Fermat's Last Theorem for Regular Primes

A foray into algebraic number theory

Sayan Das,
BSc 2nd year, Jadavpur University

17 May, 2025

1. FLT for primes p: $p \nmid h\left(\mathbb{Q}(\zeta_p)\right)$

1.1. Introduction

One of the oldest parts of number theory is the study of *Diophantine equations* - polynomial equations $P(X_1, \ldots, X_n) = 0$ with coefficients in $\mathbb{Z}$[1] (resp. $\mathbb{Q}$) to which one seeks solutions $(x_1, \ldots, x_n)$ with $x_j \in \mathbb{Z}$ (resp. $\mathbb{Q}$) for $1 \leqslant j \leqslant n$.

Perhaps the most natural example of a Diophantine equation is $x^2 + y^2 = z^2$, which, geometrically, corresponds to right triangles of integer side lengths by Pythagoras' theorem. The solutions to this equation are called Pythagorean triples.

$(0, 0, 0)$ is a trivial solution, which we do not consider as, classically, a side length of 0 is nonsensical. So when we speak of solutions to a Diophantine equation, we really mean *nontrivial* solutions. Some examples of (nontrivial) Pythagorean triples are $(3, 4, 5)$ and $(5, 12, 13)$. To find the general solution we further assume $\gcd(a, b, c) = 1$; else if $\gcd(a, b, c) = d > 1$ then we can simply divide by d to get the solution $\gcd(a/d, b/d, c/d) = 1$. Also $x, y \equiv 1 \pmod 2 \implies z^2 \equiv 2 \pmod 4$ which is absurd. This means x and y cannot have the same parity, so we assume x is even whilst y, and consequently z, is odd.

1.1.1. Theorem. — *The general solution of the Diophantine equation $x^2 + y^2 = z^2$ is*

$$\left(2ab, a^2 - b^2, a^2 + b^2\right), \quad \text{where } a > b > 0 \text{ are of opposite parity.}$$

Proof. Rewrite $x^2 + y^2 = z^2$ as $x^2 = z^2 - y^2 = (z + y)(z - y)$. Now $x, z + y, z - y$ are even so we can write $z = 2u, z + y = 2v, z - y = 2w$. Then $u^2 = vw$. Also if d is a common divisor of v and w then d divides $v + w = z$ and $v - w = y$. But $\gcd(y, z) = 1$, so $d = 1$; hence v and w are coprime. So $u^2 = vw \implies v$ and w are both squares, say $v = a^2, w = b^2$. So we get

$$(x, y, z) = \left(2ab, a^2 - b^2, a^2 + b^2\right)$$

where $a > b > 0$ and $a + b \equiv a^2 + b^2 = z \equiv 1 \pmod 2$ so a and b are of opposite parity. □

A natural follow-up question is: are there solutions to $x^3 + y^3 = z^3$, and what about $x^4 + y^4 = z^4$, and $x^n + y^n = z^n$ for $n \geqslant 3$? In fact, the answer is no. This was first conjectured by Pierre de Fermat in 1637, in the margins of his copy of Diophantus' *Arithmetica*:

1.1.2. Theorem (Fermat's Last Theorem)**.** —

$$x^n + y^n = z^n \text{ has no nontrivial solutions for } n \geqslant 3.$$

The following theorem includes the case $n = 4$, by taking $x = a, y = b, z = c^2$.

1.1.3. Theorem. — $x^4 + y^4 = z^2$ *has no nontrivial solutions.*

[1] from the German *Zahlen*, "Numbers"

Fermat's proof. Suppose there is a solution (x, y, z) in integers. Then $(x^2)^2 + (y^2)^2 = (z)^2$ so (x^2, y^2, z) forms a Pythagorean triple. From 1.1.1 we get

$$x^2 = 2rs, y^2 = r^2 - s^2, z = r^2 + s^2$$

but $y^2 + s^2 = r^2$ so (y, s, r) is another Pythagorean triple. Then again

$$s = 2ab, y = a^2 - b^2, r = a^2 + b^2$$

so $x^2 = 2rs = 4ab(a^2 + b^2)$. Also $\gcd(ab, a^2 + b^2) = 1$ as $\gcd(a, b) = 1$ so ab and $a^2 + b^2$ must be squares. Let $a^2 + b^2 = Z^2$, then a and b must be squares as well so let $a = X^2, b = Y^2$. We then have $X^4 + Y^4 = Z^2$. Now observe that $0 < Z \leqslant Z^2 = r < z$, i.e. we started with a solution $x^4 + y^4 = z^2$ and then found another solution $X^4 + Y^4 = Z^2$ with $0 < Z < z$. We can repeat this process *ad infinitum*, and get an infinite decreasing sequence of positive integers as solutions to the given equation. But that contradicts the well-ordering principle - any subset of the positive integers must have a minimum. So we have reached an absurdity. □

The method above employed by Fermat is known as *the method of infinite descent*. A similar method can be used to prove 1.1.2 for $n = 3$; however, for general n this method proves insufficient. The general case was famously settled after 350 years by Wiles [Wil95], and Wiles and Taylor [WT95] using rather sophisticated techniques such as elliptic curves and modular forms. Much before Wiles' proof for the general case, in 1850 Kummer had proven 1.1.2 for the case of $n = p$ where p is a special kind of a prime, called a *regular prime*. We aim to give a motivated exposition of his proof, following closely the first chapter of [Was97] although in a more approachable way hopefully. We will be focusing on the case $\gcd(xyz, p) = 1$; otherwise the proof is much more difficult (see [Was97, Chapter 9]).

