# SQUARES - AN INVITATION TO MODERN NUMBER THEORY

EHUD DE SHALIT

**Expanded Notes for a talk delivered at the Asian Science Camp**

**August 2012**

*These notes contain many more details than I could present at the talk, and in particular are aimed at listeners with mathematical backgound.*

## 1. Sums of two squares

1.1. **Introduction.** Pierre de Fermat was a French amateur mathematician, perhaps the most serious amateur the subject has ever witnessed. From the town of Toulouse, in the High Court of which he served as a councillor, he corresponded with other prominent mathematicians of his time. In those days it was not yet common to publish one's results in a scholarly paper submitted to a Journal or The Academy. Rather, if you discovered a good theorem, you reported it to your rivals and challenged them to find the proof themselves, supplying hints but not much more. Fermat made astounding discoveries in many branches of mathematics, from number theory to optics. For some of these discoveries, such as his famous "last theorem", he surely did not have a proof. A handful of his statements were wrong. But mostly he was right, and, as historians believe today, had convincing arguments, if short of full proofs by modern standards.

In a letter to Mersenne from 1640 Fermat made an observation which amounts to the following theorem.

**Theorem 1.1.** *An odd prime $p$ is of the form $x^2 + y^2$ for integers $x$ and $y$ if and only if $p \equiv 1 \bmod 4$.*

**Remark 1.1.** *Clearly, $2 = 1^2 + 1^2$. Since a square leaves residue 0 or 1 modulo 4, a sum of two squares is never 3 modulo 4, so the necessity of the condition $p \equiv 1 \bmod 4$ is obvious. What is not obvious is that every prime of this form is indeed a sum of two squares.*

Fermat gave only a sketch of a proof in his letter. Leonhard Euler, a Swiss-born mathematician who spent most of his career in St. Petersburg, and is considered the most prolific mathematician of all times, learned about Fermat's discoveries in number theory from Goldbach in 1729. In 1732, when he was only 25, Euler published his first paper, refuting a conjecture of Fermat that every number of the form $2^{2^n} + 1$ is prime. Euler simply showed that 641 divided $2^{32} + 1$. The theorem on sums of two squares gave him a much greater headache, and he managed to give

a full proof only in 1749. We shall now describe Euler's proof in modern language. It consists of two essentially distinct steps.

1) **The *descent* step** (the terminology will become clear later on): If the congruence $x^2+y^2 \equiv 0 \bmod p$ has a non-trivial solution, then there exists a solution in integers to $x^2+y^2=p$.

*Non-trivial* means that at least one of $x$ or $y$ (hence in fact both) is not divisible by $p$.

2) **The *reciprocity* step**: If $p \equiv 1 \bmod 4$, then the congruence $x^2+y^2 \equiv 0 \bmod p$ has a non-trivial solution.

1.2. **Review of some algebra.** All commutative rings are assumed to have 1. Recall that an ideal in a commutative ring $R$ is a proper subset $I$ of $R$ closed under addition and subtraction (i.e. it is an additive subgroup) and under multiplication by every element of $R$. Ideals show up naturally in algebra because they are the kernels of homomorphisms between rings. If $f : R \to S$ is a homomorphism between commutative rings, $\ker(f)$ is an ideal. Conversely, if $I$ is an ideal, the quotient group $R/I$ can be given a ring structure ($(a+I)(b+I) = ab+I$) so that the canonical projection $R \to R/I$ becomes a ring homomorphism with kernel $I$. Since $1 \notin I$, an ideal is *never* a subring.

If $a \in R$ is a non-invertible element, the set $(a) = aR$ is an ideal. Such ideals are called principal.

If $I$ and $J$ are ideals, $IJ$ denotes the ideal consisting of all finite sums of products of an element from $I$ with an element from $J$. This defines an associative and commutative multiplication on the set of ideals, and the product $I_1 I_2 \cdots I_r$ is therefore defined unambiguously by induction on $r$. For example, the product of two principal ideals is again principal $(b)(c) = (bc)$.

A *domain* is a subring of a field. A commutative ring is a domain if and only if it has no zero divisors: the product of two nonzero elements is nonzero. A *principal ideal domain (PID)* is a domain in which every ideal is principal: of the form

$$I = (a)$$

for some (non-unique, in general) $a$, called a *generator* of $I$. The rings $\mathbb{Z}$ and $\mathbb{R}[X]$ are PID's. In both of them this follows from the existence of a Euclidean algorithm. Let us recall how it works in $\mathbb{Z}$, because we shall soon have the occasion to immitate it in a similar example. Given $a \neq 0$ and $b$ in $\mathbb{Z}$, the Euclidean algorithm tells us that there exist $q, r$ such that $|r| < |a|$ and

$$b = qa + r.$$

Note that unless $a|b$, there are two possible values of $r$. We can make $r$ unique if we insist, for example, that $-a/2 < r \leq a/2$, but this is not necessary for our purposes. Now given a non-zero ideal $I$ in $\mathbb{Z}$, pick $0 \neq a \in I$ such that $|a|$ is minimal. It is an easy exercise to show that $I = (a)$. Although PID's are nice, you should keep in mind that they are very special rings. The ring $\mathbb{R}[X, Y]$ is already not a PID. The ideal consisting of all the polynomials vanishing at the origin $(0, 0)$ is not principal (prove it!).

A unit in a commutative ring $R$ is an element $u$ for which there exists a multiplicative inverse: a $v \in R$ such that $uv = 1$. Such a $v$ is necessarily unique, so is denoted $u^{-1}$. The units in $R$ are denoted $R^\times$ and they form a multiplicative subgroup. For example, $\mathbb{Z}^\times = \{\pm 1\}$. An element $\pi$ in a *domain* $R$ is called *irreducible* if whenever $\pi = bc$, either $b$ or $c$ is a unit. It is called *prime* if whenever

$\pi|bc$ (i.e. $bc = \pi x$ is solvable in $R$), $\pi|b$ or $\pi|c$. Both properties are well-known characterizations of primes in $\mathbb{Z}$. In general domains they need not be equivalent properties, but it is an easy fact that in a PID, *prime* $\equiv$ *irreducible* (prove it!).

An ideal $P$ in a commutative ring $R$ is called *a prime ideal* if whenever $bc \in P$, $b \in P$ or $c \in P$. Show that this is equivalent to $R/P$ being a domain. The *element* $\pi$ is prime if and only if the *ideal* $(\pi)$ is prime. This follows at once from the dictionary $b \in (\pi) \Leftrightarrow \pi|b$. Thus for PID's there isn't much new in the definition, because we can always test the primeness of an ideal by checking the primeness of any generator of it. In rings which are not PID's however, the notion of a prime ideal is much more useful than the notion of a prime element. Keep this remark in mind. It will be crucial later on in the course.

