

La preuve en mathématiques

Université de Lille III

24-28 Mai 2005

What is a complex proof:

Wiles' example.*

Jean Petitot, CREA (Ecole Polytechnique)

Contents

1	Introduction	3
2	The first proof of FLT: the case $n = 4$, the descent method and elementary arithmetics	4
3	From Euler to Kummer, the case of regular primes.	5
3.1	A brief historical survey	5
3.2	Kummer's proof (a sketch)	5
3.2.1	Case I	6
3.2.2	Case II	7
4	Faltings theorem and the Mordell-Weil conjecture	7
5	Frey's breakthrough	8
6	Elliptic curves as projective plane cubics	9
6.1	Weierstrass normal form	9
6.2	Discriminant, j -invariant, conductor, and semi-stability .	10
6.3	The group structure	11
7	Elliptic curves as complex tori	12

*This text is a revised and completed English version of a lecture given at the Mathematical Center of the Ecole des Hautes Etudes en Sciences Sociales in 1997.

8	From complex tori to cubics	13
9	From cubics to complex tori	13
10	Zeta function and Dirichlet L-functions	14
10.1	Riemann zeta function	14
10.2	Mellin transform and functional equation	15
10.3	Zeros of Zeta, distribution of primes, and Riemann hypothesis	16
10.4	Dirichlet series and Dirichlet L -functions	18
11	Modular forms	19
11.1	Two classes of L -functions	19
11.2	Modular functions for $SL(2, \mathbb{Z})$	20
11.3	Fourier expansion at infinity, modular forms, and cusp (parabolic) forms	21
11.4	L -functions of cusp forms	23
11.5	Hecke operators for $SL(2, \mathbb{Z})$	24
11.6	Petersson scalar product and Euler product of cusp forms	26
11.7	Fricke involution and generalization to the groups $\Gamma_0(N)$	27
11.8	Hecke operators for $\Gamma_0(N)$ and Euler products	29
11.9	New forms and Atkin-Lehner theorem	29
12	L-functions of elliptic curves and Eichler-Shimura theory	30
12.1	Encoding geometrico-arithmetic information into L -functions	30
12.2	The L -function of an elliptic curve E	31
12.3	The modular curve $X_0(N)$ and its Jacobian $J_0(N)$	31
12.4	Modular elliptic curves	32
12.5	The Eichler-Shimura construction	33
13	From Taniyama-Shimura-Weil to Fermat: Ribet theorem.	33
14	Encoding information in Galois representations	35
14.1	Torsion points and Galois representations	35
14.2	Modular representations and Deligne theorem	36
15	Wiles side story	36
15.1	Semi-stable modular lifting conjecture for $p = 3, 5$	37
15.2	The construction of the auxiliary elliptic curve	39
15.3	Lifting to p -adic representations: from characteristic p to characteristic 0	40
15.4	Deformation data and Barry Mazur conjectures	41
15.5	Gorenstein rings and “cotangent spaces”	43

15.6 Wiles own description of his proof	44
16 Conclusion: the categorical complexity of a proof	51

1 Introduction

In his *Panorama des Mathématiques pures. Le choix bourbachique*, Jean Dieudonné classified theorems in six classes and focus on two of them:

1. “Les problèmes qui engendrent une méthode” as analytic number theory or finite groups theory;
2. “Les problèmes qui s’ordonnent autour d’une *théorie* générale, féconde et vivante” as Lie group theory or algebraic topology;

And he gave the example of *modular forms*:

“La théories des formes automorphes et des formes modulaires est devenue un extraordinaire carrefour où viennent réagir les unes sur les autres les théories les plus variées : Géométrie analytique, Géométrie algébrique, Algèbre homologique, Analyse harmonique non commutative et Théorie des nombres.”

With many creative mathematicians, Jean Dieudonné was convinced that the mathematical interest of a proof depends upon its capacity of circulating between many heterogeneous theories and of *translating* some parts of theories into completely different other ones.

I want to develop this idea relating proofs to this very special type of “inter-expression” Albert Lautman called the *unity* of mathematics. As the best method is to scrutinize a good example, I will focus on one of the most prototypical examples of a “great” complex proof, namely Wiles-Taylor’s proof of the Taniyama-Shimura-Weil conjecture on elliptic curves, a corollary of which is, due to a theorem of Ribet, Fermat last theorem (FLT). Of course the challenge is quite impossible to be taken up in one hour, but I will try nevertheless to present some key ideas of the proof.

We will arrive at the conclusion that the best would be to work inside the framework of *category theory* since “translations” are in general *functors* from one category into another. Wiles’ proof is “complex” in the sense that it contains an impressive density of functorial changes of category.

2 The first proof of FLT: the case $n = 4$, the descent method and elementary arithmetics

Consider Fermat equation $x^n + y^n = z^n$ and suppose that (x, y, z) is a non trivial integral solution. We can suppose that x, y, z are relatively primes and that $n = 4$ or n is an odd prime.

The case $n = 4$ was proved by Fermat himself using a “descent argument” based on the fact that if (a, b, c) is a Pythagorean triple (that is a triple of positive integers such that $a^2 + b^2 = c^2$), then the area $ab/2$ of the right triangle of sides a, b, c cannot be a square.

The problem can be reduced to find non trivial integral solutions of the equation

$$u^4 + v^4 = (u^2)^2 + (v^2)^2 = w^2$$

with w a square, $\gcd(u, v) = 1$, u odd and v even. But, in a Pythagorean triple (a, b, c) the integers a and b cannot both be odd. So we can suppose that a is odd and b even, and of course $c > 0$. But a theorem of Diophantus (and in 1630 Fermat was precisely working on Diophantus’ *Arithmetica* ¹) says that there exist then integers m, n relatively prime $(m, n) = 1$ and not both odd such that $a = m^2 - n^2$, $b = 2mn$, $c = m^2 + n^2$.

Due to Diophantus’ theorem, there exist integers m, n relatively prime $(m, n) = 1$ and not both odd such that $u^2 = m^2 - n^2$, $v^2 = 2mn$, $w = m^2 + n^2$. Therefore $m^2 = u^2 + n^2$ with u odd and we can apply Diophantus *again*: there exist integers p, q relatively prime such that, $u = p^2 - q^2$, $n = 2pq$, $m = p^2 + q^2$. Therefore $v^2 = 2mn = 4pq(p^2 + q^2)$. As v is even $(\frac{v}{2})^2 = pq(p^2 + q^2)$ is a square. But as $p, q, p^2 + q^2$ are relatively prime they must all be squares: $p = r^2$, $q = s^2$, $p^2 + q^2 = k^2$. But then $r^4 + s^4 = k^2$ is a new solution and as $(\frac{v}{2})^2 = pq(p^2 + q^2) = r^2s^2(r^4 + s^4)$ implies $v = 2rs\sqrt{r^4 + s^4}$ we have $r < v$ and $s < v$. The new integral solution is therefore strictly smaller than the initial one and it is non trivial since the initial solution is not trivial. Hence Fermat’s famous “descent argument”: after a finite number of steps we would get negative solutions \Rightarrow contradiction $\Rightarrow (u, v, w)$ can’t exist.