1.2. Failure of unique factorisation in number rings

In the 19th century many tried to prove 1.1.2 by trying to factorise the general equation $x^n + y^n = z^n$ as in the case of small exponents. First note that to prove 1.1.2 for any $n \geqslant 3$ it is enough to prove 1.1.2 for $n = p$ prime. This is because if 1.1.2 holds for some $m \geqslant 3$ then it holds for all positive multiples km of m. As in 1.1.1 we may assume $\gcd(x, y, z) = 1$. At this stage it is prudent to introduce some algebraic machinery.

1.2.1. Definition (Algebraic preliminaries)**.** — A *group* is a set G equipped with an associative binary operation $\star : G \times G \to G$ such that there exists an identity element $e_G \in G$, i.e. $a \star e_G = e_G \star a = a$ for any $a \in G$, and for every $a \in G$ there exists some inverse $b \in G$ such that $a \star b = b \star a = e_G$. If the operation is commutative, then G is said to be abelian. If we do not require the existence of inverses, G is said to be a *monoid*. $(\mathbb{Z}, +), (\mathbb{Z}/m\mathbb{Z}, +)$ are groups but $(\mathbb{Z}, \cdot), (\mathbb{Z}/m\mathbb{Z}, \cdot)$ are just monoids, where $\mathbb{Z}/m\mathbb{Z}$ are the integers modulo m. A *homomorphism* between two groups $(G, \star)$ and $(H, *)$ is a structure-preserving map $\eta : G \to H$ such that

$$\eta(a \star b) = \eta(a) * \eta(b), \ \eta(e_G) = e_H.$$

If η is bijective then η is called an *isomorphism*, and we write $G \cong H$; in this case there is no distinction between G and H except the names of the elements. For example, $(\mathbb{R}, +) \cong (\mathbb{R}^+, \cdot)$ taking $\eta(x) = e^x$. If $G = H$ then the isomorphism is called an *automorphism*.

A *ring* A is[2] a set equipped with both addition and multiplication $+, \cdot : A \times A \to A$ such that $(A, +)$ is an abelian group, $(A, \cdot)$ is a monoid and multiplication distributes over addition: $a(b + c) = ab + ac$, $(b + c)a = ba + ca$. The multiplication may not be commutative, but in number theory the most common rings we deal with are (or are based on) the commutative rings $\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z}$, so we follow the convention of assuming all rings to be commutative.

Although $(A, \cdot)$ is a monoid i.e. multiplicative inverses need not exist, sometimes there is a subset $A^\times$ of nonzero elements in A that are invertible under multiplication, called *units*. In $\mathbb{Z}$, the units are -1 and 1. The units of A forms an abelian group $A^\times$ under multiplication called the *group of units* of A. So $\mathbb{Z}^\times = \{-1, 1\}$. If $a = ub$ for some unit $u \in A$ then a and b are *associates* in A. In $\mathbb{Z}$, 5 and -5 are associates.

We generalise the notion of primes from $\mathbb{Z}$ to a ring A as follows.

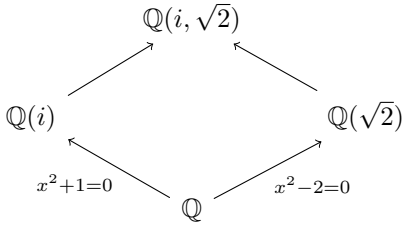
For any nonzero non-unit element $p \in A$, if $p = ab \implies a$ or b is a unit, then p is *irreducible* in A; if $p \mid ab \implies p \mid a$ or $p \mid b$ then p is *prime* in A.

A *field* is a ring with $1 \neq 0$ in which every nonzero element is a unit. The rationals $\mathbb{Q}$, reals $\mathbb{R}$ and complex numbers $\mathbb{C}$ are the most common examples of (infinite) fields. For any prime p, $(\mathbb{Z}/p\mathbb{Z})$ is an example of a finite field, denoted $\mathbb{F}_p$.

An *integral domain* is a ring where for any $a, b \in A$ we have $ab = 0 \implies a = 0$ or $b = 0$. Every finite integral domain is a field. The *characteristic* of a field is the smallest positive integer $n > 0$ such that $n \cdot 1 = 0$.

