# Fermat's Last Theorem

Cyrill Graf and James Guillan

12.12.2023

## Introduction

In this section we will report about Fermat's last theorem, which is a famous theorem from number theory that, until fairly recently, remained a conjecture that attracted the attention of mathematicians for over three centuries. Our section will closely follow the paper by Kramer [1].

Pierre de Fermat was born on the $20^{\text{th}}$ of August 1601 in France. At the insistence of his father, Fermat chose to lead a judicial career which resulted in him becoming a lawyer and, in 1631, the *Conseiller au Parlement de Toulouse* as well. In place of having any large political ambitions, Fermat fervently studied mathematics in his spare time, in particular number theory, which at the time, still mainly consisted of the works collected in Diophantus' book *Arithmetica* which was written in the $3^{\text{rd}}$ century. And so it was that Fermat studied the then by Claude Gaspar Bachet newly issued *Arithmetica* and wrote many notes on the margins on his copy. Most of his notes were observations for which he sometimes made little incomplete proof sketches. His son Samuel recognised the importance of his father's observations and thus published a new version of Diophantus' *Arithmetica* five years after Fermat's death in 1665, now containing the marginal notes from Fermat's copy, making his observations available to the following generations of mathematicians. Until 1995, all but one of these observations were rigorously proven by notable mathematicians such as Leonhard Euler (1707-1783) for instance. The problem that remained, eluded mathematicians for three and a half centuries, becoming known as *Fermat's Last Problem* or Fermat's Last Theorem, it being the last of his observations yet to be proven then or contradicted. The solution to this riddle came from Andrew Wiles in 1995, who worked on the problem for more than seven years with Richard Taylor. We will look at the main parts of the proof in the latter half of this report.

Fermat's conjecture came about after studying the part of Diophantus' book about *Pythagorean triples*, i.e. triples of integers $(a, b, c)$ that satisfy the equation

$$a^2 + b^2 = c^2. \tag{1}$$

Some of these numbers were already known to the Babylonians in 1600 B.C. and were revered by the so-called Pythagoreans in ancient Greece. In *Arithmetica*, Diophantus concerned himself with the question of how to systematically construct such Pythagorean triples. If $(a, b, c)$ is a Pythagorean triple then so is $(|a|, |b|, |c|)$, hence it suffices to restrict ourselves to the *primitive Pythagorean triples*. These are Pythagorean triples $(a, b, c)$ satisfying $a, b, c > 0$,

$\gcd(a, b, c) = 1$ and $a$ even, so all other solutions are obtained by changing signs, permuting a,b and by multiplying with some non-zero integer. A complete description of all the primitive triples, of which there are then infinitely many, are given in the following theorem.

**Theorem 1.** *A triple of positive integers $(a, b, c)$ is a primitive Pythagorean triple if and only if there exist two integers $x, y$ with $x > y > 0$, $gcd(a, b, c) > 1$ and $x, y$ not of the same parity such that*

$$a = 2xy$$
$$b = x^2 - y^2$$
$$c = x^2 + y^2$$

Fermat asked himself how many such solution triples of positive natural numbers exist when one replaces the exponent 2 in the Pythagorean equation by an integer $n \geq 3$. After his attempts to find such triples, Fermat wrote on the margin of his copy of Diophantus' book the following statement:

> It is impossible to separate a cube into two cubes, or a biquadrate into two biquadrates, or in general any power higher than the second into powers of like degree; I have discovered a truly remarkable proof which this margin is too small to contain.

In modern language, Fermat's statement means:

**Theorem 2** (Fermat's Last Theorem)**.** *For every integer $n \geq 3$, there exist no three integers $a, b, c$ with $abc \neq 0$ satisfying:*

$$a^n + b^n = c^n.$$

No proof of this statement was ever found among Fermat's papers. He did, however, write a proof for the case when $n = 4$. Mathematicians have debated whether Fermat indeed possessed the proof of the theorem. Perhaps, at one point, he mistakenly believed he had found such a proof. Despite Fermat's honesty and frankness in acknowledging imperfect conclusions, it is very difficult to understand today, how some of the most distinguished mathematicians could have failed to rediscover his proof, given that he did not have the tools of modern number theory.