Fermat’s proof contains the deep idea of descent formalized later by Mordell. But it is “elementary” in the sense it uses only elementary arithmetic computations.

¹It is in the margin of Diophantus’ treatise that Fermat wrote his remark: “I have discovered a truly marvelous proof of this proposition that this margin is too narrow to contain”.

3 From Euler to Kummer, the case of regular primes.

3.1 A brief historical survey

The history of the successive proofs of FLT for odd primes p is a true Odyssey. We cannot summarize it here. In a nutshell we can say that, during what could be called an “Eulerian” period, many particular cases were successively proved by Sophie Germain, Dirichlet, Legendre, Lamé, etc. using a fundamental property of *unique factorization of integers in prime factors* in algebraic extensions of \mathbb{Q} . But this property is *not* always satisfied. In 1844 Ernst Kummer was able to abstract the property for a prime p to be *regular*, proved FLT for all regular primes and shew that the *irregularity* of primes was the main obstruction to a natural algebraic proof.

It must be strongly emphasized that it is for this proof that Kummer invented the concept of “ideal” number which will become with Dedekind the founding concept of *ideal* of a ring (the basis of commutative algebra) and proved his outstanding result that unique factorization in prime factors remains valid for “ideal” numbers.

After this breakthrough, a lot of particular cases of irregular primes were proved which enabled to prove FLT up to astronomical p ; a lot of false proofs were also published, and a lot of computational verifications were made; but no *general* proof was found. It is to its extraordinary resistance to purely arithmetic and algebraic proofs than FLT owes his legendary celebrity.

3.2 Kummer’s proof (a sketch)

Kummer’s basic idea was to *factorize* Fermat equation in the ring $\mathbb{Z}[\zeta]$ where ζ is a primitive p th root of unity and to work in the *cyclotomic extension* $\mathbb{Z}[\zeta] \subset \mathbb{Q}(\zeta)$ of the elementary arithmetic $\mathbb{Z} \subset \mathbb{Q}$. In $\mathbb{Z}[\zeta]$ we have the factorization

$$x^p - 1 = \prod_{j=0}^{j=p-1} (x - \zeta^j),$$

the polynomial

$$\Phi(x) = x^{p-1} + \dots + x + 1 = \prod_{j=1}^{j=p-1} (x - \zeta^j)$$

is irreducible over \mathbb{Q} and is the minimal polynomial defining ζ . The conjugates of ζ are $\zeta^2, \dots, \zeta^{p-1}$, $\mathbb{Q}(\zeta)$ is the splitting field of $\Phi(x)$ over \mathbb{Q} and

$\mathbb{Q}(\zeta)/\mathbb{Q}$ is a Galois extension. We note that $\Phi(1) = p$. The prime p is ramified in $\mathbb{Z}[\zeta]$ (and in fact is the only ramified prime). More precisely, $(1 - \zeta)$ is a prime ideal of $\mathbb{Z}[\zeta]$ and there exists some unit u s.t.

$$p = u(1 - \zeta)^{p-1}$$

In $\mathbb{Z}[\zeta]$ we have the decomposition

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

In $\mathbb{Z}[\zeta]$, unique factorization of an integer in prime factors is no longer necessarily true. But Kummer shew it remains valid for ideals.

It is traditional to distinguish two cases in the tentative proof of FLT. We suppose that a solution (x, y, z) exists and we look at its relations with the prime power p .

3.2.1 Case I

Suppose first that x and y are *prime to p* . This implies that the ideals $(x + \zeta^j y)$ are *relatively prime*.

As the product of the $(x + \zeta^j y)$ is the p th power $(z)^p$, each $(x + \zeta^j y)$ is therefore a p th power and we have in particular

$$(x + \zeta y) = \mathfrak{a}^p$$

which shows that \mathfrak{a}^p is a *principal* ideal.

It is here that the property of regularity arises.

Intuitive definition. p is a regular prime if when a p th power \mathfrak{a}^p is principal \mathfrak{a} is already itself a principal ideal.

Technical definition. p is a regular prime if it doesn't divide the class number h_p of the cyclotomic field $\mathbb{Q}(\zeta)$.

As \mathfrak{a}^p is principal, if p is a regular prime, \mathfrak{a} is principal: $\mathfrak{a} = (t)$, $(x + \zeta y) = (t)^p$ and there exists therefore some unit u in $\mathbb{Z}[\zeta]$ s.t.

$$x + \zeta y = ut^p.$$

The idea is then to compare $x + \zeta y$ with its complex conjugate $x + \bar{\zeta}y$ using congruences mod p in $\mathbb{Z}[\zeta]$.

Using the fact that $\{1, \zeta, \dots, \zeta^{p-2}\}$ is an integral basis of $\mathbb{Z}[\zeta]$ over \mathbb{Z} and developing t as $t = \sum_{i=0}^{i=p-2} \tau_i \zeta^i$, one shows first that $t^p \equiv \bar{t}^p \pmod{p\mathbb{Z}[\zeta]}$. Secondly, using a lemma of Kronecker, one shows that u being a unit, there exists j s.t. $\frac{u}{\bar{u}} = \zeta^j$. One concludes that

$$x + \zeta y = ut^p = \zeta^j \bar{u} t^p \equiv \zeta^j \bar{u} \bar{t}^p \pmod{p\mathbb{Z}[\zeta]} \equiv \zeta^j (x + \bar{\zeta}y) \pmod{p\mathbb{Z}[\zeta]} \quad ((C))$$

We get therefore mod $p\mathbb{Z}[\zeta]$ a *linear relation* between $1, \zeta, \zeta^j, \zeta^{j-1}$ (we use $\zeta^j \bar{\zeta} = \zeta^{j-1}$) with integral coefficients x, y coming from the supposed solution (x, y, z) of Fermat equation.

But the congruence (C) is impossible. Indeed if $1, \zeta, \zeta^j, \zeta^{j-1}$ are different powers then they are independent in $\mathbb{Z}[\zeta]$ over \mathbb{Z} . When it is not the case ($j = 0, j = 1, j = 2, j = p - 1$), one proves the particular cases.

3.2.2 Case II

The real difficulty is the case II when one of x, y, z is divided by p . We will skip it here.

Kummer's proof is marvelous and played a fundamental role in the elaboration of modern arithmetical tools. Its essential achievement is to do arithmetic no longer in \mathbb{Z} but in the ring of integers $\mathbb{Z}[\zeta]$ of the cyclotomic field $\mathbb{Q}(\zeta)$. But it remains a proof developed inside a *single* theory, namely algebraic number theory.