PID's are *unique factorization domains* (UFD's): every $a \in R$ is a finite product of primes (irreducibles), and this product is unique up to a change of order and the multiplication of the prime factors by units. Stated explicitly, if $a = \pi_1 \dots \pi_r$ and $a = \pi'_1 \dots \pi'_s$ are two decompositions as products of primes, then $r = s$, and after a reordering of the indices $\pi'_i = u_i \pi_i$ for $u_i \in R^\times$. Note that already here the language of ideals is better suited: an equivalent formulation is that every ideal $I$ in $R$ is a finite product $I = P_1 P_2 \cdots P_r$ of primes, and the $P_i$ are unique up to order. In fact $(\pi) = (\pi')$ if and only if $\pi' = u\pi$ for a unit $u$. Thus if $I = (a)$, to the decomposition $a = \pi_1 \dots \pi_r$ as a product of prime elements, which is unique up to ordering and the replacement of $\pi_i$ by $\pi'_i = u_i \pi_i$ ($u_i \in R^\times$, and of course $u_1 \dots u_r = 1$), we attach the decomposition $I = P_1 P_2 \cdots P_r$ with $P_i = (\pi_i)$ which is unique up to ordering only! Once you become accustomed to the formulation of the UFD property in PID's in terms of ideals, it will not be surprising that when we lose the PID property, *unique factorization of ideals* becomes the right notion to look at!

1.3. **The descent step.** We look at the ring $\mathbb{Z}[i] = \{x + yi | \, x, y \in \mathbb{Z}\}$ called the ring of Gaussian integers. It has 4 units $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$, and may be depicted as the lattice $\mathbb{Z}^2$ in the Euclidean plane.

**Lemma 1.2.** *$\mathbb{Z}[i]$ is a PID, hence a UFD.*

*Proof.* As for $\mathbb{Z}$, this follows from the existence of a Euclidean algorithm: if $a \neq 0$, $b \in \mathbb{Z}[i]$ then there are $q, r \in \mathbb{Z}$ with $b = qa + r$ and $|r| < |a|$. (Prove this! How small can we make $|r|/|a|$?). If $I$ is a non-zero ideal in $\mathbb{Z}[i]$ and $0 \neq a \in I$ is an element with $|a|$ minimal (why does such an $a$ exist?), show that $I = (a)$. ∎

**Lemma 1.3.** *If $q$ is a rational prime (i.e. a prime in $\mathbb{Z}$) and $q = x^2 + y^2$, then $x + iy$ is a prime in $\mathbb{Z}[i]$.*

*Proof.* Let $\kappa = x + iy$. Assume that $\kappa = \lambda\mu$ in $\mathbb{Z}[i]$. Then $q = \kappa\bar{\kappa} = (\lambda\bar{\lambda})(\mu\bar{\mu})$ is a decomposition in $\mathbb{Z}$, hence either $\lambda\bar{\lambda} = 1$ and $\lambda$ is invertible or $\mu$ is invertible. ∎

**Lemma 1.4.** *If $N = a^2 + b^2$, $q = x^2 + y^2$ is prime ($a, b, x, y \in \mathbb{Z}$) and $q|N$, then there exist $a_1, b_1 \in \mathbb{Z}$ such that $N/q = a_1^2 + b_1^2$, and any rational prime $p$ which divides both $a_1$ and $b_1$, divided already $a$ and $b$.*

*Proof.* In $\mathbb{Z}[i]$,

$$x + yi | (a + bi)(a - bi).$$

Since we have just seen that $x+yi$ is prime, it must divide one of the factors, let's say

$$a+bi = (x+yi)(a_1+b_1 i).$$

Multiplying by the conjugates this gives $N = q(a_1^2+b_1^2)$. If $p$ divides both $a_1$ and $b_1$ then it clearly divided both $a$ and $b$. ∎

We can now conclude the descent step. Suppose, by way of contradiction, that the assertion was wrong, and consider the collection of all pairs $(p,N)$ where $p$ is a prime which is *not* a sum of two squares, and $N$ is an integer divisible by $p$ which may be written as $a^2+b^2$, for integers $a$ and $b$ not divisible by $p$. Since we assumed that our assertion was wrong, this collection is non-empty. Order the collection lexicographically: $(p,N) < (p',N')$ if $p<p'$ or $p=p'$ and $N<N'$. Pick a minimal $(p,N)$. We shall reach a contradiction.

If $|a| > p/2$, we may replace it with $a_1 \equiv a \bmod p$, $|a_1| < p/2$, let $N_1 = a_1^2+b^2$ and get a smaller pair $(p,N_1)$. Hence $|a| < p/2$, and similarly $|b| < p/2$, so $N < p^2/2$.

Clearly, $N$ is not $p$. If $q$ is any prime factor of $N/p$, then $q<p$. Pick such a $q$. If $q$ is a sum of two squares, then by the last lemma, we may replace the pair $(p,N)$ by the pair $(p,N/q)$, which is smaller.

If, on the other hand, $q$ is not a sum of two squares, then either the pair $(q,N)$ is in our collection and is smaller than $(p,N)$, or both $a$ and $b$ are divisible by $q$, in which case we replace $(p,N)$ by $(p,N/q^2)$.

In all cases we have managed to *descend* from a pair $(p,N)$ to a smaller pair. Since we have assumed our initial pair was minimal (alternatively, since such a descent can not go on indefinitely), we reach a contradiction. Hence our collection of pairs $(p,N)$ must have been empty to begin with.

1.4. **The reciprocity step.** This step was the more difficult one for Euler. In modern language, however, it is rather simple. Suppose $p-1=4k$. Consider the field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ and the equation

$$x^{4k}-1 = (x^{2k}-1)(x^{2k}+1) = 0$$

over it. By Fermat's little theorem it has $4k$ distinct solutions in $\mathbb{F}_p$, namely all the non-zero elements of the field. Nevertheless, at most $2k$ of them annihilate $x^{2k}-1$, so there must be a solution of $x^{2k}+1=0$. If $a$ is any integer representing $x$ then

$$a^2+1^2 \equiv 0 \bmod p$$

which is what we wanted to get, in fact with $b=1$.