# From 1637 to 1980

Quite clearly, any proof of Fermat's theorem will rely on a proof by contradiction. We suppose there exists $m > 2$ and a triple of positive natural numbers $(a, b, c)$ such that

$$a^m + b^m = c^m.$$

and try to produce a contradiction. We observe the following:

*Remark.* If Fermat's last theorem holds for $p \in \mathbb{N}$ then it also holds for $n = pl$, for every $l \in \mathbb{N}$

*Proof.* Suppose Fermat's last theorem does not hold for $n = pl$, that is, there exist non-zero integers $a, b, c$ such that $a^n + b^n = c^n$. But this means that

$$\left(a^l\right)^p + \left(b^l\right)^p = a^{lp} + b^{lp} = a^n + b^n = c^n = c^{lp} = \left(c^l\right)^p.$$

Since $a^l, b^l, c^l$ are non-zero integers, Fermat's theorem also does not hold for $p$. □

By the remark it is sufficient to suppose that $m$ is a prime number $m = p > 2$ or $m = 4$: If $m > 2$ is an integer, then either $m = pl$ for a prime $p > 2$ and some $l \in \mathbb{N}$ or $m = 4l$ for some $l \in \mathbb{N}$.

Furthermore, without loss of generality, we may assume that $a, b, c$ are relatively coprime. If not, i.e. $d := \gcd(a, b, c) > 1$, then we can divide the triple by $d$ to get a new solution triple $(a', b', c')$ which share no common divisor. Lastly, observe that exactly one of the numbers $a, b, c$ is even when $a, b, c$ are coprime. To see this, we reduce the equation modulo 2 and use the fact that $a^2 \equiv a \mod 2$ to get

$$a + b \equiv c \mod 2.$$

If $a$ is even, then $b$ can not be even as then $c$ would be even too, which is a contradiction to the assumption that $a, b, c$ share no common factor. So if $a$ is even then $b$ and $c$ are odd. Similarly we get for $b$ even that $a$ and $c$ must be odd. If $a$ and $b$ are both odd, then $c$ must naturally be even. Hence exactly one of the numbers $a, b, c$ must be divisible by 2. So we suppose $a$ is even.

We begin with Fermat's proof for the case $n = 4$. We prove a slightly more general statement first, in which the so-called *infinite descent* technique is utilised which consists of constructing a smaller solution to the problem from an existing one we assume to exist. Repeating the argument infinitely many times yields a sequence of positive integers $a_0 > a_1 > \cdots > 0$ and thus a contradiction.

**Theorem 3.** *There exist no integers $a, b, c$ with $abc \neq 0$ such that*

$$a^4 + b^4 = c^2. \tag{2}$$

*Proof.* Suppose that $(a, b, c)$ is a triple of positive integers satisfying (2). Similarly as discussed beforehand, we may assume without loss of generality that $\gcd(a, b, c) = 1$, as well as that $a$ is even. Then $(a^2, b^2, c)$ is a primitive Pythagorean triple: $(a^2)^2 + (b^2)^2 = c^2$. By theorem 1 there exist integers $x, y$ of different parity with $x > y > 0$, $\gcd(x, y) = 1$ and

$$\begin{aligned} a^2 &= 2xy \\ b^2 &= x^2 - y^2, \\ c &= x^2 + y^2. \end{aligned}$$

By the second equation $b^2 + y^2 = x^2$ and $\gcd(b, y, x) = 1$, hence by theorem 1 again there exist integers $w, z$ such that they are of different parity, $w > z > 0$, $\gcd(w, z) = 1$ and

$$\begin{aligned} b &= 2wz \\ y &= w^2 - z^2, \\ x &= w^2 + z^2. \end{aligned}$$

Hence

$$a^2 = 2xy = 4wz(w^2 + z^2)$$

Since $w, z, (w^2 + z^2)$ are pairwise coprime and by the uniqueness of the factorisation into primes, we conclude that $w, z, (w^2 + z^2)$ are squares of positive integers $e, f, g$:

$$\begin{aligned} w &= e^2, \\ z &= f^2, \\ w^2 + z^2 &= g^2. \end{aligned}$$

which satisfy $e^4 + f^4 = g^2$. We have that $c = x^2 + y^2 = (w^2 + z^2)^2 + (w^2 - z^2)^2 > g^4 > g > 0$, thus we have found a new triple $(e, f, g)$ satisfying (2) with $c > g$. In this way, we can construct infinitely many triples of integers $(a_i, b_i, c_i)$ satisfying (2) and $c_i > c_{i+1}$, yielding an infinite decreasing sequence of positive integers $c_0 > c_1 > c_2 > \cdots > 0$, but this is absurd. □

Fermat's Last Theorem for $n = 4$ follows immediately from this: Suppose there exist natural numbers $a, b, c$ such that $a^4 + b^4 = c^4$, then setting $d = c^2$, it holds that $a^4 + b^4 = d^2$, contradicting the statement of theorem 3.