4 Faltings theorem and the Mordell-Weil conjecture

The natural context of a proof of FLT seems to be algebraic geometry since Fermat equation

$$x^n + y^n = z^n$$

is the homogeneous equation of a projective plane curve F . The equation has rational coefficients and FLT says that for $n \geq 3$ F has no rational points. So FLT is a particular case of computing the cardinal $|F(\mathbb{Q})|$ of the set of rational points of a projective plane curve F defined over the rational field \mathbb{Q} . To solve the problem, one needs a deep knowledge of the arithmetic properties of *infinitely many* types of projective plane curves since the genus g of F is

$$g = \frac{(n-1)(n-2)}{2}$$

and increases quadratically with the degree n . We note that for $n \geq 4$ we have $g \geq 3$. But of course it is extremely difficult to prove *general* arithmetic theorems valid for infinitely many sorts of classes of curves.

The greatest achievement in this direction was the demonstration by Gerd Faltings of the celebrated *Mordell-Weil conjecture*.

Theorem (Faltings). Let C be a smooth connected projective curve defined over a number field K and let $K \subset K'$ be an algebraic extension of the base field K . Let g be the genus C .

1. If $g = 0$ (sphere) and $C(K') \neq \emptyset$, then C is isomorphic over K' to the projective line \mathbb{P}^1 and there exist *infinitely many* rational points over K' .
2. If $g = 1$ (elliptic curve), either $C(K') = \emptyset$ (no rational points over K') or $C(K')$ is a finitely generated \mathbb{Z} -module (Mordell-Weil theorem, a deep generalization of Fermat descent method).
3. If $g \geq 2$, $C(K')$ is *finite* (Mordell-Weil conjecture, Faltings theorem).

Faltings theorem is an extremely difficult one which won him the Fields medal in 1986. But for FLT we need to go from “ $C(K')$ finite” to “ $C(K') = \emptyset$ ”.

5 Frey’s breakthrough

In 1986, Gerhard Frey introduced a completely new idea which led to Wiles-Taylor proof in 1994. The idea is to use an hypothetical solution $a^p + b^p + c^p = 0$ of Fermat equation (p an odd prime) *as parameters for an elliptic curve E* , namely what is now called a Frey curve:

$$y^2 = x(x - a^p)(x + b^p).$$

The idea is that, as far as (a, b, c) is a solution of Fermat equation and is supposed to be too exceptional to exist, the associated Frey curve E must also be in some sense “too exceptional” to exist.

We meet here a spectacular example of a *translation strategy* which consists in coding solutions of a first equation into parameters of a second equation of a completely different type and using the properties of the solutions of the second equation for gathering informations on those of the first equation. G. Frey was perfectly aware of the originality of his trick. In his paper he explains:

“In the following paper we want to relate conjectures about solutions of the equation $A - B = C$ in global fields with conjectures about elliptic curves.”

“An overview over various conjectures and implications discussed in this paper (...) should show how ideas of many mathematicians come together to find relations which could give a new approach towards Fermat’s conjecture.”

And indeed, the advantages of Frey’s “elliptic turn” are multifarious:

1. Whatever the degree p could be, we work always on an elliptic curve and we shift therefore from the full universe of algebraic plane curves to a *single* class of curves.
2. Elliptic curves are the best known of all curves and their fine Diophantine and arithmetic structures can be investigated using *non elementary* techniques from analytic number theory.
3. For elliptic curves we dispose of a strong criterion of “normality”: “good” elliptic curves are *modular* in the sense they can be parametrized by modular curves.
4. A well known conjecture, the *Taniyama-Shimura-Weil conjecture*, says in fact that *every* elliptic curve is modular.

From Frey’s idea we can derive a schema of proof for FLT:

1. Prove that Frey elliptic curves are not modular.
2. Prove the Taniyama-Shimura-Weil conjecture.

Step 1 was achieved by Kenneth Ribet who proved that Taniyama-Shimura-Weil implies Fermat and triggered a revolutionary challenge, and step 2 by Andrew Wiles and Richard Taylor for the so called “semistable” case, which is sufficient for FLT.

6 Elliptic curves as projective plane cubics

6.1 Weierstrass normal form

As a plane algebraic curve, an elliptic curve E is a projective cubic of equation

$$F(X, Y, T) = C_{X^3}X^3 + \dots + C_{T^3}T^3 = 0.$$

Its affine part in the complement of the line at infinity $T = 0$ is the affine curve of affine coordinates $x = X/T$, $y = Y/T$:

$$F(x, y) = C_{X^3}x^3 + \dots + C_{T^3} = 0$$

We can simplify this expression and reduce it to what is called a Weierstrass form by controlling the behavior of E at infinity and by using appropriate changes of variables. We get:

$$y^2 = x^3 - \frac{c_4}{48}x - \frac{c_6}{864} \tag{W_3}$$

which is of the form $y^2 = x^3 + px + q$, a typical cubic polynomial ($p = -c_4/48$, $q = -c_6/864$).

The curve E is symmetric relative to the x axis. It will be *singular* at (x, y) if and only if the *discriminant* $4p^3 + 27q^2 = 0$, that is if $\frac{1}{12(48)^2} ((c_4)^3 - (c_6)^2) = 0$.

1. If the double root is not a triple root the singular point is a normal crossing of two branches (ordinary double point or node) reducible to the normal form $y^2 = x^2(x + 1)$;
2. if it is a triple root the singular point is a cusp reducible to the normal form $y^2 = x^3$.

It must be emphasized that singular elliptic curves are “trivial” in the sense they can be *parametrized* by a projection onto a line from the singular point in such a way that rational points correspond to rational points.

1. For $y^2 = x^2(x + 1)$ take $r = y/x$ which yields the parametrization $x = r^2 - 1$, $y = r^3 - r$;
2. for $y^2 = x^3$ take the parametrization $x = r^2$, $y = r^3$.

Traditionally, the discriminant of the form W_3 has been defined as the (homogeneous of weight 12) expression $-16(4p^3 + 27q^2)$

$$\Delta = \frac{1}{1728} ((c_4)^3 - (c_6)^2)$$

(note that $1728 = 12^3$).

The discriminant of a Frey elliptic curve (which is not of the normal form W_3)

$$y^2 = x(x - a^p)(x + b^p) = x^3 + (b^p - a^p)x^2 - a^p b^p x$$

is

$$\Delta = 16(a^p b^p c^p)^2$$

6.2 Discriminant, j -invariant, conductor, and semi-stability

The discriminant of an elliptic curve E defined over \mathbb{Q} is particularly important because it encodes the properties of the reduction of E modulo a prime p . Let S_E (S for “singular”) be the set of primes p s.t. E has *bad reduction* at p (i.e. E is singular modulo p).

Proposition. $S_E = \{p \in \mathcal{P} \mid \text{has bad reduction at } p\}$ is the set

$$\{p \in \mathcal{P} \mid p \text{ divides } \Delta\}.$$

We will use also a finer invariant called the *conductor* of E .