An *embedding* of a field E in F is a map $\sigma : E \to F$ such that the mapping restricted to each subgroup $\sigma : (E, +) \to (F, +)$ and $\sigma : (E^\times, \cdot) \to (F^\times, \cdot)$ is a group homomorphism. If $E = F$ and σ is bijective then it is called an *automorphism*.

1.2.2. Field extensions. — Often, when solving Diophantine equations, we encounter equations that don't have rational or even real solutions. For example $x^2 - 2 = 0$ has no rational solutions and $x^2 + 1 = 0$ has no real solutions. It is often useful to consider some extension of $\mathbb{Q}$ containing the roots of these equations in the intermediate step even if we are interested in solutions in $\mathbb{Q}$. In this way we construct the field $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ corresponding to $x^2 - 2 = 0$ and the field of *Gaussian numbers* $\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\}$ corresponding to $x^2 + 1 = 0$, containing all the roots of these respective equations. We can even construct $\mathbb{Q}(i, \sqrt{2})$ containing all the roots of both $x^2 - 2 = 0$ and $x^2 + 1 = 0$.



Fields constructed by adjoining (finitely many) roots of a polynomial equation in $\mathbb{Q}$ to $\mathbb{Q}$ are called (algebraic) number fields - these are examples of (finite) field extensions. If α is the root of some polynomial equation $P(x)$ with coefficients in $\mathbb{Q}$ (we call such an α algebraic) then the smallest field containing all the roots of $P(x)$ is $\mathbb{Q}(\alpha)$ defined as

$$\mathbb{Q}(\alpha) = \left\{ \sum_{n=0}^{n-1} a_k \alpha^k : a_k \in \mathbb{Q} \right\}.$$

[2]from the French *Anneau*, "Ring"

We will now formalise the preceding notions in the following definitions.

1.2.3. Definition. — Let K be a field[3]. A field $L \supseteq K$ is called an *extension* of K, and the extension is denoted by L/K ("L over K"). In a field extension L/K, L forms a K-vector space with $[L : K] = \dim_L(K)$ called the *degree* of the extension. A *finite field extension* is a field extension of finite degree.

Suppose $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for L/K, and $\alpha \in L$, then $\alpha\alpha_i = \sum_j a_{ij}\alpha_j$, with $a_{ij} \in K$. The *norm* and *trace* of α are defined, respectively, as

$$\mathrm{N}_{L/K}(\alpha) = \det(a_{ij}), \quad \mathrm{Tr}_{L/K}(\alpha) = \mathrm{tr}(a_{ij}).$$

An element $\alpha \in L/K$ is *algebraic* over K if α is the root of some polynomial with coefficients in K. If $[L : K]$ is finite then every $\alpha \in L/K$ is algebraic.

1.2.4. Definition. — A (algebraic) *number field* K is a finite field extension $K/\mathbb{Q}$ of $\mathbb{Q}$.

1.2.5. Definition (Galois group)**.** — Let L/K be a finite extension with $[L : K] = n$. The set of automorphisms $\sigma : L \to L$ with $\sigma(x) = x$ for all $x \in K$ forms a group under function compositon, called the *Galois group* $\mathrm{Gal}(L/K)$. For example, $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ consists of two elements: the identity automorphism and complex conjugation automorphism. Informally, the Galois group is a measure of how much the elements in the extension are indistinguishable from each other.

A rational root of a monic polynomial with integer coefficients is an integer; assuming p/q to be a rational root with $\gcd(a, b) = 1$ we get

$$0 = \frac{p^n}{q^n} + a_{n-1}\frac{p^{n-1}}{q^{n-1}} + \cdots + a_1\frac{p}{q} + a_0$$

from which, multiplying by q^n, we get $0 = p^n + a_{n-1}p^{n-1}q + \cdots + a_1pq^{n-1} + a_0q^n$. This implies that, for $n \geqslant 1$, $-p^n = q(\cdots) \implies q \mid p^n \implies q \mid p \implies q = \pm 1$. Hence an integer. We use this to generalise the notion of "integers" to an algebraic number field.

1.2.6. Definition. — An algebraic number $\alpha \in K/\mathbb{Q}$ is an *algebraic integer* if it is the root of a monic polynomial with integer coefficients. The ring of integers of an algebraic number field K is called the *number ring* $\mathcal{O}_K$.

For example, the ring of *Gaussian integers* $\mathbb{Z}[i]$ is the number ring for the field of Gaussian numbers $\mathbb{Q}(i)$. Note that $-i, i, -1$ and 1 are the units in $\mathbb{Z}[i]$ and $[\mathbb{Q}(i) : \mathbb{Q}] = 2$. Something interesting happens when we look at the behaviour of primes in $\mathbb{Z}$ when going from $\mathbb{Z}$ to $\mathbb{Z}[i]$.

(1) Some primes, like 3, remain prime in $\mathbb{Z}[i]$. We call these *inert primes*.

(2) Some primes, like 5, factor as $5 = (2 + i)(2 - i)$. We call these *split primes*.

(3) Some primes, like 2, factor as $2 = -i(1 + i)^2$. We call these *ramified primes*.