The same technique can be used for proving similar theorems such as: The following equations have no solution in non-zero integers: $x^4 - y^4 = \pm z^2$, $x^4 + 4y^4 = z^2$, $x^4 - 4y^4 = \pm z^2$.

Indeed the technique of infinite descent was also used by Leonhard Euler in this proof for the case $n = 3$. An important step in Euler's proof uses the divisibility properties of integers of the form $a^2 + 3^2$ thus being more technical than the case $n = 4$. Later, Peter Gustave Lejeune Dirichlet (1805-1859) and Adrien-Marie Legendre (1752-1833) proved the case for $n = 5$ in 1825. An important contribution was made by Sophie Germain (1776-1833): *If $p$ is an odd prime such that $2p+1$ is also prime then Fermat's theorem holds for $p$ when $p \nmid abc$.* By relating the problem to cyclotomic fields, Ernst Eduard Kummer (1810-1893) managed to prove the theorem for all prime exponents smaller than 100, except for 37, 59, 67. During the $20^{\text{th}}$ century only minor progress was made, building on the work of Kummer. Number theory also seemed to be drifting away from Fermat's problem at the time in spite there being offered a lucrative sum of 100'000 DM (or approximately 50'000 CHF). Taking advantage of the improving computer technology, in 1976 Wagstaff confirmed numerically via a computer that Fermat's theorem held true for all primes smaller than 125'000.

## Modern Approach: Frey curves and their conductor

Thanks to the work of Fermat, Euler and Legendre or Dirichlet, and the considerations from before, we know that it is enough to suppose that there exists a triple $(a, b, c)$ and a prime $l > 5$ such that

$$a^l + b^l = c^l$$

holds. We already saw that without loss of generality we can assume that $a, b, c$ are coprime. For each such a triple we define a corresponding elliptic curve, the Frey curve, given by

$$E_{a,b,c} : Y^2 = X\left(X - a^l\right)\left(X + b^l\right) = X^3 + \left(b^l - a^l\right)X^2 - (ab)^l X.$$

Furthermore, the discriminant of a Frey curve is

$$\Delta_{E_{a,b,c}} = 16(abc)^{2l}.$$

If we now can show that such an elliptic curve cannot exist, then we are done as each Frey curve corresponds to a solution of Fermat's last theorem. In order to do so, we define the notion of a good and a bad reduction.
Let

$$\tilde{E} : Y^2 = X^3 + \tilde{a}_1 X^2 + \tilde{a}_2 X + \tilde{a}_3 = f(X)$$

be the reduction modulo $p$ (for some prime $p$) of the elliptic curve

$$E : Y^2 = X^3 + a_1 X^2 + a_2 X + a_3.$$

and $\Delta$ be the discriminant of the elliptic function. If the prime $p$ does not divide $\Delta$, then we call the reduction good, else we call it bad. We have seen in a previous talk that if a reduction is bad then either of the following two cases happen

- Two of the zeroes of $f(X)$ are equal. In this case we call the reduction a multiplicative reduction and the curve $\tilde{E}$ has a node.
- All three zeroes of $f(X)$ are the same, then we call the reduction an additive reduction and $\tilde{E}$ has a cusp.