Definition. $N_E = \prod_{p|\Delta} p^{n(p)}$ where

1. $n(p) = 1$ if E_p is a node;
2. $n(p) = 2$ if $p > 3$ and E_p is a cusp;
3. $n(p)$ is given by Tate's algorithm for $p = 2, 3$.

E is called *semi-stable* if its conductor N_E is without square factors i.e. if $n(p) = 1 \forall p$. If E is semi-stable its conductor has a very simple form: $N_E = \prod_{p|\Delta} p$.

Another fundamental invariant of E is the *modular invariant* of weight 0 defined by

$$j = \frac{c_4^3}{\Delta}$$

6.3 The group structure

One of the fundamental properties of elliptic curves is to possess a structure of *algebraic abelian group*. One can define a commutative (additive) group structure on their points using only algebraic operations. Let P and Q be two points of E . As the equation is cubic, the line PQ intersects E in a third point R . The group law is then defined by setting $P + Q + R = 0$. The neutral element 0 is the point at infinity in the y direction and the opposite $-P$ of P is therefore the symmetric of P relative to the x axis.

If we take the tangent to E at P and if it intersects E in R , we have $2P + R = 0$ (limit case of the general formula when $P = Q$).

The points on the x axis are exactly the points of order 2 s.t. $P = -P$ or $2P = 0$ (torsion points).

Fermat's descent for $n = 4$ is a particular case of inverse duplication. Write $x^4 + y^4 = z^4$ as $u^4 + v^4 = w^2$. With $X = \frac{u}{v}$ and $Y = \frac{w}{v^2}$ the equation becomes $Y^2 = X^4 + 1$. The new change of variables $X = \frac{y}{2x}$ and $Y = \frac{y^2 + 8x}{4x^2}$ yields the elliptic curve

$$y^2 = x^3 - 4x$$

Let (r, s) be the new solution obtained from an initial solution (u, v) by the descent method and let (c, d) be the point on E corresponding

to (r, s) . As $(c, d) \in E$ we have $d^2 = c^3 - 4c$. If $c = x(P)$, the initial solution (u, v) corresponds on E to the point $P' = 2P$. The descent method consists therefore in starting with a point P' and in constructing $P = P'/2$: it is a *division algorithm*.

7 Elliptic curves as complex tori

There is a completely different way of looking at elliptic curves, the equivalence of the two perspectives being one of the greatest achievements of mathematics in the first half of the XIXth century (Abel, Jacobi, etc.). It belongs to another theory, namely the theory of analytic complex functions. The problem is to study *doubly periodic* analytic functions on the complex plane \mathbb{C} . Let (ω_1, ω_2) be the two periods. We look for analytic functions $f(z)$ such that $f(z + m\omega_1 + n\omega_2) = f(z)$ for all $m, n \in \mathbb{Z}$. As ω_1 and ω_2 cannot be colinear, $\text{Im}(\omega_1/\omega_2) \neq 0$ and changing eventually a sign we can suppose $\text{Im}(\omega_1/\omega_2) > 0$. Let Λ be the lattice $\{m\omega_1 + n\omega_2\}_{m, n \in \mathbb{Z}}$ in \mathbb{C} and E the quotient space $E = \mathbb{C}/\Lambda$; E is a complex torus and f is defined on E . f is called an *elliptic function*. E being compact, f cannot be holomorphic without being constant according to Liouville theorem; f can only be a *meromorphic* function if it is not constant.

Applying residue theorem successively to f , f'/f , and zf'/f we can show:

1. f possesses at least 2 poles.
2. If the m_i are the order of the singular points a_i (poles and zeroes) of f , $\sum m_i = 0$ (this says that the *divisor* $\text{div}(f)$ is of degree 0).
3. $\sum m_i a_i \equiv 0 \pmod{\Lambda}$.

One elliptic function is of particular interest since it generates with its derivative the field of all elliptic functions. It is the Weierstrass function $\wp(z)$ which is the most evident even function having a double pole at the points of the lattice Λ . Let $\Lambda' = \Lambda - \{0\}$, the definition is:

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda'} \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right)$$

The derivative $\wp'(z)$ is an odd function possessing triple poles at the points of Λ :

$$\wp'(z) = -2 \sum_{\omega \in \Lambda} \frac{1}{(z - \omega)^3}$$

Theorem. $\wp(z)$ and $\wp'(z)$ generate the field of elliptic functions on the elliptic curve $E = \mathbb{C}/\Lambda$.

8 From complex tori to cubics

What are the relations between these two definitions of elliptic curves, one algebraic and the other analytic? In one sense, from complex tori to cubics, the relation is quite simple. Indeed $\wp(z)^3$ and $\wp'(z)^2$ have both a pole of order 6 at 0 and must be related. Some (tedious) computations on their Laurent expansions show that there exists effectively an algebraic relation between $\wp(z)$ and $\wp'(z)$, namely

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3$$

with $g_2 = 60G_4$ and $g_3 = 140G_6$, G_m being the *Eisenstein series*

$$G_m = \sum_{\omega \in \Lambda'} \frac{1}{\omega^m}$$

This means that $(\wp(z), \wp'(z))$ is on the elliptic curve E_{cub} of equation²

$$y^2 = 4x^3 - g_2x - g_3$$

which discriminant is:

$$\Delta = (g_2)^3 - 27(g_3)^2$$

the lattice Λ corresponding to the point at infinity in the y direction.

One can verify that $\Delta \neq 0$ and that E is therefore *regular*.

One of the great advantage of the torus representation is that the group structure become evident. Indeed $E_{\text{tor}} = \mathbb{C}/\Lambda$ inherits the additive group structure of \mathbb{C} and through the parametrization by $\wp(z)$ and $\wp'(z)$ this group structure is transferred to E_{cub} .

9 From cubics to complex tori

On the other direction, from projective regular plane cubics to complex tori, the relation is deeper and comes from the theory of *Riemann surfaces*. Let $E = E_{\text{cub}}$ a regular cubic. Topologically it is a torus and it is endowed with a complex structure making it a compact Riemann surface. Let γ_1 and γ_2 two loops corresponding to a parallel and a meridian of E (they constitute a \mathbb{Z} -basis of the first integral homology group $H_1(E, \mathbb{Z})$). Up to a factor, there exists a single *holomorphic* 1-form ω on E . Its periods $\omega_i = \int_{\gamma_i} \omega$ generate a lattice Λ in \mathbb{C} and we can consider the torus

²We will generically note E an elliptic curve. When it will be necessary to distinguish between its cubic algebraic representation and its toric analytic representation we will use the notations E_{cub} and E_{tor} .

$E_{\text{tor}} = \mathbb{C}/\Lambda$ which is called the *Jacobian* of E . If a_0 is a base point in E , the integration of the 1-form ω defines a map

$$\begin{aligned} \Phi : E_{\text{cub}} &\rightarrow E_{\text{tor}} \\ a &\mapsto \int_{a_0}^a \omega \end{aligned}$$

(the map is well defined since two paths from a_0 to a differ by a \mathbb{Z} -linear combination of the γ_i and the values of ω differ by a point of the lattice Λ).