We will come back to this later in §1.2.10.

A nice property of both $\mathbb{Z}$ and $\mathbb{Z}[i]$ is that both are *unique factorisation domains* (UFDs): this means an integral domain D where any nonzero element n that is not a unit can be written uniquely (upto units) as a finite product of irreducible elements $q_1 \ldots q_k \in D$ as

$$n = q_1 \cdots q_k \quad \text{upto units.}$$

In a UFD, every irreducible element is prime.

[3] from the German *Körper*, "Body"

1.2.7. The cyclotomic field and ideal numbers. — A special number field is the cyclotomic field $\mathbb{Q}(\zeta_n)$ which is a generalisation of the Gaussian number field $\mathbb{Q}(i)$. The units in $\mathbb{Q}(i)$ are all roots of unity: $i^4 = (-i)^2 = (-1)^2 = 1$. We define the complex number ζ_n to be a primitive nth root of unity $\zeta_n = e^{2i\pi/n}$, i.e., $\zeta_n^n = 1$ and $\zeta_n^k \neq 1$ for $1 \leqslant k \leqslant n-1$. We've already seen $\zeta_4 = i$. The *nth cyclotomic field* is obtained by adjoining ζ_n to $\mathbb{Q}$,

$$\mathbb{Q}(\zeta_n) = \left\{ \sum_{k=0}^{n-1} a_k \zeta_n^k : a_k \in \mathbb{Q} \right\}.$$

Geometrically, $1, \zeta_n, \ldots, \zeta_n^{n-1}$ are points on the unit circle in the complex plane dividing it into exactly n parts. This is where the name *cyclotomic* originates[4]. The reason why we're interested in the cyclotomic fields is because we can use it to factorise Fermat's equation. This is because $1, \zeta_n, \ldots, \zeta_n^{n-1}$ are all distinct nth roots of unity, so they are all the roots of the polynomial $x^n - 1 = 0$, so $x^n - 1 = (x-1)(x-\zeta_n)\cdots(x-\zeta_n^{n-1})$. So for prime p,[5]

$$x^p + y^p = (x+y)(x+\zeta_p y)\cdots(x+\zeta_p^{p-1} y) = z^p. \tag{1.2.1}$$

This naturally leads us to consider the number ring $\mathbb{Z}[\zeta_p]$. In 1847, Lamé announced a proof of 1.1.2 for any odd prime p, by using the factorisation in (1.2.1). His idea was to show that each of the factors $(x + \zeta_p^k y)$ on the right are pairwise coprime, whence once concludes that each of these factors is itself a pth power, and one tries to reach an absurdity by infinite descent in the same spirit as 1.1.3's proof. This approach, however, is fundamentally flawed as unique factorisation does not hold in $\mathbb{Z}[\zeta_p]$, so the factorisation in (1.2.1) may not be unique ! In 1847 Kummer showed that the first p for which $\mathbb{Z}[\zeta_p]$ fails to be UFD is $p = 23$. To see this, first note that 2 is irreducible in $\mathbb{Z}[\zeta]$ (setting $\zeta = \zeta_{23}$). Also,

$$\begin{aligned} &2 \nmid (1 + \zeta^2 + \zeta^4 + \zeta^5 + \zeta^6 + \zeta^{10} + \zeta^{11}) = a \\ &2 \nmid (1 + \zeta + \zeta^5 + \zeta^6 + \zeta^7 + \zeta^9 + \zeta^{11}) = b \\ \text{but } &2 \mid ab = 2\zeta^{17} + 2\zeta^{16} + 2\zeta^{15} + 2\zeta^{13} + 2\zeta^{12} + 6\zeta^{11} + 2\zeta^{10} + 2\zeta^9 + 2\zeta^7 + 2\zeta^6 + 2\zeta^5. \end{aligned}$$

To calculate the product ab we brute force distribute and multiply, then use the fact that $1 + \zeta + \cdots + \zeta^{22} = 0$,

$$\begin{aligned} ab = 1 + \zeta + \zeta^2 + \zeta^3 + \zeta^4 + 3\zeta^5 + 3\zeta^6 + 3\zeta^7 + \zeta^8 + 3\zeta^9 + 3\zeta^{10} + 7\zeta^{11} + 3\zeta^{12} + 3\zeta^{13} + \zeta^{14} \\ +3\zeta^{15} + 3\zeta^{16} + 3\zeta^{17} + \zeta^{18} + +\zeta^{19} + \zeta^{20} + \zeta^{21} + \zeta^{22} \\ = 2\zeta^{17} + 2\zeta^{16} + 2\zeta^{15} + 2\zeta^{13} + 2\zeta^{12} + 6\zeta^{11} + 2\zeta^{10} + 2\zeta^9 + 2\zeta^7 + 2\zeta^6 + 2\zeta^5. \end{aligned}$$

As $2 \mid ab$ but $2 \nmid a$ and $2 \nmid b$, it is not prime yet irreducible so $\mathbb{Z}[\zeta_{23}]$ is not UFD. This failure of unique factorisation rendered the lines of attack by Lamé and many others essentially irrecoverable. Kummer was able to get some partial results, however, and his work led to the development of much of algebraic number theory.