Modern language allows to give a shorter argument for the whole proof, concealing the descent argument, as follows. If $p \equiv 1 \bmod 4$ then

$$\begin{aligned}\mathbb{Z}[i]/(p) &= \mathbb{Z}[X]/(p, X^2+1) \\ &= \mathbb{F}_p[X]/(X^2+1).\end{aligned}$$

As we have just seen in the reciprocity step, the polynomial $X^2+1$ splits in $\mathbb{F}_p[X]$ into a product of two relatively prime (linear) factors, so $\mathbb{Z}[i]/(p) = \mathbb{F}_p \times \mathbb{F}_p$ and $p$ is not prime in $\mathbb{Z}[i]$. If $\pi$ is a prime of $\mathbb{Z}[i]$ dividing $p$, then $\bar{\pi}|\bar{p}=p$ as well. But $\bar{\pi}$ and $\pi$ are distinct primes: they do not differ by a unit, or they would be $\pm 1 \pm i$, and $p$ would have to be 2 (check it!). By unique factorization, $\pi\bar{\pi}$ divides $p$ in $\mathbb{Z}[i]$. Being a rational integer, it must divide $p$ in $\mathbb{Z}$ (why?), hence must be equal to it: $p=\pi\bar{\pi}$. Writing $\pi = x+yi$ we get $p=x^2+y^2$. We have chosen to go through the

more elaborate proof outlined above both for historical reasons, and to emphasize the principle of descent, which plays a prominent role in many areas in number theory.

1.5. **Where Euler failed.** By the same method Euler proved the following two "theorems" of Fermat.

**Theorem 1.5.** *A prime* $p = x^2 + 2y^2$ *if and only if* $p = 2$ *or* $p \equiv 1, 3 \bmod 8$.

**Theorem 1.6.** *A prime* $p = x^2 + 3y^2$ *if and only if* $p = 3$ *or* $p \equiv 1 \bmod 3$.

For the first theorem work in the ring $\mathbb{Z}[\sqrt{-2}]$. For the second, there is a little twist. The ring $\mathbb{Z}[\sqrt{-3}]$ is not a PID. One has to work in the slightly larger ring $\mathbb{Z}[\frac{-1+\sqrt{-3}}{2}]$, and to bother about 2's in the denominators, but other than that, the arguments are the same.

We leave the equation $p = x^2 + 4y^2$ as an exercise. It follows at once from the study of $p = x^2 + y^2$ (why?).

Fermat also conjectured the following theorem.

**Theorem 1.7.** *A prime* $p = x^2 + 5y^2$ *if and only if* $p = 5$ *or* $p \equiv 1, 9 \bmod 20$.

When Euler tried to follow the same two-step path he rather quickly found that reciprocity step led to the following.

**Theorem 1.8.** *Non-trivial solutions to* $x^2+5y^2 \equiv 0 \bmod p$ *exist if and only if* $p = 5$ *or* $p \equiv 1, 3, 7, 9 \bmod 20$.

However, the descent step did not work any more. The ring $\mathbb{Z}[\sqrt{-5}]$ turned out not to be a PID, nor a UFD. Indeed,

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

and $2, 3, 1 \pm \sqrt{-5}$ are all irreducible elements, so factorization is not unique. Related problems are that the ideal $(2, 1+\sqrt{-5})$ is not principal, and that 3 is irreducible but not prime (because it does not divide $1 \pm \sqrt{-5}$ although it divides their product). And $x^2 + 5y^2 \equiv 0 \bmod 3$ has the solution $(1, 1)$, but the equation $x^2 + 5y^2 = 3$ is clearly insolvable.

Euler could not solve the puzzle, and turned his attention to generalizations of the reciprocity step that we shall consider in the next lecture. It was Lagrange, who in his memoire *Recherches d'Arithmétique* (1773-1775) developed the theory of quadratic forms to deal with the problem and showed that the quadratic form $x^2 + 5y^2$ had a companion one, $2x^2 + 2xy + 3y^2$ of the same discriminant $D = -20$ (and up to equivalence of quadratic forms, no other). The right generalization of the descent step led to the the following theorem.

**Theorem 1.9.** *If the congruence* $x^2 + 5y^2 \equiv 0 \bmod p$ *has a non-trivial solution then* either $p = x^2 + 5y^2$ or $p = 2x^2 + 2xy + 3y^2$. *The first case occurs when* $p \equiv 1, 9 \bmod 20$ *and the second when* $p \equiv 3, 7 \bmod 20$.

1.6. **Where Lagrange failed.** The next surprise came with $x^2 + 27y^2$. As far as the reciprocity step goes, this is no different than $x^2 + 3y^2$. In fact, since $-27 = (-3) \cdot 3^2$, asking whether $-27$ is a square modulo $p$ is the same as asking whether $-3$ is a square, and Euler (perhaps Fermat) already knew that this is the case if an only if $p \equiv 1 \bmod 3$ (assume $p \neq 3$).

The descent step led, as before, to a (unique, up to equivalence) companion quadratic form, and showed that if $p \equiv 1 \bmod 3$, *either* (i) $p = x^2 + 27y^2$ *or* (ii) $p = 4x^2 + 2xy + 7y^2$. Both quadratic forms have $D = -4 \cdot 27 = -108$. Examples of primes satisfying (i) are $p = 31, 43, 109$. Examples of primes satisfying (ii) are $7, 13, 19, 37, 61, 67, 73, 79, 97, 103$. However, it is impossible to distinguish between the primes satisfying (i) and the primes satisfying (ii) by a congruence on $p \bmod 108$, or modulo any other modulus! It took Gauss, the *prince of mathematics*, to prove the following conjecture of Euler in 1805.

**Theorem 1.10.** *A prime $p = x^2+27y^2$ if and only if $p \equiv 1 \bmod 3$ and $x^3 \equiv 2 \bmod p$ is solvable. A prime $p = 4x^2+2xy+7y^2$ if and only if $p \equiv 1 \bmod 3$ and $x^3 \equiv 2 \bmod p$ is non-solvable.*

Surprisingly, cubic equations over finite fields intervene!

The appearance of several non-equivalent quadratic forms of the same discriminant $D = -20$ or $D = -108$ is related to the failure of certain rings to be PID's. In both cases we have claimed that the two quadratic forms of discriminant $D$ were the only ones, up to equivalence. If you look up tables of class numbers (the number of inequivalent quadratic forms of a given discriminant is called the class number of the discriminant and denoted $h(D)$) you find indeed $h(-20) = 2$, but $h(-108) = 3$. This is because there are *two* notions: equivalence and *proper equivalence*, and the class number usually refers to the stricter notion of proper equivalence. The two quadratic forms $4x^2 \pm 2xy + 7y^2$ are equivalent (under the obvious change of variables $(x, y) \mapsto (-x, y)$) but not properly equivalent.

For a discussion of the Descent Step for the general quadratic form $x^2 + ny^2$, $n > 0$, we refer to D.Cox's beautiful book, *Primes of the form $x^2 + ny^2$*. In the next lecture we turn our attention, as Euler did, to generalizations of the Reciprocity Step.