Theorem. Φ is an *isomorphism* between E_{cub} and E_{tor} .

This is the beginning of the great story of *Abelian varieties*.

It is in this context, *where algebraic structures are translated and coded in analytic ones*, that one can develop an extremely deep theory of *arithmetic* properties of elliptic curves. Its “deepness” comes from *the analytic coding of arithmetics*.

10 Zeta function and Dirichlet L-functions

10.1 Riemann zeta function

One of the most beautiful way of encoding arithmetic properties in analytic structures comes from the outstanding works of Riemann on the zeta function $\zeta(s)$. The initial definition of the zeta function is extremely simple and gave rise to a lot of computations at Euler time:

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

which is a series absolutely convergent for integral exponents $s > 1$. Euler already proved $\zeta(2) = \pi^2/6$ and $\zeta(4) = \pi^4/90$. A trivial expansion shows that in the convergence domain the sum is equal to an infinite product, called an *Euler product*, containing a factor for each prime p (we note \mathcal{P} the set of primes):

$$\zeta(s) = \prod_{p \in \mathcal{P}} \left(1 + \frac{1}{p^s} + \dots + \frac{1}{p^{ms}} + \dots \right) = \prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{1}{p^s}}.$$

The zeta function is a symbolic expression associated to the distribution of primes, which is well known to be a very mysterious structure. But its fantastic strength as a tool comes from the fact that *it can be extended by analytic continuation to the complex plane*. First s can be extended to *reals* > 1 , secondly s can be extended to *complex* numbers s of real part $\Re(s) > 1$, and thirdly s can be extended by analytic continuation to a meromorphic function on the entire complex plane \mathbb{C} .

10.2 Mellin transform and functional equation

The zeta function encodes very deep arithmetic properties. Riemann proved in his celebrated 1859 paper “Über die Anzahl der Primzahlen unter einer gegebenen Grösse” (“On the number of prime numbers less than a given quantity”) that it manifests beautiful properties of symmetry. This can be made explicit noting that $\zeta(s)$ is related by a transformation called the *Mellin transform* to the *theta function* which possesses beautiful properties of automorphy, where “automorphy” means invariance of a function $f(\tau)$ defined on the Poincaré plane \mathcal{H} (complex numbers τ of positive imaginary part $\Im(\tau)$) relatively to a countable subgroup of the group acting on \mathcal{H} by homographies (also called Möbius transformations) $\gamma(\tau) = \frac{a\tau+b}{c\tau+d}$.

The theta function $\Theta(\tau)$ is defined on the half plane \mathcal{H} as the series

$$\Theta(\tau) = \sum_{n \in \mathbb{Z}} e^{in^2\pi\tau} = 1 + 2 \sum_{n \geq 1} e^{in^2\pi\tau}$$

$\Im(\tau) > 0$ is necessary to warrant the convergence of $e^{-n^2\pi\Im(\tau)}$. We will see later that $\Theta(\tau)$ is what is called a *modular form* of level 2 and weight $\frac{1}{2}$. Its automorphic symmetries are

1. Symmetry under translation: $\Theta(\tau+2) = \Theta(\tau)$ (level 2, trivial since $e^{2i\pi} = 1$ implies $e^{in^2\pi(\tau+2)} = e^{in^2\pi\tau}$).
2. Symmetry under inversion: $\Theta\left(\frac{-1}{\tau}\right) = \left(\frac{\tau}{i}\right)^{\frac{1}{2}} \Theta(\tau)$ (weight $\frac{1}{2}$, proof from Poisson formula).

If $f : \mathbb{R}^+ \rightarrow \mathbb{C}$ is a complex valued function defined on the positive reals, its *Mellin transform* $g(s)$ is defined by the formula:

$$g(s) = \int_{\mathbb{R}^+} f(t)t^s \frac{dt}{t}$$

Let us compute the Mellin transform of $\Theta(it)$ or more precisely, using the formula $\Theta(\tau) = 1 + 2\tilde{\Theta}(\tau)$, of $\tilde{\Theta}(it) = \frac{1}{2}(\Theta(it) - 1)$:

$$\Lambda(s) = \frac{1}{2}g\left(\frac{s}{2}\right) = \frac{1}{2} \int_0^\infty (\Theta(it) - 1) t^{\frac{s}{2}} \frac{dt}{t} = \sum_{n \geq 1} \int_0^\infty e^{-n^2\pi t} t^{\frac{s}{2}} \frac{dt}{t}$$

In each integral we make the change of variable $x = n^2\pi t$. The integral becomes:

$$n^{-s}\pi^{-\frac{s}{2}} \int_0^\infty e^{-x} x^{\frac{s}{2}-1} dx$$

But $\int_0^\infty e^{-x} x^{\frac{s}{2}-1} dx = \Gamma\left(\frac{s}{2}\right)$ where $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$ is the *gamma function*, and therefore

$$\Lambda(s) = \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \left(\sum_{n \geq 1} \frac{1}{n^s}\right) = \zeta(s) \Gamma\left(\frac{s}{2}\right) \pi^{-\frac{s}{2}}$$

This remarkable expression enables to use the automorphic symmetries of the theta function to derive a *functional equation* satisfied by the lambda function, and therefore by the zeta function. Indeed, let us write $\Lambda(s) = \int_0^\infty = \int_0^1 + \int_1^\infty$ and use the change of variable $t = \frac{1}{u}$ in the first integral. Since $\frac{i}{u} = -\frac{1}{iu}$ and

$$\Theta\left(\frac{i}{u}\right) = \Theta\left(-\frac{1}{iu}\right) = \left(\frac{iu}{i}\right)^{\frac{1}{2}} \Theta(iu) = u^{\frac{1}{2}} \Theta(iu)$$

due to the symmetry of Θ under inversion, we verify that the \int_0^1 part of $\Lambda(s)$ is equal to the \int_1^∞ part of $\Lambda(1-s)$ and vice-versa and therefore the lambda function satisfies the functional equation

$$\Lambda(s) = \Lambda(1-s)$$

As $\zeta(s)$ is well defined for $\Re(s) > 1$, it is also well defined, via the functional equation of Λ , for $\Re(s) < 0$, the difference between the two domains coming from the difference of behavior of the gamma function Γ .

We can easily extend $\zeta(s)$ to the domain $\Re(s) > 0$ using the fact that $\zeta(s)$ has a pole of order 1 at $s = 1$ and computing $\zeta(s)$ as

$$\zeta(s) = \frac{1}{s-1} + \dots$$

$\Lambda(s)$ being now defined on the half plane $\Re(s) > 0$, the functional equation can be interpreted as a symmetry relative to the line $\Re(s) = \frac{1}{2}$, hence the major role of this line which is called the *critical line* of $\zeta(s)$.