With this in mind, we can now set up the notion of a semistable elliptic curve, which is an elliptic curve $E$ over $\mathbb{Q}$ for which the reduction $\tilde{E}$ modulo a prime $p$ of $E$ has either a node or is good for all possible primes.
For semistable elliptic curves $E$ over $\mathbb{Q}$ we define the conductor $N_E$ as the product of all primes $p$ for which the reduction $\tilde{E}$ modulo $p$ of $E$ has a node. Since $E$ is semistable the only primes which give us a node in the reduction are exactly the primes which divide the discriminant $\Delta$. Therefore we get

$$N_E := \prod_{\substack{p|\Delta \\ p \text{ prime}}} p.$$

The conductor $N_E$ is well defined, because we have only finitely many primes $p$ that divide $\Delta$.
We now come back to the Frey-curves and show that they are semistable. It is obvious from the definition that the roots of $E_{a,b,c}$ are $\{0, a^l, -b^l\}$. If a prime $p$ divides $a$, then it cannot divide $b$, as otherwise it would also divide $a^l + b^l = c^l$, which is a contradiction to the assumption that $a, b, c$ are coprime. Similarly we get that every prime $p$ that divides $b$ cannot divide $a$. Hence for every $p$ that divides either $a$ or $b$ the reduction $\mathcal{E}$ mod $p$ given by $X(X - \tilde{a}^l)(X + \tilde{b}^l)$ has two distinct roots, namely $\tilde{a}^l$ and $\tilde{b}^l$, and therefore cannot be an additive reduction. If the prime $p$ divides $c$, then we must have $a^l + b^l \equiv 0 \mod p$. Then we must also have $a^l \equiv -b^l \mod p$, which cannot be zero as otherwise we would have $\gcd(a, b, c) \geq p > 1$, which is a contradiction. Therefore the reduction $\tilde{E}$ for such a $p$ cannot be additive as it has two distinct roots, namely 0 and $\tilde{a}^l$. Therefore we can conclude that the Frey curves are semistable ellitpic curves.
Since Frey curves are semistable we can calculate the conductor $N_{E_{a,b,c}}$ and get

$$N_{E_{a,b,c}} = \prod_{\substack{p|\Delta_{E_{a,b,c}} \\ p \text{ prime}}} p = 2 \cdot \prod_{\substack{p|abc \\ p \text{ prime} \\ p \neq 2}} p.$$

# Modular to the level $N_E$ and Ribet's Reduction

Before we can state the definition of a modular elliptic curve of level $N$, we first need to recall some things.
An element $\tau \in \mathbb{Q} \cup \{\infty\}$ is a cusp if and only if there exists a matrix $g$ with trace $\pm 2$ such that $g\tau = \tau$. $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ has only one cusp, namely $\infty$.
A cusp form $\xi$ of integral weight $k$ is a holomorphic function defined on the upper complex half plane $\mathbb{H}$ which satisfies the relation

$$\xi \mid_k g = (c\tau + d)^{-k} \xi(g\tau) = \xi(\tau)$$

for all fractional linear transformations $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = g \in \Gamma$ and which vanishes at the cusp $\infty$. We can use the invariance of $\xi$ with respect to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma$ and the fact that $\xi$ vanishes at $\infty$ to deduce that $\xi$ has a Fourier expansion of the form

$$\xi(\tau) = \sum_{n=1}^{\infty} c_n e^{2\pi i \tau}.$$

We call $\xi$ normed if $c_1 = 1$ holds.
We can now generalise this to the notion of a generalised cusp form by replacing $\Gamma$ in the above definition by an arbitrary subgroup $\Pi$ of $\Gamma$ and by requiring $\xi(\tau)$ to vanish at all cusps of $\Pi$. We then call such a cusp form a generalised cusp form with respect to $\Pi$ of integral weight $k$, however some authors drop the generalised part and refer to this as a cusp form, when it is inherently clear which subgroup is used in the definition.
In the last talk we saw that the cusp forms of weight $k$ for $\Gamma$ form a complex vector space $\mathcal{S}_k$ whose dimension is given by

$$\dim_{\mathbb{C}}(\mathcal{S}_k) = \begin{cases} \lfloor \frac{k}{12} \rfloor - 1 & \text{if } k \equiv 2 \mod 12 \\ \lfloor \frac{k}{12} \rfloor & \text{if } k \equiv 2 \mod 12. \end{cases}$$

Unfortunately we cannot do this for a generalised cusp forms. Even though they form complex vector space $\mathcal{GS}_k(\Pi)$, we cannot give a formula for its dimension due to the fact that, the smaller the subgroup, the less restrictions we impose on $\mathcal{GS}_k(\Pi)$. To demonstrate this, let us pick $\Pi = \{I\}$, where $I$ is the $2 \times 2$ identity matrix. We then clearly have $\xi \mid_k I = (0\tau + 1)^k \xi = \xi$ for all functions $\xi$ on the upper half plane.
So if we let $\Pi$ be any subgroup of $\Gamma$ we will run into problems, as the definition is too broad. We therefore choose a special set of subgroups, the so called modular subgroups of weight $N$ given by