## 2. Quadratic Reciprocity

2.1. **Gauss' law of quadratic reciprocity.** Let $p$ be an odd prime. Euler asked the question when does the congruence

$$x^2 + ny^2 \equiv 0 \bmod p$$

have a non-trivial solution. The question makes sense for any $n$, positive or negative, and clearly depends only on $n \bmod p$. However, Euler fixed $n$ and asked the question for a variable $p$. There is no a-priori reason why this should be determined by congruence relations on $p$. That this is indeed the case (as we have seen for $n = 1$) is a deep and wonderful theorem.

Interpreted as an equation in the field $\mathbb{F}_p$, Euler's question is the same as asking whether $-n$ is a square in $\mathbb{F}_p$.

For $(p, n) = 1$, Legendre defined the symbol

$$\left(\frac{-n}{p}\right)$$

to be 1 if $-n$ is a square modulo $p$ and $-1$ if it is not a square modulo $p$. If $p|n$ the symbol is defined to be 0. Note that in terms of Legendre's symbol, the number of solutions in $\mathbb{F}_p$ to the equation $x^2 = -n$ is $1 + (-n/p)$.

Euler's empirical discovery, based on the study of many numerical examples, was that for a fixed $n$, whether or not $(-n/p) = 1$ is determined by congruences on $p$ modulo $4|n|$. Here are some entries from his tables, as communicated to Goldbach.

| **n** | **condition for $(-\mathbf{n}/\mathbf{p}) = \mathbf{1}$** |
|---|---|
| 3 | $p \equiv 1, 7 \bmod 12$ ($\Leftrightarrow p \equiv 1 \bmod 3$) |
| 5 | $p \equiv 1, 3, 7, 9 \bmod 20$ |
| 7 | $p \equiv 1, 9, 11, 15, 23, 25 \bmod 28$ ($\Leftrightarrow p \equiv 1, 2, 4 \bmod 7$) |
| $-3$ | $p \equiv \pm 1 \bmod 12$ |
| $-5$ | $p \equiv \pm 1, \pm 9 \bmod 20$ ($\Leftrightarrow p \equiv 1, 4 \bmod 5$) |
| $-7$ | $p \equiv \pm 1, \pm 3, \pm 9 \bmod 28$ |

Note that in half the cases the modulus can be taken $|n|$.

Euler did not specify in his conjecture which congruences modulo $4|n|$ should appear. The reason for not being able to predict them was that he was interested in general $n$. Gauss' insight was that for $n$ an odd prime there is a very neat formula. That this formula suffices follows from the following lemma.

**Lemma 2.1.** *(i) The Legendre symbol is multiplicative in the numerator*

$$\left(\frac{nm}{p}\right) = \left(\frac{n}{p}\right)\left(\frac{m}{p}\right).$$

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

*(iii)* $(2/p) = (-1)^{(p^2-1)/8}$

*Proof.* (i) Consider the homomorphism $\phi : \mathbb{F}_p^\times \to \mathbb{F}_p^\times$, $\phi(x) = x^2$. Its kernel consists of $\{\pm 1\}$ so the image $H \subset \mathbb{F}_p^\times$ has $(p-1)/2$ elements, hence is a subgroup of index 2. We therefore have a canonical homomorphism $\mathbb{F}_p^\times \to \{\pm 1\}$ which is 1 on $H$ and $-1$ on its other coset. But $H$ is just the set of square residues modulo $p$, so this homomorphism is the Legendre symbol.

(ii) This is a restatement of the reciprocity step in Fermat's theorem on sums of two squares.

(iii) This is equivalent to $(-2/p) = 1$ if and only if $p \equiv 1, 3 \bmod 8$ which was mentioned above. ∎

**Theorem 2.2.** *(Gauss' law of quadratic reciprocity). Let $p$ and $q$ be two distinct odd primes. Then*

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

Using (i) an (ii) of the lemma, this is the same as $(q/p) = (p^*/q)$ where $p^* = (-1)^{(p-1)/2}p$.

Here is an example how Gauss' law leads to a quick computation of the Legendre symbol:

$\left(\frac{104}{163}\right) = \left(\frac{2^3.13}{163}\right) = \left(\frac{2}{163}\right)^3 \left(\frac{13}{163}\right) = (-1)^3(-1)\left(\frac{163}{13}\right) = \left(\frac{7}{13}\right) = -\left(\frac{13}{7}\right) = -\left(\frac{-1}{7}\right) = 1$.

Gauss gave several proofs of this celebratred theorem (there are many more known today). We shall give two proofs: one elementary, that sheds little light on its true meaning and possible generalizations. The second one uses some algebra and basic number theory, but is more transparent.

2.2. **The elementary proof.**

**Lemma 2.3.** *(Gauss) For every integer $a$ denote by $\{a\}_p$ (or just $\{a\}$ if $p$ is fixed) the unique integer congruent to $a$ modulo $p$ satisfying*

$$-p/2 < \{a\} < p/2.$$

*Let $a$ be relatively prime to $p$. Let $\mu$ be the number of elements in*

$$S = \{\{a\}, \{2a\}, \ldots, \{(p-1)a/2\}\}$$

*which are negative. Then*

$$\left(\frac{a}{p}\right) = (-1)^\mu.$$

*Proof.* In absolute values, $S$ covers $1, 2, \ldots, (p-1)/2$. Indeed, it contains $(p-1)/2$ numbers whose absolute values are all in the range $[1, (p-1)/2]$, and no two are opposite for if $ia \equiv -ja \bmod p$ $(1 \leq i, j \leq (p-1)/2)$ then $p$ divides $(i+j)a$, but $i + j \leq p - 1$. It follows that modulo $p$

$$\begin{aligned} \prod_{i=1}^{(p-1)/2} i &\equiv (-1)^\mu \prod (ai) \\ &\equiv (-1)^\mu a^{(p-1)/2} \prod i. \end{aligned}$$

Dividing the congruence by $\prod i$ (legitimate, since it is nonzero modulo $p$) we get that $(-1)^\mu$ is congruent modulo $p$ to $a^{(p-1)/2}$. Now consider the homomorphism

$$\psi : \mathbb{F}_p^\times \to \mathbb{F}_p^\times$$

given by $\psi(a) = a^{(p-1)/2}$. Its kernel contains at most $(p-1)/2$ elements, but on the other hand it contains all the squares (denoted $H$ above) by Fermat's little theorem, which are $(p-1)/2$ in number. It therefore contains half of the elements of $\mathbb{F}_p^\times$, hence the image is of order 2, namely $\{\pm 1\}$. Moreover, $\psi(a) = 1$ if and only if $a \in H$, if and only if $(a/p) = 1$. This shows that $(-1)^\mu = 1$ if and only if $(a/p) = 1$. Since both expressions are $\pm 1$, they are equal. ∎