First, Kummer explored some notion of "ideal numbers" in a number ring for which unique factorisation is restored. He first considered the basic divisibilty properties that would be required for an ideal number: if an ideal number divides a in some number ring $\mathcal{O}_K$ then it must also divide ab for every $b \in \mathcal{O}_K$; and if it divides $a, b \in \mathcal{O}_K$ it must also divide $a + b \in \mathcal{O}_K$. Kummer's ideas were formalised for any general ring by Dedekind and Noether.

[4] from the Greek *Cyclos*, "Circle" + *Tomos*, "Cutting"

[5] recall that to prove Fermat's Last Theorem for any $n \geqslant 3$ it is sufficient to prove it for any odd prime p

1.2.8. Definition (Ideal theory)**.** — An *ideal* $\mathfrak{a}$ of a ring A is a subset of A such that

- $(\mathfrak{a}, +)$ is a subgroup of $(A, +)$;
- if $x \in A$ and $y \in \mathfrak{a}$, then $xy \in \mathfrak{a}$.

The notation $x \equiv y \pmod{\mathfrak{a}}$ means $x - y \in \mathfrak{a}$. A *principal ideal* is an ideal generated by one element: $(x) = \{ax : a \in A\}$, for some fixed $x \in A$. Also $x \in A^\times \iff (x) = A = (1)$. An integral domain where every ideal is principal is called a *principal ideal domain* (PID).

An ideal $\mathfrak{p}$ in A is *prime* if $\mathfrak{p} \neq (1)$ and $xy \in \mathfrak{p} \implies x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. An ideal $\mathfrak{m}$ in A is *maximal* if $\mathfrak{m} \neq (1)$ and there is no ideal $\mathfrak{a}$ such that $\mathfrak{m} \subset \mathfrak{a} \subset A$.

A ring A is *Noetherian* if every ascending chain $\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \mathfrak{a}_3 \subset \cdots$ of ideals in A terminates, i.e. there is an $N > 0$ such that $\mathfrak{a}_m = \mathfrak{a}_{m+1}$ for all $m \geqslant N$.

Let A_1, A_2 be rings such that $A_2 \supseteq A_1$. Then $\alpha \in A_2$ is said to be *integral over* A_1 if there exists a monic polynomial $f(X)$ with coefficients in A_1 such that $f(\alpha) = 0$. *Integral closure* of A_1 in A_2 is the set of all $\alpha \in A_2$ that are integral over A_1. A_1 is *integrally closed over* A_2 if it equals its integral closure in A_2.

A *Dedekind domain* is a Noetherian, integrally closed integral domain in which every nonzero prime ideal is maximal.

The ring of integers $\mathcal{O}_K$ of a number field K is an example of a Dedekind domain. A Dedekind domain is not a UFD in general, but the following theorem is almost as good.

1.2.9. Theorem. — *[Neu92, Theorem 3.3] Every ideal* $\mathfrak{a}$ *of a Dedekind domain* D *different from* (0) *and* (1) *admits a factorisation*

$$\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_k \text{ upto units}$$

into nonzero prime ideals $\mathfrak{p}_i$ *of* D *which is unique upto the order of the factors.*

This theorem is certainly perfectly in line with Kummer's notion of "ideal numbers." For example, consider the simpler number ring $\mathbb{Z}[\sqrt{-5}]$. It is not UFD as

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

But if we consider the ideals we have

$$\begin{aligned}
(2) &= (2, 1 + \sqrt{-5})(2, 1 - \sqrt{-5}) = \mathfrak{p}_1\mathfrak{p}_2 \\
(3) &= (3, 1 + \sqrt{-5})(3, 1 - \sqrt{-5}) = \mathfrak{p}_3\mathfrak{p}_4 \\
(1 + \sqrt{-5}) &= (2, 1 + \sqrt{-5})(3, 1 + \sqrt{-5}) = \mathfrak{p}_1\mathfrak{p}_3 \\
(1 - \sqrt{-5}) &= (2, 1 - \sqrt{-5})(3, 1 - \sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_4
\end{aligned}$$

where $\mathfrak{p}_i$ are prime ideals and thus the ideal factorisation

$$(6) = (2, 1 + \sqrt{-5})(2, 1 - \sqrt{-5})(3, 1 + \sqrt{-5})(3, 1 - \sqrt{-5}) = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_4$$

is unique. However note that none of the $\mathfrak{p}_i$'s are principal which means $\mathbb{Z}[\sqrt{-5}]$ is not PID. We want to measure the deviation of a number ring from being a PID, which leads us towards the notion of an *ideal class group*.