10.3 Zeroes of Zeta, distribution of primes, and Riemann hypothesis

Riemann zeta function is one of the most beautiful objects in mathematics. Since Euler's time, an impressive amount of computations have been performed by the greatest mathematicians and a universe of relations with other functions has been discovered.

Due to the functional equation

$$\zeta(s) \Gamma\left(\frac{s}{2}\right) \pi^{-\frac{s}{2}} = \zeta(1-s) \Gamma\left(\frac{1-s}{2}\right) \pi^{-\frac{1-s}{2}},$$

the behavior of $\zeta(s)$ depends upon that of the gamma function $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$ which extended the factorial function $\Gamma(n) = (n-1)!$. $\Gamma(s)$ has no zeroes but has poles exactly on negative integers $-k$ ($k \geq 0$) where it has residue $\frac{(-1)^k}{k!}$.

For $s = -2k$ with $k > 1$, the functional equation reads

$$\zeta(-2k)\Gamma(-k)\pi^k = \zeta(1+2k)\Gamma\left(\frac{1+2k}{2}\right)\pi^{-\frac{1+2k}{2}}$$

and as the rhs is finite (the only pole of $\zeta(s)$ is $s = 1$) while $\Gamma(-k)$ is a pole, we must have $\zeta(-2k) = 0$. These are called the *trivial zeroes* of the zeta function.

The main interest of $\zeta(s)$ is to have *non trivial zeroes which encode the distribution of primes* in the following sense. For x a positive real, let $\pi(x)$ be the number of primes $p \leq x$. From Gauss (1792, 15 years old) and Legendre (1808) to Hadamard (1896) and De La Vallée Poussin (1896) it has been proved the asymptotic formula called the *prime number theorem*:

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

A better approximation, due to Gauss (1849), is $\pi(x) \sim \text{Li}(x)$ where the logarithmic integral is $\text{Li}(x) = \int_2^x \frac{dx}{\log(x)}$.³

The prime number theorem is a consequence of the fact that $\zeta(s)$ has no zeroes on the line $1 + it$ (recall that 1 is the pole of $\zeta(s)$). It as been improved with better approximations by many great arithmeticians.

In his 1859 paper, Riemann proved the fantastic result that $\pi(x)$ can be computed as the sum of a series whose terms are indexed by the non trivial zeroes of $\zeta(s)$.

It can be proved easily that all the non trivial zeroes of $\zeta(s)$ must lie inside the critical strip $0 < \Re(s) < 1$. Due to the functional equation they are symmetric relatively to the critical line and it is known that there exist an infinity of zeroes on the critical line and that the zeroes “concentrate” in a precise sense on the critical line. An enormous amount of computations from Riemann time to actual supercomputers (ZetaGrid: more than 10^{12} zeroes in 2005) via Gram, Backlund, Titchmarsh, Turing, Lehmer, Lehman, Brent, van de Lune, Wedeniwski, Odlyzko, Gourdon, and others, shows that all computed zeroes lie on the critical line $\Re(s) = \frac{1}{2}$.

The celebrated *Riemann hypothesis*, one of the deepest unsolved problem (8th Hilbert problem), claims that in fact they all lie on the critical

³For small n , $\pi(x) < \text{Li}(x)$, but Littelwood proved in 1914 that the inequality reverses an infinite number of times.

line. It is equivalent to the conjecture that for some constant c

$$|\text{Li}(x) - \pi(x)| \leq c\sqrt{x} \log(x)$$

that is

$$\pi(x) = \int_2^x \frac{dx}{\log(x)} + O(\sqrt{x} \log(x))$$

(recall that the prime number theorem is equivalent to the fact that no non trivial zeroes lie on the line $\Re(s) = 1$ limiting the critical strip).

10.4 Dirichlet series and Dirichlet L -functions

For many reasons, *generalized* zeta functions are important. They have the general form

$$\sum_{n \geq 1} \frac{a_n}{n^s}$$

and under some conditions on the a_n can be factorized in Euler products

$$\prod_{p \in \mathcal{P}} \left(1 + \frac{a_p}{p^s} + \dots + \frac{a_{p^m}}{p^{ms}} + \dots \right)$$

1. The condition is of course that the coefficients a_n are *multiplicative* in the sense that $a_1 = 1$ and, if $n = \prod p_i^{r_i}$, $a_n = \prod a_{p_i^{r_i}}$.
2. Moreover if the a_n are *strictly multiplicative* in the sense that $a_{p^m} = (a_p)^m$ then the series can be factorized in a *first degree* (or linear) Euler product

$$\prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{a_p}{p^s}}.$$

3. If $a_1 = 1$ and if for every prime p there exists an integer d_p s.t.

$$a_{p^m} = a_p a_{p^{m-1}} + d_p a_{p^{m-2}}$$

then the series can be factorized in a *second degree* (or quadratic) Euler product

$$\prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{a_p}{p^s} - \frac{d_p}{p^{2s}}}$$

The most important examples of Dirichlet series are given by Dirichlet L -functions where the a_n are the values $\chi(n)$ of a *character* mod m , that is of a multiplicative morphism

$$\chi : (\mathbb{Z}/m\mathbb{Z})^* \rightarrow \mathbb{C}$$

$$L_\chi = \sum_{n \geq 1} \frac{\chi(n)}{n^s}$$

As χ is multiplicative, the a_n are strictly multiplicative and the series can be factorized in a *first degree* Euler product. The theory of the zeta function can be straightforwardly generalized (theta function, automorphy symmetries, lambda function, functional equation).

11 Modular forms

11.1 Two classes of L -functions

We have just seen that L -functions such as Riemann zeta function encode in a very subtle way deep arithmetic informations. We will now see that we meet naturally *two classes* of L -functions, those associated to elliptic curves and those associated to what are called *modular forms*. A great discovery of Shimura has been that in the case of *modular elliptic curves*, the two L -functions are *equal*. The Taniyama-Shimura-Weil conjecture that every elliptic curve over \mathbb{Q} is modular says therefore that the two classes are identical. It is a conjecture on the equivalence between two completely different ways of constructing objects of a certain type (L -functions). Its deepness has been very well formulated by Anthony Knapp who explained that XXth century mathematics discovered

“a remarkable connection between automorphy and arithmetic algebraic geometry. This connection first shows up in the coincidence of L -functions that arise from some very special modular forms (“automorphic” L -functions) with L -functions that arise from number theory (“arithmetic” or “geometric” L -functions, also called “motivic”).”

“The automorphic L -functions have manageable analytic properties, while the arithmetic L -functions encode subtle number-theoretic information. The fact that the arithmetic L -functions are automorphic enables one to bring a great deal of mathematics to bear on extracting the number-theoretic information from the L -function.”

“Automorphic L -functions have more manageable analytic properties, but they initially have little to do with algebraic number theory or algebraic geometry. The fundamental objective is to prove that motivic L -function are automorphic.”