We now conclude the proof. Let $\mu$ be the number of negative elements among $\{q\}_p, \ldots, \{(p-1)q/2\}_p$ and likewise $\nu$ the number of negative elements among $\{p\}_q, \ldots, \{(q-1)p/2\}_q$. We must show that

$$\mu + \nu \equiv \frac{p-1}{2}\frac{q-1}{2} \bmod 2.$$

Let

$$A = \left\{(i,j) | 1 \leq i \leq \frac{p-1}{2},\ 1 \leq j \leq \frac{q-1}{2}\right\}.$$

Decompose it into 4 parts (it is useful to draw them in the $(i,j)$-plane):

$$\begin{aligned} A_1 &= \{(i,j) \in A |\, pj - qi < -q/2\} \\ A_2 &= \{(i,j) \in A |\, -q/2 < pj - qi < 0\} \\ A_3 &= \{(i,j) \in A |\, 0 < pj - qi < p/2\} \\ A_4 &= \{(i,j) \in A |\, p/2 < pj - qi\}. \end{aligned}$$

Mapping $(i,j)$ to $(i',j') = (\frac{p+1}{2} - i, \frac{q+1}{2} - j)$ is a premutation of $A$ that maps $A_1$ bijectively onto $A_4$. Thus

$$\frac{p-1}{2}\frac{q-1}{2} = |A| \equiv |A_2| + |A_3| \bmod 2.$$

We claim that $|A_2| = \nu$ and $|A_3| = \mu$. Let us check the assertion for $A_2$, the other one being symmetrical. For a given $j \in [1,(q-1)/2]$, there is a unique $i \in \mathbb{Z}$ such that $-q/2 < pj - qi < q/2$, and then $pj - qi = \{pj\}_q$. If $(i,j) \in A_2$, we conclude that $\{pj\}_q$ is negative. If, on the other hand, $\{pj\}_q$ is negative, then from $-q/2 < pj - qi < 0$ we get that $i \in [1,(p-1)/2]$ and so $(i,j) \in A_2$.

2.3. **More background in algebra, and a second proof.** We shall assume knowledge of Galois theory. Let $\zeta = e^{2\pi i/p}$ and $F = \mathbb{Q}(\zeta)$. The field $F$ is a splitting field of the polynomial

$$f(X) = (X^p - 1)/(X - 1) = X^{p-1} + \cdots + X + 1$$

since all the roots of $f$, which are the $\zeta^k$ for $k = 1, \ldots, p-1$, are in $F$. We claim that $f$ is irreducible. The proof, due to Eisenstein, uses the change of variables

$$g(Y) = f(Y+1) = [(Y+1)^p - 1]/Y.$$

The coefficients of $g$ are the binomial coefficients $\binom{p}{n}$ and except for the leading coefficient, are all divisible by $p$ (since $p$ divides the numerator of $\binom{p}{n}$ but not the denominator, when $0 < n < p$). Moreover, the last coefficient of $g$ is just $p$. Now suppose we had a decomposition $g = g_1 g_2$ where $g_1$ and $g_2$ are in $\mathbb{Q}[X]$, of positive degree, and monic. Since $g$ is primitive, Gauss' lemma on polynomials implies that $g_i \in \mathbb{Z}[X]$. Reading the equation modulo $p$ gives $Y^{p-1} = \bar{g}_1 \bar{g}_2$ in $\mathbb{F}_p[X]$, which implies that all the non-leading coefficients of $g_i$ are also divisible by $p$. But then the free term of $g$ would be divisible by $p^2$, contradiction. This shows that $g$ is irreducible, hence $f$ (why?).

Now $F$ is obtained from $\mathbb{Q}$ by adjoining a single root of $f(X)$, so is of degree $p-1$, and as we have seen, is Galois over $\mathbb{Q}$.

Let $G = Gal(F/\mathbb{Q})$. By Galois theory, its order is $|G| = [F : \mathbb{Q}] = p - 1$. We are going to find its structure. There is a homomorphism (the cyclotomic character modulo $p$)

$$\chi : G \to (\mathbb{Z}/p\mathbb{Z})^\times$$

defined by assigning to $\sigma \in G$ the unique integer modulo $p$ such that

$$\sigma(\zeta) = \zeta^{\chi(\sigma)}.$$

Check that this is well defined, that $\chi(\sigma\tau) = \chi(\sigma)\chi(\tau)$ and that $\chi$ is injective. Since both domain and range have the same cardinality $p-1$, it is an isomorphism. If $a$ is an integer prime to $p$ we let $\sigma_a$ be the unique $\sigma$ for which $\chi(\sigma) = a \bmod p$. If $q$ is prime, $\sigma_q$ is called the *Frobenius* substitution attached to $q$.

A *number field* is a subfield of $\mathbb{C}$ which is finite dimensional over $\mathbb{Q}$. We have already seen that number theoretic questions phrased entirely in $\mathbb{Q}$ are better studied in number fields such as $\mathbb{Q}(i)$. Moreover, the subring $\mathbb{Z}[i]$ played a special role in the question of representing primes as sums of two squares. The subring $\mathbb{Z}[i]$ is a lattice in $\mathbb{Q}(i)$ - it is the $\mathbb{Z}$-span of a basis of $\mathbb{Q}(i)$ as a vector space over $\mathbb{Q}$. A lattice in a number field which is also a subring (contains 1 and is closed under multiplication) is called an *order*. It turns out that any number field $F$ has a unique maximal order

that contains any other order. It is called the *ring of algebraic integers in* $F$ and denoted $\mathcal{O}_F$.

The ring of algebraic integers in $\mathbb{Q}$ is just $\mathbb{Z}$ (prove it!). If $F = \mathbb{Q}(\zeta)$ as above, $\mathcal{O}_F = \mathbb{Z}[\zeta]$, and $1, \zeta, \ldots, \zeta^{p-2}$ is a basis for it over $\mathbb{Z}$. If $F = \mathbb{Q}(\sqrt{d})$ where $d \in \mathbb{Z}$ is square-free (positive or negative) then $\mathcal{O}_F = \mathbb{Z}[\sqrt{d}]$ if $d \equiv 2, 3 \bmod 4$ but $\mathcal{O}_F = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ if $d \equiv 1 \bmod 4$.

The ring $\mathcal{O}_F$ is of key importance. It is in general not a PID, nor a UFD (recall the example of $\mathbb{Z}[\sqrt{-5}]$). However, every ideal of $\mathcal{O}_F$ has a unique (up to ordering) decomposition into a product of prime ideals.