1.2.10. Definition (Ideal class group)**.** — Two nonzero ideals $\mathfrak{a}, \mathfrak{b} \subset \mathcal{O}_K$ are said to be equivalent, $A \sim B$, if there exist nonzero $\alpha, \beta \in \mathcal{O}_K$ such that $(\alpha)\mathfrak{a} = (\beta)\mathfrak{b}$. This is an equivalence relation and the equivalence classes are called ideal classes. The set of these ideal classes form a finite abelian group called the *ideal class group* of K, denoted C_K. The number of ideal classes is called the *class number* $h(K) = |C_K|$.

Note that $h(K) = 1 \iff \mathcal{O}_K$ is PID. So the class number, in some sense, measures the deviation of $\mathcal{O}_K$ from being a PID.

1.2.11. Ramifications. — Notice that in 1.2.9 we did not assume each of the $\mathfrak{p}_i$'s to be distinct; grouping together the same prime ideals yields something that resembles the Fundamental Theorem of Arithmetic in $\mathbb{Z}$ generalised to ideals in a Dedekind domain D,

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}, \quad e_i > 0. \tag{1.2.2}$$

In (1.2.2) the $\mathfrak{p}_i's$ are now distinct. Now suppose $\mathfrak{a} = (p)$ for some prime $p \in \mathbb{Z}$ and $D = \mathcal{O}_K$. Then $\mathbb{Z}/(p) = \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a finite field with p elements. For each $\mathfrak{p}_i$ in (1.2.2), $\mathcal{O}_K/\mathfrak{p}_i$ is a finite extension of $\mathbb{Z}/(p)$. We define $f_i = [\mathcal{O}_K/\mathfrak{p}_i : \mathbb{Z}/(p)]$ as the degree of this extension. We call each exponent e_i in (1.2.2) the *ramification index* of $\mathfrak{p}_i$ over p and we call f_i the *inertial degree* of $\mathfrak{p}_i$ over p. We then say, in $\mathcal{O}_K$

- p is *ramified* if there is some $i = j : e_j > 1$; if $f_j = r = 1$, then p is *totally ramified*.
- p is *totally split* if $e_i = f_i = 1$ for all $1 \leqslant i \leqslant r$ and $r = [K : \mathbb{Q}]$.
- p is *inert* if it does not split at all.

Coming back to $K = \mathbb{Q}(i)$, we see that $(2) = (1+i)^2$ so 2 is totally ramified in $\mathbb{Z}[i]$. On the other hand $(5) = (2+i)(2-i)$ so it is totally split in $\mathbb{Z}[i]$. Lastly, $(3) = \mathfrak{p}$ does not split so it is inert in $\mathbb{Z}[i]$.

1.3. Kummer's approach using regular primes

1.3.1. Definition. — An odd prime p is *regular* if $p \nmid h_p = h\left(\mathbb{Q}(\zeta_p)\right)$.

1.3.2. Theorem (Fermat's Last Theorem for regular primes)**.** — *Let p be a regular prime. Then the Diophantine equation $x^p + y^p = z^p$ has no (nontrivial) solutions for $p \nmid xyz$.*

The key idea behind Kummer's proof is the *regularity assumption*: $p \nmid h_p$. The previous proofs by Lamé and others failed due to $\mathbb{Z}[\zeta_p]$ not being UFD (for $p = 23$ for example) but even considering ideal factorisation as Kummer did, we have

$$(z)^p = \prod_{k=0}^{p-1} \left(x + \zeta_p^k y\right)$$

where the ideal factors $(x + \zeta_p^k y)$ are coprime so each must be a pth power, so $(x + \zeta_p y) = \mathfrak{a}^p$ for some ideal $\mathfrak{a}$, so $\mathfrak{a}^p$ is trivial in the class group. But we can only conclude that $\mathfrak{a}$ itself is trivial in the class group if $p \nmid h_p$. This is why Kummer defined a regular prime the way he did.

The next natural question is: how do we determine whether a prime p is regular i.e. whether p divides $h_p = h\left(\mathbb{Q}(\zeta_p)\right)$ or not ? Kummer gave a nice criterion for this.

1.3.3. Definition. — The *Bernoulli numbers* B_n are defined as the coefficients in the expansion

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}.$$

In particular, $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_3 = 0$ and in fact $B_{2k+1} = 0$ for all $k \geqslant 1$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}$, $B_8 = -\frac{1}{30}$, $B_{10} = \frac{5}{66}$, $B_{12} = -\frac{691}{2730}$ etc. (cf. [Was97, p. 6]).

1.3.4. Theorem (Kummer's criterion)**.** — *A prime p is regular if and only if p does not divide the numerator of any of the Bernoulli numbers B_k for $k = 2, 4, 6, \ldots, p-3$.*

For example, the prime $p = 691$ is irregular as 691 divides the numerator of B_{12}. The first few irregular primes are $37, 59, 67, 101, 103, 131, 149$, and 157. In fact, it is known that there are infinitely many irregular primes while it is not known whether there are infinitely many regular primes. However, Siegel conjectured that approximately $1 - e^{-1/2} \approx 39\%$ of primes are irregular and $e^{-1/2} \approx 61\%$ are regular. So Kummer's proof is still a significant partial result.