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \middle| c \equiv 0 \mod N \right\}.$$

We can now finally set up our definition of a modular elliptic function.
We call an elliptic curve $E = E(\mathbb{Q}) : Y^2 = f(X)$ over $\mathbb{Q}$ modular to the level $N$ if every prime $p$ coprime to $N$ results in a good reduction $E(\mathbb{F}_p)$ and there exists a non-zero normed generalised cusp form $\xi$ of weight 2 with respect to $\Gamma_0(N)$, such that the $p$-th Fourier coefficient satisfies

$$c_p = p - \mid E(\mathbb{F}_p) \mid = p - \left| \left\{ (x, y) \in \mathbb{F}_p^2 | y^2 - f(x) = 0 \right\} \right|.$$

At first glance this definition seems to be quite restrictive in the sense that it seems to exclude quite a lot of elliptic curves. However, this is not the case. As already conjectured in 1957 by Yutaka Taniyama and Goro Shimura and later proven by Andrew Wiles, all elliptic curves are modular to some level. This surprising result is summarised in the modularity theorem, which can a bit loosely stated as

**Theorem 4** (Modularity theorem)**.** *Every (rational) elliptic curve $E$ is modular to the level $N = N_E$.*

Since we now know what it means for an elliptic curve to be modular to the level $N$, we can now take a look at Ribet's reduction (or at least a simplified version, which applies to the Frey curves).

**Theorem 5** (Ribet's reduction)**.** *Suppose there is a non-trivial integral triple $a, b, c$ together with an exponent $l > 5$ such that $a^l + b^l = c^l$. Now, given that the associated Frey-curve*

$$E_{a,b,c} : Y^2 = X\left(X - a^l\right)\left(X + b^l\right)$$

*is modular to the level $N = N_E$, then it is also modular to the level $\frac{N}{p}$ for any odd prime $p$ dividing $N$.*

Some early work of Wiles and Taylor shows that the Frey curve is indeed modular to the level $N_{E_{a,b,c}}$, where $N_{E_{a,b,c}}$ is the conductor of $E_{a,b,c}$. But then we must have by Ribet, that we can divide out all odd prime factors of the conductor. Doing so leads to the fact that the Frey-curve $E_{a,b,c}$ is indeed modular to the level 2, hence if we can show that there are no such curves, then we are done.

## Modular elliptic curves of weight 2

Before we can finish the proof of Fermat's last theorem with the proof that there are no elliptic curves that are modular to the level 2, we first want to understand the space of generalised cusp forms for $\Gamma_0(N)$ a bit better.
For this, we realise that the quotient $\Gamma_0(N)\backslash\mathbb{H}$ is an open Riemann surface (often called the modular curve $Y(\Gamma_0(N))$ ), that is an open one-dimensional complex manifold. Since this space is Tychonoff we can compactify it and get the Riemann surface $\overline{\Gamma_0(N)\backslash\mathbb{H}}$ (one can show that adding the cusps of $\Gamma_0(N)$ gives said compact Riemann surface), which is often called the compactified modular curve $X(\Gamma_0(N))$ or $X_0(N)$ and for which we can define its geometric genus.
The geometric genus of a complex manifold $M$ of dimension $k$ is the dimension of the space of holomorphic $k$-forms on $M$.
We will omit the calculations, but one can find the genus $g_N$ of $\overline{\Gamma_0(N)\backslash\mathbb{H}}$ by studying the quotient maps $\overline{\Gamma_0(N)\backslash\mathbb{H}} \to \overline{\Gamma_0(1)\backslash\mathbb{H}}$ and using the Hurwitz genus formula and the fact that $\overline{\Gamma_0(1)\backslash\mathbb{H}}$ has genus 0. One finds for $p > 3$ prime

$$g_p = \begin{cases} n-1 & \text{if } p = 12n+1 \\ n & \text{if } p = 12n+5, 12n+7 \\ n+1 & \text{if } p = 12n+11. \end{cases}$$

Moreover, one has

$$\begin{cases} g_N = 0 & \text{if } N = 1, 2, 3, \ldots, 10, 12, 13, 16, 18, 25; \\ g_N = 1 & \text{if } N = 11, 14, 15, 17, 19, 20, 21, 24, 27, 32, 36, 49; \\ g_N = 2 & \text{if } N = 22, 23, 26, 28, 29, 31, 37, 50. \end{cases}$$

Since all those definitions are rather technical and complex we want to gain some visual intuition before we continue. Let us first take a the elliptic curve $Y^2 = X^3 - X$. If we take a look at the real cross section of its plot, then we have the following image.