Going back to $F = \mathbb{Q}(\zeta)$, whose Galois group $G$ was identified with $(\mathbb{Z}/p\mathbb{Z})^\times$, we recall that the group $H$ of squares in $(\mathbb{Z}/p\mathbb{Z})^\times$ is of index 2. We let $K$ be the quadratic extension of $\mathbb{Q}$ which is fixed by $H$, $K = F^H$. We then have, for $a \in (\mathbb{Z}/p\mathbb{Z})^\times$

$$\sigma_a|_K = 1 \Leftrightarrow a \in H \Leftrightarrow (a/p) = 1.$$

**Lemma 2.4.** *$K = \mathbb{Q}(\sqrt{p^*})$ where $p^* = (-1)^{(p-1)/2}p$.*

*Proof.* Consider Gauss' sum

$$g = \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \zeta^a.$$

Clearly, $g \in F$. We now compute its square.

$$\begin{aligned} g^2 &= \sum_a \left(\frac{a}{p}\right) \zeta^a \sum_b \left(\frac{ab}{p}\right) \zeta^{ab} \\ &= \sum_b \left(\frac{b}{p}\right) \sum_a \zeta^{a(1+b)}. \end{aligned}$$

We seperate the term with $b = -1$, which is equal $(p-1)\left(-1/p\right)$. For any other $b$ the inner sum equals $-1$ (for if we added the term with $a = 0$ we would get 0). However, there are as many quadratic residues as there are non-residues, so

$$\sum_{b \neq -1} \left(\frac{b}{p}\right)(-1) = \left(\frac{-1}{p}\right).$$

Altogether, $g^2 = (-1/p)p = p^*$ so $\mathbb{Q}(\sqrt{p^*})$ is indeed a subfield of $F$. Moreover,

$$\begin{aligned} \sigma_b(g) &= \sum_{a=1}^{p-1} \left(\frac{a}{p}\right) \zeta^{ab} \\ &= \left(\frac{b}{p}\right) \sum_{a=1}^{p-1} \left(\frac{ab}{p}\right) \zeta^{ab} \\ &= \left(\frac{b}{p}\right) g, \end{aligned}$$

showing that $\sigma_b(g) = g$ if and only if $b \in H$. This proves the lemma. ∎

Consider the prime $q$, which was assumed to be distinct from $p$ and odd too. We know that $\sigma_q|_K = 1$ if and only if $(q/p) = 1$.

Recall that in any commutative ring $R$, $(x+y)^q \equiv x^q + y^q \bmod qR$ because all the binomial coefficients $\binom{q}{n}$ are divisible by $q$ if $0 < n < q$. From the fact that

$\sigma_q(\zeta) = \zeta^q$ and $n^q \equiv n \bmod q$ for every rational integer $n$ (Fermat's little theorem) we deduce that

$$\begin{aligned}\left(\sum_{k=1}^{p-2} n_k \zeta^k\right)^q &\equiv \sum_{k=1}^{p-2} n_k^q \zeta^{qk} \bmod q \\ &\equiv \sum_{k=1}^{p-2} n_k \zeta^{qk} = \sigma_q(\sum_{k=1}^{p-2} n_k \zeta^k).\end{aligned}$$

Thus, for $\alpha \in \mathcal{O}_F$,

$$\sigma_q(\alpha) \equiv \alpha^q \bmod q\mathcal{O}_F.$$

The same therefore holds in $\mathcal{O}_K$ modulo $q\mathcal{O}_K$ (this follows at once from $\mathcal{O}_K = \mathcal{O}_F \cap K$). But from the explicit description of $\mathcal{O}_K$ given above (and the fact that $q$ is odd) one sees that

$$\begin{aligned}\mathcal{O}_K/q\mathcal{O}_K &= \mathbb{Z}[\frac{1+\sqrt{p^*}}{2}]/(q) \\ &= \mathbb{Z}[\sqrt{p^*}]/(q) \\ &= \mathbb{Z}[X]/(X^2 - p^*, q) \\ &= \mathbb{F}_q[X]/(X^2 - p^*).\end{aligned}$$

If $(p^*/q) = -1$, the polynomial $X^2 - p^*$ is irreducible over $\mathbb{F}_q$, so $\mathcal{O}_K/q\mathcal{O}_K = \mathbb{F}_{q^2}$ is a field, and $q\mathcal{O}_K = Q$ remains prime. In this case raising to power $q$, hence $\sigma_q$, is a *non-trivial* automorphism of $\mathcal{O}_K/q\mathcal{O}_K$, so a-fortiori $\sigma_q|_K$ is non-trivial, and $(q/p) = -1$.

If on the other hand $(p^*/q) = 1$, the polynomial $X^2 - p^*$ is reducible, equal to $(X-\alpha)(X+\alpha)$ for some $\alpha \in \mathbb{F}_q$. This easily implies that there is a ring isomorphism

$$\mathcal{O}_K/q\mathcal{O}_K \simeq \mathbb{F}_q \times \mathbb{F}_q$$

and that $q\mathcal{O}_K = Q_1Q_2$ is a product of two prime ideals in $\mathcal{O}_K$. (Let $Q_i$ be the inverse images of $(X \pm \alpha) \bmod (X^2 - p^*)$ under the homomorphism

$$\mathcal{O}_K \to \mathcal{O}_K/q\mathcal{O}_K = \mathbb{F}_q[X]/(X^2 - p^*)).$$

Raising to power $q$ is then the *trivial* automorphism of $\mathcal{O}_K/q\mathcal{O}_K$. It follows that $\sigma_q$ induces the trivial automorphism on $\mathcal{O}_K/q\mathcal{O}_K$.

We contend that the non-trivial automorphism of $K$ must *exchange* the two prime ideals $Q_i$ which apear in the decomposition of $q\mathcal{O}_K$. In fact, the $Q_i$ are maximal: $Q_1 + Q_2 = \mathcal{O}_K$ (prove it!) so there exists an element $x \in Q_1$ which is congruent to 1 modulo $Q_2$. If $x'$ denotes the Galois conjugate of $x$, then $xx'$ is in $Q_1 \cap \mathbb{Z} = q\mathbb{Z}$, so it lies in $Q_2$ as well. Since $x \notin Q_2$, we must have $x' \in Q_2$, or $x \in Q_2'$, which implies that $Q_2' \neq Q_2$, so we must have $Q_2' = Q_1$.

Modulo $q$, the non-trivial automorphism of $K$ must therefore exchange the two copies of $\mathbb{F}_q$, and can not act trivially. Since we have seen that $\sigma_q|_K$ acts trivially on $\mathcal{O}_K$ modulo $q$, we must have $\sigma_q|_K = 1$, or, equivalently, $(q/p) = 1$. The proof of the reciprocity law is complete.

The fact that the way $q$ decomposes in $K$ as a product of prime ideal is determined by a congruence on $q \bmod p$ is another way to formulate of the law of quadratic reciprocity.

## 3. Generalizations

3.1. **Class Field Theory over $\mathbb{Q}$ and the reciprocity law.** By the end of the 19th century the algebraic theory of number fields and their rings of integers was well developed, thanks to work of Kummer, Dedekind, Kronecker, Weber and many others. It became clear that what mattered in the above discussion was not so much that we were dealing with *quadratic* extensions, but that $K$ was an *abelian* extension of $\mathbb{Q}$.