Throughout the remainder of this article we write $\zeta = \zeta_p$. We'll first need some basic results about $\mathbb{Z}[\zeta]$.

1.3.5. Proposition. — *$\mathbb{Z}[\zeta]$ is the ring of integers of $\mathbb{Q}(\zeta)$, thus $\mathbb{Z}[\zeta]$ is a Dedekind domain so that we have unique factorisation into prime ideals and so on.*

If $\mathcal{O}$ is the ring of integers of $\mathbb{Q}(\zeta)$, then it is easy to see that $\mathbb{Z}[\zeta] \subseteq \mathcal{O}$. Showing the reverse inclusion $\mathcal{O} \subseteq \mathbb{Z}[\zeta]$ is more involved; for brevity we skip it, but the full proof is present in (p. 2, [Was97]). Another result that we'll use without proving is the following:

1.3.6. Proposition. — *[Was97, Proposition 1.5] Let ε be a unit in $\mathbb{Z}[\zeta]$. Then there exist $\varepsilon_1 \in \mathbb{Q}(\zeta + \zeta^{-1})$ and $r \in \mathbb{Z}$ such that $\varepsilon = \zeta^r \varepsilon_1$.*

Proof of Fermat's Last Theorem for regular primes (Theorem 1.3.2). First, observe that for $p \leqslant 5$ the proof is trivial. For, if $3 \nmid xyz$ then $x^3 \equiv y^3 \equiv z^3 \equiv \pm 1 \pmod 9$. Then $x^3 + y^3 \equiv -2, 0, 2 \not\equiv \pm 1 \pmod 9$ so $x^3 + y^3 \neq z^3$. Similarly, if $5 \nmid xyz$ then $x^5 \equiv y^5 \equiv z^5 \equiv 1, 3 \pmod{25}$. Then $x^5 + y^5 \equiv 2, 4, 6 \pmod{25} \not\equiv 1, 3 \pmod{25}$ so $x^5 + y^5 \neq z^5$. But we must stop at $p = 7$, for

$$1^7 + 30^7 \equiv 31^7 \pmod{49}.$$

So we must treat the rest of the cases $p \geqslant 7$ separately. Note that if $x \equiv y \equiv -z \pmod p$ then $x^p + y^p \equiv -2z^p \equiv z^p \pmod p$ which is absurd since $p \nmid 3z^p$. Hence we may rewrite the equation if necessary, as $x^p + (-z)^p = (-y)^p$, to obtain $x \not\equiv y \pmod p$.

The ideals $(x + \zeta^i y)$ for $0 \leqslant i < p$ are pairwise coprime. For, if that weren't the case, we'd have a prime ideal $\mathfrak{p}$ such that $\mathfrak{p} \mid (x + \zeta^i y)$ and $\mathfrak{p} \mid (x + \zeta^j y)$ for some $i \neq j$. Then $\mathfrak{p} \mid (\zeta^i y - \zeta^j y) = (\text{unit})(1 - \zeta)y$ so either $\mathfrak{p} = (1 - \zeta)$ or $\mathfrak{p} \mid y$.

Similarly, $\mathfrak{p} \mid \zeta^j(x + \zeta^i y) - \zeta^i(x + \zeta^j y) = (\text{unit})(1 - \zeta)x$ so either $\mathfrak{p} = (1 - \zeta)$ or $\mathfrak{p} \mid x$. If $\mathfrak{p} \neq (1 - \zeta)$ then $\mathfrak{p} \mid x$ and $\mathfrak{p} \mid y$ which is absurd as $\gcd(x, y) = 1$. If $\mathfrak{p} = (1 - \zeta)$ then $x + y \equiv x + \zeta^i y \equiv 0 \mod \mathfrak{p}$. As $x + y \in \mathbb{Z}$, we have $x + y \equiv 0 \pmod p$ but

$$z^p = x^p + y^p \equiv x + y \equiv 0 \pmod p$$

so $p \mid z$ which is absurd.

Now let $\beta \in \mathbb{Z}[\zeta]$, write $\beta = b_0 + b_1\zeta + \cdots + b_{p-2}\zeta^{p-2}$. Then

$$\beta^p \equiv (b_0)^p + (b_1\zeta)^p + \cdots + (b_{p-2}\zeta^{p-2})^p = b_0^p + b_1^p + \cdots + b_{p-2}^p \pmod p$$

so every $\beta \in \mathbb{Z}[\zeta]$ raised to the pth power is congruent mod p to some rational integer. Now let $\alpha = a_0 + a_1\zeta + \cdots + a_{p-1}\zeta^{p-1}$ with each $a_i \in \mathbb{Z}$ and at least one $a_i = 0$. If $n \in \mathbb{Z}$ and n divides α, then n divides each a_j. Similarly, suppose each $a_i \in \mathbb{Z}/p\mathbb{Z}$ with at least one $a_i = 0$, then $p \mid \alpha \implies p$ divides each a_j.