M. Ram Murty also emphasized the point:

“In its comprehensive form, an identity between an automorphic L -function and a “motivic” L -function is called a reciprocity law. (...) The conjecture of Shimura-Taniyama that every elliptic curve over \mathbb{Q} is “modular” is certainly the most intriguing reciprocity law of our time. The “Himalayan peaks” that hold the secrets of this non abelian reciprocity law challenge humanity.”

11.2 Modular functions for $SL(2, \mathbb{Z})$

We start from the representation of elliptic curves as complex tori $E = \mathbb{C}/\Lambda$ with Λ a lattice $\{m\omega_1 + n\omega_2\}_{m,n \in \mathbb{Z}}$ in \mathbb{C} with \mathbb{Z} -basis $\{\omega_1, \omega_2\}$. If $\tau = \omega_2/\omega_1$, we can suppose $\text{Im}(\tau) > 0$, that is $\tau \in \mathcal{H}$ where \mathcal{H} is the Poincaré upper half complex plane. To correlate *univocally* E and its “module” τ we must look at the transformation of τ when we change the \mathbb{Z} -basis of Λ . Let $\{\omega'_2, \omega'_1\}$ another \mathbb{Z} -basis. We have $\begin{pmatrix} \omega'_2 \\ \omega'_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_2 \\ \omega_1 \end{pmatrix}$ with $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ an integral matrix. But γ must be invertible and its inverse must therefore be also an integral matrix, so $\gamma \in SL(2, \mathbb{Z})$. γ acts on τ via Möbius transformations:

$$\gamma(\tau) = \frac{a\tau + b}{c\tau + d}$$

The concept of modular form arises naturally when we consider *holomorphic $SL(2, \mathbb{Z})$ -invariant differentials* on the Poincaré half-plane \mathcal{H} . Let $f(\tau)d\tau$ be a 1-form on \mathcal{H} with f an holomorphic function and consider $f(\tau')d\tau'$ with $\tau' = \gamma(\tau)$. We have

$$\begin{aligned} f(\tau')d\tau' &= f\left(\frac{a\tau + b}{c\tau + d}\right) \frac{(c\tau + d)a - (a\tau + b)c}{(c\tau + d)^2} d\tau \\ &= f\left(\frac{a\tau + b}{c\tau + d}\right) \frac{1}{(c\tau + d)^2} d\tau \text{ since } ad - bc = 1 \end{aligned}$$

We see that in order to get the invariance $f(\tau)d\tau = f(\tau')d\tau'$ we need $f\left(\frac{a\tau + b}{c\tau + d}\right) \frac{1}{(c\tau + d)^2} = f(\tau)$, i.e. $f(\gamma(\tau)) = (c\tau + d)^2 f(\tau)$. Hence the general definition:

Definition. An holomorphic function on \mathcal{H} is a *modular function of weight k* if $f(\gamma(\tau)) = (c\tau + d)^k f(\tau)$ for every $\gamma \in SL(2, \mathbb{Z})$.

We note that the definition implies $f = 0$ for *odd* weights since $-I \in SL(2, \mathbb{Z})$ and if k is odd

$$f(-I\tau) = f\left(\frac{-\tau}{-1}\right) = f(\tau) = (-1)^k f(\tau) = -f(\tau)$$

The weight 0 means that f is $SL(2, \mathbb{Z})$ -invariant.

A modular function of weight k can also be interpreted as an homogeneous holomorphic function of degree $-k$ defined on the lattices Λ . If we define $\tilde{f}(\Lambda)$ by $\tilde{f}(\Lambda) = \omega_1^{-k} f(\tau)$ we see that for f to be modular of weight k is equivalent to $\tilde{f}(\alpha\Lambda) = \alpha^{-k} \tilde{f}(\Lambda)$.

To be modular f has only to be modular on generators of $SL(2, \mathbb{Z})$, two generators being the translation $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ acting by $\tau \rightarrow \tau + 1$ and the inversion $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ acting by $\tau \rightarrow -1/\tau$. Therefore f is modular of weight k iff

$$\begin{cases} f(\tau + 1) = f(\tau) \\ f(-\frac{1}{\tau}) = (-\tau)^k f(\tau) \end{cases}$$

We already met modular functions in the theory of elliptic curves: Eisenstein series G_{2k} of weight $2k$, the *elliptic invariants* which are the coefficients g_2 of weight 4 and g_3 of weight 6 of the Weierstrass equation associated to a complex torus, the discriminant $\Delta = (g_2)^3 - 27(g_3)^2$ of weight 12, the modular invariant j of weight 0.

11.3 Fourier expansion at infinity, modular forms, and cusp (parabolic) forms

The fact that a modular function f is invariant by the translation $\tau \rightarrow \tau + 1$ means that it is *periodic* of period 1 and therefore can be expanded in a *Fourier series*

$$f(\tau) = \sum_{n \in \mathbb{Z}} c_n e^{2i\pi n\tau} = \sum_{n \in \mathbb{Z}} c_n q^n \text{ with } q = e^{2i\pi\tau}$$

The variable $q = e^{2i\pi\tau}$ uniformizes f at infinity. It is called the *nome*.

If we use this representation, the property of modularity $f(-\frac{1}{\tau}) = (-\tau)^k f(\tau)$ imposes very strict constraints on the Fourier coefficients c_n and therefore modular functions generate very special series $\{c_n\}_{n \in \mathbb{Z}}$.

For controlling the holomorphy of f at infinity one introduces two restrictions for the general concept of a modular function of weight k .

Definition. f is called a modular *form* of weight k if f is *holomorphic at infinity*, that is if its Fourier coefficients $c_n = 0$ for $n < 0$.

Definition. Moreover f is called a *parabolic* modular form, or a *cusp form*, if f vanishes at infinity, that is if $c_0 = 0$ (then $c_n = 0$ for $n \leq 0$).

It is traditional to note M_k the space of modular forms of weight k , and $S_k \subset M_k$ the space of cusp forms of weight k . Eisenstein series

$$G_k(\tau) = \sum_{(m,n) \in \mathbb{Z} \times \mathbb{Z} - \{0,0\}} \frac{1}{(m\tau + n)^k}$$

(the power k must be even ($k = 2r$) for if k is odd the $(-m, -n)$ and (m, n) terms cancel) are modular forms. The discriminant Δ of elliptic curves, $\Delta(\tau) = (g_2(\tau))^3 - 27(g_3(\tau))^2$ with $g_2(\tau) = 60G_4(\tau)$ and $g_3(\tau) = 140G_6(\tau)$, is a modular function of weight 12 which expands as

$$\Delta(\tau) = q - 24q^2 + 252q^3 - 1472q^4 + \dots$$

One can show that it is given by the infinite product

$$\Delta(\tau) = q \prod_{r=1}^{r=\infty} (1 - q^r)^{24}$$

It is therefore a *cusp form* $\Delta \in S_{12}$. We note that $\Delta(\tau) = 0$ exactly for $q^r = 1$, that is $e^{2i\pi r\tau} = 1$, that is $r\tau \in \mathbb{Z}$, that is $\tau \in \mathbb{Q}$, that is for the rational points on the boundary of \mathcal{H} (which are called cusp points). $\Delta(\tau)$ vanishes nowhere on \mathcal{H} .

