ is obtained by expressing $P_{z,w}(x)$ in terms of an "automorphic" kernel function which can then be studied by means of spectral behaviour of Maass forms.

Reference. An excellent treatment of how $P_{z,w}(x)$ is computed and its consequences is H. Iwaniec's book: Spectral methods of automorphic forms, GTM, Vol. 53, AMS, 2002.

We can now address Problem 1 and explain how intrinsically it is connected to Problem 2 and "Langlands Program".

Consider the equality

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

Thus if all the prime numbers dividing m_1 can be written as sum of two squares, then m will. The problem is then to check for prime p

$$a^2 + b^2 \equiv 0 (\text{mod } p).$$

We thus only need to look at a and b module p, i.e., as members of the field $\mathbb{Z}/p\mathbb{Z}$ in which every element has an inverse, reducing us to

$$(ab^{-1})^2 + 1 \equiv 0 (\text{mod } p).$$

If we now set $x := ab^{-1}$, then we need to know for what p, the equation $x^2 + 1 = 0$ has a solution in $\mathbb{Z}/p\mathbb{Z}$.

Suppose $x^2 + 1 = 0$ has a solution. Then, p being odd, $(p-1)/2$ is an integer and thus

$$x^{p-1} = (x^2)^{(p-1)/2} \equiv (-1)^{(p-1)/2} (\text{mod } p).$$

Now consider the product

$$x \cdot 2x \cdot 3x \ldots (p-1)x = (p-1)! x^{p-1}.$$

Note that no two of the factors are equal mod p, since every non–zero element of $\mathbb{Z}/p\mathbb{Z}$ is invertible and the numbers $1, 2, \ldots, p-1$ are different modulo p.

Thus
$x \cdot 2x \ldots (p-1)x = (p-1)!x^{p-1} \equiv (p-1)!(\text{mod } p)$

or simply

$$x^{p-1} \equiv 1(\text{mod } p).$$

Thus

$$(-1)^{(p-1)/2} = 1$$

or $p \equiv 1(4)$.

For the converse, let

$$x = ((p-1)/2)!$$

which can be shown to satisfy $x^2+1 \equiv 0(\text{mod } p)$. This problem is equivalent to whether the polynomial x^2+1 when reduced modulo p decomposes to two distinct linear factors and the answer depends on congruence class of p modulo 4.

One can ask this about an arbitrary monic polynomial with integral coefficients, whether it "splits completely" modulo a prime p. As we just saw with $f(x) = x^2+1$, f splits for half the primes ($p \equiv 1(4)$) and remains irreducible for the other half ($p \equiv 3(4)$) and these are fairly refined, although elementary, facts.

The answer for an arbitrary f is very subtle. One wants to come up with tools that addresses the problem for a general f. We can consider the field of rational numbers $\mathbb{Q}$ and build a very large field $\overline{\mathbb{Q}}$ containing $\mathbb{Q}$ which contains all the roots of all the polynomials over $\mathbb{Q}$. We then let Gal($\overline{\mathbb{Q}}/\mathbb{Q}$) be the group of all the automorphisms of $\overline{\mathbb{Q}}$ which act like identity on $\mathbb{Q}$, the "Galois group" of $\overline{\mathbb{Q}}/\mathbb{Q}$.

For a given monic integral irreducible polynomial f, we can adjoin all the roots of f to $\mathbb{Q}$ to get a field L and then consider Gal$(L/\mathbb{Q})$ defined the same way and call it Gal (f). If this Galois group is abelian, then the problem of splitting mod p can be resolved through class field theory and Artin reciprocity map.

Let us explain this through a simple example. Assume $L = Q(\sqrt{d})$, and thus L contains the roots of $f(x) = x^2 - d$, d not a square. Then Galois group Gal (f) is easy to determine. The roots of f are $\pm\sqrt{d}$. Note that

$$L = \{a + b\sqrt{d} | a, b \in \mathbb{Q}\}.$$

An automorphism σ of L which acts like identity on $\mathbb{Q}$, will send

$$a + b\sqrt{d} \mapsto a + b\sigma(\sqrt{d}).$$

and thus $\sqrt{d} \mapsto \sigma(\sqrt{d})$. Note that

$$(\sigma(\sqrt{d}))^2 = \sigma(d) = d$$

and thus $\sigma(\sqrt{d}) = \sqrt{d}$ or $-\sqrt{d}$.

Thus

$$\text{Gal } (L/\mathbb{Q}) = \text{ Gal}(f) = \{1, \sigma\},$$

where $\sigma(\sqrt{d}) = -\sqrt{d}$.

If we stay away from the few primes that divide d, the splitting of $f(x)$ modulo a prime p is equivalent to

$$x^2 \equiv d(\text{mod } p)$$

having a solution or not. We define the Legendre symbol $\left(\frac{d}{p}\right)$ by

$$\left(\frac{d}{p}\right) = \begin{cases} 1 & \text{if } d \text{ is square modulo } p \\ -1 & \text{otherwise.} \end{cases}$$

The famous quadratic reciprocity law says that for two odd primes p and q

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

For almost all p, we will attach to the pair

$$(Q(\sqrt{d}), p) \mapsto \left(\frac{d}{p}\right).$$

If we now consider the field $\mathbb{Q}_p(\sqrt{d})$ in which $\mathbb{Q}_p$ is the completion of Q with respect to the ultra–metric

$$|xp^m|_p = p^{-m},$$

where p does not divide the rational number $x \in \mathbb{Q}$, then

$$\sigma_p(\sqrt{d}) = \left(\frac{d}{p}\right)\sqrt{d}$$

where σ_p is the extension of σ to $Q_p(\sqrt{d})$. The automorphism σ_p will act trivially if and only if $x^2 - d$ splits modulo p.

Then for almost all p, the symbol $\left(\frac{d}{p}\right)$ defines a character χ_p of Q_p^* by

$$\chi_p(xp^m) = \left(\frac{d}{p}\right)^m \qquad (p \nmid x).$$

The correspondence

$$\sigma_p \mapsto \chi_p$$

for all $p \nmid d$ is an example of Artin reciprocity map. It can be defined for all Galois extensions $L/\mathbb{Q}$ for which Gal$(L/\mathbb{Q})$ is abelian.

One of the major steps in Langlands program is to extend this reciprocity to any Galois extension. Matters will be more complicated. One will have to consider continuous representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and attach to them generalizations of objects already introduced as Maass forms in $L^2(\Gamma\backslash H)$ or modular forms of arbitrary weight.

While this has been established for function fields in many settings, the case of number fields is wide open.

Reference for reciprocity laws: Jared Weinstein, Bulletin of AMS, Vol. 53, Number 1, 2016.

Conclusion: Two problems are connected through spectral analysis of modular forms. Their generalizations, the space of automorphic forms, will then be the target for resolving these central problems in number theory. This is an important part of the Langlands Program.

Thank you!