Observe that $1 + \zeta + \cdots + \zeta^{p-1} = 0$, so we may use any subset of $\{1, \zeta, \ldots, \zeta^{p-1}\}$ with $p - 1$ elements as a basis of the $\mathbb{Z}$-module[6] $\mathbb{Z}[\zeta]$.

[6] A module is the generalisation of a vector space from a *base field* to a *base ring*. So a $\mathbb{Z}$-module satisfies all the vector space axioms, just over the ring of integers rather than a base field.

Since at least one $a_i = 0$, the other a_j's give the coefficients with respect to a basis, and the first claim follows. For the second claim we repeat the same argument but replace $\mathbb{Z}$ with $\mathbb{Z}/p\mathbb{Z}$.

We now return to the FLT equation (1.2.1). Consider

$$\prod_{i=0}^{p-1}(x+\zeta^i y) = (z)^p$$

as an equality of ideals. We prove that the ideals $(x+\zeta^i y)$ are pairwise coprime for $0 \leqslant i < p$. This means each of these ideals must be the pth power of an ideal,

$$(x+\zeta^i y) = A_i^p.$$

Each A_i^p is a principal ideal. This is the crucial part where p being a regular prime becomes useful. By definition of regular prime, we have that p does not divide the class number of $\mathbb{Q}(\zeta)$. So A_i^p being principal implies that A_i itself is principal, say $A_i = (\alpha_i)$. Hence,

$$(x+\zeta^i y) = (\alpha_i^p) \implies x+\zeta^i y = (\text{unit})\cdot \alpha_i^p.$$

We were able to achieve this result without having to falsely assume that $\mathbb{Z}[\zeta]$ is UFD.

Let $i = 1$ and omit the subscripts, so $x+\zeta y = \varepsilon\alpha^p$ for some unit ε. Then by Proposition 1.3.6 we have $\varepsilon = \zeta^r\varepsilon_1$ for some integer r and $\varepsilon_1 = \overline{\varepsilon_1}$. As $\alpha \in \mathbb{Z}[\zeta]$, there is an $a \in \mathbb{Z}$ such that $\alpha^p \equiv a \pmod p$. So $x+\zeta y = \zeta^r\varepsilon_1\alpha^p \equiv \zeta^r\varepsilon_1 a \pmod p$. Since $\overline{a} = a$ and $\overline{p} = p$, we have $x+\zeta^{-1}y = \zeta^{-r}\varepsilon_1\overline{\alpha}^p \equiv \zeta^{-r}\varepsilon_1\overline{a} \pmod{\overline{p}} = \zeta^{-r}\varepsilon_1 a \pmod p$. Hence,

$$\begin{aligned} \zeta^{-r}(x+\zeta y) &\equiv \zeta^r(x+\zeta^{-1}y) \pmod p, \text{ or} \\ x+\zeta y-\zeta^{2r}x-\zeta^{2r-1}y &\equiv 0 \pmod p. \end{aligned} \tag{1.3.1}$$

If $1, \zeta, \zeta^{2r}, \zeta^{2r-1}$ are distinct then, as $p \geqslant 5$, then (1.3.1) implies that $p \mid x$ and $p \mid y$ which is absurd. So they cannot be distinct. Noting that $1 \neq \zeta$ and $\zeta^{2r} \neq \zeta^{2r-1}$, we get 3 cases:

(1) $1 = \zeta^{2r}$. From (1.3.1) we have $x+\zeta y-x-\zeta^{-1}y \equiv 0 \pmod p$ so $\zeta y-\zeta^{p-1}y \equiv 0 \pmod p$. This implies that $y \equiv 0 \pmod p$, which is absurd.

(2) $1 = \zeta^{2r-1}$, or equivalently $\zeta = \zeta^{2r}$. From (1.3.1) we have $(x-y)-\zeta(x-y) \equiv 0 \pmod p$ so $x-y \equiv 0 \pmod p$ which is absurd as x and y are coprime.

(3) $\zeta = \zeta^{2r-1}$. From (1.3.1) we have $x-\zeta^2 x \equiv 0 \pmod p$ so $x \equiv 0 \pmod p$ which is absurd.

Hence, Fermat's Last Theorem is proved for all regular primes p such that $p \nmid xyz$. □

Bibliography

[Neu92] Jürgen Neukirch. *Algebraische Zahlentheorie*. Springer Berlin Heidelberg, 1992.

[Was97] Lawrence Clinton Washington. *Introduction to Cyclotomic Fields*. Springer New York, 1997.

[Wil95] Andrew John Wiles. "Modular elliptic curves and Fermat's Last Theorem". In: *Annals of Mathematics* 141.3 (1995).

[WT95] Andrew John Wiles and Richard Lawrence Taylor. "Ring theoretic properties of certain Hecke algebras". In: *Annals of Mathematics* 141.3 (1995).