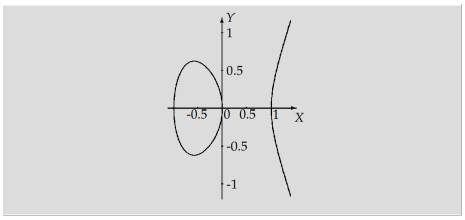


Figure 1: the real cross section of the elliptic curve $Y^2 = X^3 - X$

If we add the point $\mathcal{O}$ as an infinitely far point to the complex plane, the two ends of the curve going to $\pm\infty$ are glued together, so when we consider the elliptic curve up to homotopy or up to "continuous deformations" we get the following picture over the reals.

Figure 2: the real image of the elliptic curve $Y^2 = X^3 - X$ after adding the point $\mathcal{O}$

Over the complex numbers it appears as a torus.

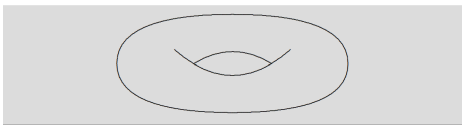

Figure 3: The complex image of an elliptic curve: A Tours

Figure 4 shows an example of a surface of genus 3. In simple terms, the value of an orientable surface's genus is equal to the number of "holes" it has.

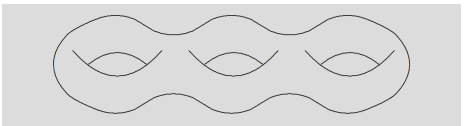

Figure 4: The complex image of a modular curve of genus $g_N = 3$

We do exactly this kind of procedure, but instead of an elliptic curve we use a fundamental domain $\Gamma_0(N)\backslash\mathbb{H}$ and instead of the point $\mathcal{O}$ we use the cusps of $\Gamma_0(N)$ to arrive at a complex manifold. Since we now have a somewhat visual image, we can finish the proof using the following theorem.

**Theorem 6.** *There is a natural isomorphism of vector spaces between the cusp forms of weight* 2 *with respect to* $\Gamma_0(N)$ *and the regular differential* 1*-forms on the complex manifold* $\overline{\Gamma_0(N)\backslash\mathbb{H}}$*. The isomorphism* $\mathcal{GS}_2(\Gamma_0(N)) \to \Omega^1\left(\overline{\Gamma_0(N)\backslash\mathbb{H}}\right)$ *is characterised by* $\xi(\tau) \mapsto \xi(\tau)d\tau$*. Thus*

$$\dim_{\mathbb{C}}\left(\mathcal{GS}_2(\Gamma_0(N))\right) = g_N.$$

Using this theorem we can immediately conclude that there are no modular elliptic curves to the level 2, since we have

$$\dim_{\mathbb{C}}\left(\mathcal{GS}_2(\Gamma_0(2))\right) = g_2 = 0.$$

This finishes the proof of Fermat's last theorem as we have successfully shown that Frey curves, which are a direct consequence from the existence of solutions to the equation $a^l + b^l = c^l$, cannot be modular of level 2 (as no such curves can exist) and therefore cannot exist.

# References


[1] Kramer, J. (2000). *Der grosse Satz von Fermat - die Lösung eines 300 Jahre alten Problems.* In: Aigner, M., Behrends, E. (eds) Alles Mathematik. Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-96366-6_12.

[2] P.Ribenboim, *13 Lectures on Fermat's last Theorem*,Springer-Verlag, New York-Heidelberg-Berlin (1979).

[3] J.H. Silverman and J.T. Tate *Rational Points on Elliptic Curves*, Springer-Verlag (1992).

[4] J.S. Milne. *Elliptic curves*, BookSurge Publishers, (2006)