If $K$ is any finite Galois extension of $\mathbb{Q}$, there is a finite set $S$ of "bad" (technically speaking, *ramified*) primes such that any rational prime $q \notin S$ decomposes in $\mathcal{O}_K$ into a product of $g$ distinct prime ideals

$$q\mathcal{O}_K = Q_1 \dots Q_g$$

and the Galois group $G = Gal(K/\mathbb{Q})$ permutes the $Q_i's$ transitively. If furthermore $K$ is an *abelian* extension of $\mathbb{Q}$ (i.e. $G$ is abelian) there is a unique element $\sigma_q \in G$ which stabilizes all the $Q_i$ and induces $x \mapsto x^q$ on any $\mathcal{O}_K/Q_i$ (a finite field containing $\mathbb{F}_q$). This $\sigma_q$ is called the *Frobenius substitution (automorphism)* attached to $q$. If $G$ is not abelian, the automorphism $\sigma$ may depend on $Q_i$, a choice of a different $Q_i$ above the same $q$ leading to a conjugate of $\sigma$, so we can still talk of a well-defined *Frobenius conjugacy class* attached to $q$. Although this is important and useful, for our present discussion we must limit $G$ to be abelian.

A celebrated theorem of Kronecker and Weber, a far-reaching generalization of the reciprocity law, says that when $K/\mathbb{Q}$ is abelian, there exists a positive integer $m$ (called the *conductor* of $K$), divisible only by the primes in $S$, such that $\sigma_q$ depends only on the residue of $q$ modulo $m$. In fact, there is a surjective homomorphism (*the Artin reciprocity map*)

$$(\mathbb{Z}/m\mathbb{Z})^\times \to G$$

mapping $q \bmod m$ to $\sigma_q$. This is achieved by showing that $K$ can be embedded in $\mathbb{Q}(e^{2\pi i/m})$, as we explicitly did for $K = \mathbb{Q}(\sqrt{p^*})$ using a Gauss sum.

In the early 20th century Hilbert, Artin, Hasse, Furtwangler, Takagi and others generalized these statements to base fields other than $\mathbb{Q}$. Class Field Theory attempts to classify all abelian Galois extensions $K/E$ of a fixed number field $E$, and their Galois groups, in terms of data encoded in $E$ alone. One major difficulty encountered by the people who worked on the foundations of CFT was that except for $E = \mathbb{Q}$ or $E$ a quadratic imaginary extension of $\mathbb{Q}$, one did not have an explicit description of special fields like $F = \mathbb{Q}(e^{2\pi i/m})$, into which every abelian extension of $E$ could be embedded. There is a general existence theorem for such "ray class fields", and many of their properties are known, but one does not have - in general - a way of writing them down.

3.2. **Non-abelian class field theory.** For a long time Class Field Theory, the ultimate generalization of Gauss' reciprocity law, was thought to be a theory of abelian extensions. Relations between CFT and certain complex analytic functions $L(s)$, called $L$-series (which are beyond the scope of this introduction), lead Artin in the 1930's to make a conjecture about the analytic continuation of $L(s)$ that made sense for *any* Galois extension $K/\mathbb{Q}$ (or $K/E$), not necessarily abelian, and that followed from CFT in the abelian case.

From another perspective, class field theory, as well as Artin's conjectures in the non-abelian case, concerned the variation of the number of solutions of $f(X) \equiv$

$0 \bmod p$ with $p$, when $f$ was a fixed monic polynomial in $\mathbb{Z}[X]$. What happens if we go up in the dimension and consider a polynomial equation $f(X,Y) = 0$ in two variables (which describes an *algebraic curve*)? One can still look at the same equation modulo $p$ and count solutions $(x,y) \in \mathbb{F}_p^2$. The variation of the number of solutions with $p$ is a fascinating subject, that was taken up by Artin, Hasse and Weil in the first half of the 20th century.

Loosely speaking, it is nowadays believed that to any collection of $m$ polynomial equations in $n$ unknowns with coefficients from $\mathbb{Z}$ (technically, to any *motive*), there corresponds an "*automorphic*" object, belonging to another world, which tells us everything that we want to know about solutions of this system of equations modulo $p$, as $p$ varies.

In the simple-minded example where $f = X^2 - q^*$, $q$ a prime different from $p$, the automorphic object is the Legendre symbol $(\cdot/q)$ and the fact that the number of solutions to $f \equiv 0 \bmod p$ is $1 + (p/q)$ is the quadratic reciprocity law.

It requires a great deal of machinery even just to define the terms *motivic* or *automorphic*, let alone to describe the conjectured correspondence, which usually goes under the name of Langlands. Instead, we shall give two examples.

3.3. **A 0-dimensional non-abelian example.** We follow an example of van der Blij from 1952. Consider the equation

$$h(X) = X^3 - X - 1,$$

a cubic polynomial of discriminant $D = -23$. (The discriminant of a cubic polynomial with roots $\alpha_1, \alpha_2$ and $\alpha_3$ is the expression

$$D = (\alpha_1 - \alpha_2)^2(\alpha_1 - \alpha_3)^2(\alpha_2 - \alpha_3)^2$$

which can be also expressed as a certain polynomial in the coefficients of the polynomial. Note that $\sqrt{D}$ belongs to the splitting field of the polynomial.) Let $K$ be the splitting field of $h$. Since $h$ has a unique real root $\alpha$, $K$ is an $S_3$-extension of $\mathbb{Q}$ and $K = \mathbb{Q}(\alpha, \beta)$ where $\beta = \sqrt{-23}$. Let $F = \mathbb{Q}(\alpha)$ and $E = \mathbb{Q}(\beta)$. Then $[E : \mathbb{Q}] = 2$, $[F : \mathbb{Q}] = 3$ and $G = \langle \sigma, \tau \rangle$ where $\sigma$ is a generator of $Gal(K/E)$, while $\tau$ is complex conjugation, the non-trivial automorphism of $K/F$. We have $\tau^2 = 1$, $\sigma^3 = 1$ and $\tau\sigma\tau^{-1} = \sigma^2$.

Consider the homomorphism (representation)

$$\rho : G \to GL_2(\mathbb{C})$$

given explicitly by

$$\rho(\tau) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \rho(\sigma) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

(check that such a representation exists). Let $\chi_\rho = Tr\rho$ be the character of $\rho$. There are three conjugacy classes in $G$, represented by $1, \sigma$ and $\tau$, and we immediately find out that $\chi_\rho(1) = 2$, $\chi_\rho(\sigma) = -1$ and $\chi_\rho(\tau) = 0$.

The only bad (ramified) prime for $K/\mathbb{Q}$ is 23. It is easy to show that for $p \neq 23$, $N_p$, the number of solutions to $X^3 - X - 1 \equiv 0 \bmod p$, is given by

$$N_p = 1 + \chi_\rho(\sigma_p)$$

where $\sigma_p$ is any element in the Frobenius conjugacy class of $p$ (a conjugacy class in $G$). Indeed, $N_p = 3$ if $p$ splits completely in $K$, $N_p = 1$ if $p$ splits into a product of 3 primes in $K$, each of relative degree 2 (hence into a product of two primes, of

relative degrees 1 and 2, in $F$), and $N_p = 0$ if $p$ splits into a product of 2 primes in $K$, each of relative degree 3 (hence remains inert in $F$).

Compare with the number of solutions of $X^2 + 23 \equiv 0 \bmod p$, which is given by $1 + (-23/p)$. Quadratic reciprocity taught us that the "automorphic" object attached to this (abelian 0-dimensional) "motive" is the Legendre symbol $(p/23)$. What is the "automorphic" object attached to the representation $\rho$ (i.e. to the 0-dimensional, non-abelian motive given by $X^3 - X - 1$)?

Let $\varepsilon : (\mathbb{Z}/23\mathbb{Z})^\times \to \{\pm 1\}$ be the Legendre symbol $(\cdot/23)$. Observe that $\det \rho(\sigma_p) = \varepsilon(p)$, since $\det \rho$ is equal to the quadratic character of $E$. Consider the power series

$$\eta(z) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n)$$

where $q = e^{2\pi i z}$, which represents a holomorphic function in the upper half plane $\mathrm{Im}(z) > 0$. Let

$$\begin{aligned} f(z) &= \eta(z)\eta(23z) = q \prod_{n=1}^{\infty} (1 - q^n)(1 - q^{23n}) \\ &= \sum_{n=1}^{\infty} a_n q^n \ (a_1 = 1). \end{aligned}$$

Then $f \in S_1(\Gamma_0(23), \varepsilon)$ is a *weight-one cuspidal modular form of level* 23 *and nebentypus* $\varepsilon$ (in fact, it is the only such cusp-form). This means (i) that for any matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$$

with $c \equiv 0 \bmod 23$,

$$f(\frac{az+b}{cz+d}) = \varepsilon(d)(cz+d)f(z)$$

and (ii) that for every $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, $(cz+d)^{-1} f(\frac{az+b}{cz+d})$ tends to 0 as $z = iy$ and $y \to \infty$. The modular form $f$ *is* the automorphic object attached to $\rho$, as the following theorem shows.

**Theorem 3.1.** *For every $p \neq 23$, we have $\chi_\rho(\sigma_p) = a_p$.*

In what sense are the Legendre symbol and the cusp-form $f$ similar "automorphic" objects? To understand it, one has to study *automorphic representations* (whatever that means). Both the Legendre symbol and the cusp-form $f$ give rise to automorphic representations, the first to a representation of the group $GL_1$ (over $\mathbb{Q}$) and the second to a representation of the group $GL_2$.

On the other hand, the variation of the number of solutions of $h(X) \equiv 0 \bmod p$, where $h$ is a polynomial in one variable over $\mathbb{Z}$, is governed, as we have seen in the two examples, by a *Galois representation.* In our first example this Galois representation sent $\sigma \in Gal(K/\mathbb{Q})$ to $\pm 1$, depending on whether $\sigma$ was trivial or not. In our second example the representation was the one we denoted by $\rho$. The Langlands Correspondence is a largely conjectural correspondence between certain $n$-dimensional Galois representations on one side, and certain automorphic representations of the group $GL_n$ on the other side.

We have gone a long way from the original Reciprocity Law of Gauss, through Class Field Theory, to the non-abelian generalizations embodied in the Langlands Correspondence. It is sometimes hard to see the original theorems as special cases of these far-reaching generalizations. We shall end this brief introduction by exhibiting yet another example, this time in dimension 1.

3.4. **A 1-dimensional non-abelian example.** Consider the equation

$$Y^2 + Y \equiv X^3 - X^2 - 10X - 20 \bmod p$$

and let $N_p$ be the number of solutions in $\mathbb{F}_p^2$. Heuristic arguments make one believe that this time $N_p$ should not remain bounded, but rather be of the order of magnitude $p$, so it makes sense to study the deviation $p - N_p$. Here is some numerical data, that you can check for yourself

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $p =$ | 2 | 3 | 5 | 7 | 11 | 13 | 17 | 19 |
| $p - N_p =$ | $-2$ | $-1$ | 1 | $-2$ | 1 | 4 | $-2$ | 0 |

It is not an accident that $p - N_p$ is relatively small. Already in the 1920's Artin conjectured, and later Hasse proved, the estimate

$$|p - N_p| \leq 2\sqrt{p}.$$

The equation we have written down is an equation of an *elliptic curve*. Its complex points (the set of points in $\mathbb{C}^2$ satisfying the equation, compactified by adding one point "at infinity") looks like a torus ($\mathbb{C}$ modulo a lattice) and has an abelian group structure (the point at infinity serving as the neutral element). In fact the set of solutions in $\mathbb{F}_p$ inherits a similar group structure, as long as $p \neq 11$. What we are counting is the number of points on this elliptic curve over $\mathbb{F}_p$, as $p$ varies.

The Shimura-Taniyama conjecture (proved by Taylor and Wiles) suggests the following. Look at ($q = e^{2\pi i z}$)

$$\begin{aligned} f &= \eta(z)^2\eta(11z)^2 \\ &= q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 \\ &= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 + \cdots \\ &= \sum_{n=1}^{\infty} a_n q^n. \end{aligned}$$

**Theorem 3.2.** *For every $p \neq 11$ we have $p - N_p = a_p$.*

Once again, $f \in S_2(\Gamma_0(11))$ is a *weight-two cuspidal modular form of level* 11 *and trivial nebentypus,* i.e. $f$ satisfies (i)

$$f(\frac{az+b}{cz+d}) = (cz+d)^2 f(z)$$

whenever $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$ and $c \equiv 0 \bmod 11$, and (ii) for every $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, $(cz+d)^{-2} f(\frac{az+b}{cz+d})$ tends to 0 as $z = iy$ and $y \to \infty$.

The relation between the elliptic curve$E : Y^2 + Y = X^3 - X^2 - 10X - 20$ and the cusp-form $f$ is a very deep one. In one direction, it tells us that the $L$-series $L_E(s)$ has analytic continuation to the whole complex plane, something that is impossible

to prove by any other method. In the other direction it allows us to use Hasse's estimate on $p - N_p$ to prove Ramanujan's conjecture that $a_n = o(n^{1/2+\varepsilon})$ for every $\varepsilon > 0$.