On the contrary, the modular invariant j of weight 0 expands as

$$j(\tau) = \frac{1}{q} + 744 + 196\,884q + 21\,493\,760q^2 + \dots$$

It has a pole at infinity and fails to be a modular form.

The fundamental importance of the Eisenstein series and the discriminant is that they enables to determine the spaces M_k and S_k .⁴

1. $M_0 \simeq \mathbb{C}$ since an f which is $SL(2, \mathbb{Z})$ -invariant and holomorphic on \mathcal{H} and at infinity is holomorphic on the quotient $(\mathcal{H}/SL(2, \mathbb{Z})) \cup \{\infty\}$ which is compact. f is therefore constant by Liouville theorem.
2. $M_k = 0$ for $k < 0$ since if $f \neq 0 \in M_k$, then f^{12} is of weight $12k$, Δ^{-k} is of weight $-12k$, and $f^{12}\Delta^{-k} \in M_0$ but is without constant term. Therefore $f = 0$.
3. $M_k = 0$ for k odd since, if we take $\gamma = -I$, $f(\gamma(\tau)) = f(\tau) = -f(\tau)$, and $f \equiv 0$.
4. $M_k = 0$ for $k = 2$.
5. For k even $k > 2$, $M_k = \mathbb{C}G_k \oplus S_k$ since S_k is of codimension 1 in M_k and G_k has a constant term.

⁴We will see later that they are eigenvectors of the Hecke operators defined on the spaces M_k and S_k .

6. $S_k \simeq M_{k-12}$. Indeed if $f \in S_k$, $f/\Delta \in M_{k-12}$. Since $\Delta \neq 0$, f/Δ (which is of weight $k-12$) is holomorphic on \mathcal{H} and, as $c_n = 0$ for $n \leq 0$ for f and Δ , $c_n = 0$ for $n < 0$ for f/Δ and $f/\Delta \in M_{k-12}$. Reciprocally, if $g \in M_{k-12}$ then $g\Delta \in S_k$. $S_k \simeq M_{k-12}$ implies, via (2), $\dim(S_k) = 0$ for $k < 12$ and, via (5), $\dim(M_k) = 1$ for $k < 12$.

It is therefore easy to compute the dimension of M_k : e.g. for $k = 12$, via (6) and (1), $\dim(S_k) = \dim(M_0) = 1$ and, via (5), $\dim(M_k) = 2$.

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\dim(M_k)$	1	0	0	0	1	0	1	0	1	0	1	0	2	0	1	0

Such dimensions imply a lot of deep arithmetical relations because every time we can associate to d situations d modular forms of M_k and we have $d > \dim(M_k)$, then, as was emphasized by Don Zagier,

“We get “for free” information – often highly non trivial – relating these different situations.” (p. 240)

Moreover we will see that the M_k are spanned by modular forms whose Fourier series have *rational* coefficients c_n . As D. Zagier also explains:

“It is this phenomenon which is responsible for the richness of the arithmetic applications of the theory of modular forms.”

We have seen that $\Delta(\tau) = (60G_4(\tau))^3 - 27(140G_6(\tau))^2$. It is a general fundamental fact:

Theorem. Every modular form can be expressed in a unique way as a *polynomial* in G_4 and G_6 .

11.4 L -functions of cusp forms

If f is a cusp form of weight k , $f \in S_k$, then $f(\tau) = \sum_{n \geq 1} c_n q^n$ with the nome $q = e^{2i\pi\tau}$. We associate to f the L -function:

$$L_f(s) = \sum_{n \geq 1} \frac{c_n}{n^s}$$

having the same coefficients. These L -functions encode a lot of arithmetical information. They come essentially as *Mellin transform* of their generating cusp form.

Paralleling the case of Riemann ζ function for which the function

$$\Lambda(s) = \zeta(s) \Gamma\left(\frac{s}{2}\right) \pi^{-\frac{s}{2}}$$

was the Mellin transform of the theta function $\frac{1}{2}(\Theta(it) - 1)$, we introduce the Mellin transform

$$\Lambda_f(s) = \int_0^\infty f(it) t^s \frac{ds}{s}$$

of the cusp form f on the positive imaginary axis and we compute

$$\Lambda_f(s) = \frac{1}{(2\pi)^s} \Gamma(s) L_f(s)$$

The modular invariance of f and its good behavior at infinity imply that the c_n are bounded in norm by $n^{k/2}$ and therefore $L_f(s)$ is absolutely convergent in the half-plane $\Re(s) > \frac{k}{2} + 1$.

As the Riemann ζ function, the L -functions $L_f(s)$ satisfy a *functional equation*. It is the content of a deep theorem due to Hecke:

Hecke theorem. $L_f(s)$ and $\Lambda_f(s)$ are *entire* functions and $\Lambda_f(s)$ satisfies the functional equation

$$\Lambda_f(s) = (-1)^{k/2} \Lambda_f(k - s)$$

11.5 Hecke operators for $SL(2, \mathbb{Z})$

We have just seen that the L -functions $L_f(s)$ of modular forms have a behavior very similar to that of Riemann zeta function: they are Mellin transforms of functions with $SL(2, \mathbb{Z})$ symmetries, and they satisfy a functional equation. But up to now, there exists a major difference: the zeta function $\zeta(s)$ is not only a series $\sum_{n \geq 1} \frac{1}{n^s}$, it is also an *Euler product* $\prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{1}{p^s}}$ and in that sense encodes information *prime by prime*. This property is so important that it is natural to ask if and how it can be generalized to $L_f(s)$.

The problem is quite difficult. Hecke's very beautiful idea was to solve it in two steps:

1. find linear operators on the vector spaces M_k of modular forms which satisfy the relations of an Euler product;
2. look at their eigenfunctions.

The simplest way of defining Hecke operator is to start with the free group \mathcal{L} generated by the lattices Λ of \mathbb{C} (recall that it is the origin of the $SL(2, \mathbb{Z})$ action on the Poincaré half-plane \mathcal{H}). If we consider a lattice Λ and magnify it in the sublattice $n\Lambda$, there will exist *intermediary* lattices Λ' s.t. $n\Lambda \subseteq \Lambda' \subseteq \Lambda$. In that case the larger torus $\mathbb{C}/n\Lambda$ projects onto

the smaller one \mathbb{C}/Λ' . We write $[\Lambda : \Lambda'] = n$. If $\{\omega'_1, \omega'_2\}$ and $\{\omega_1, \omega_2\}$ are respective \mathbb{Z} -basis of Λ' and Λ we have

$$\begin{pmatrix} \omega'_2 \\ \omega'_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_2 \\ \omega_1 \end{pmatrix}$$

with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(n)$ the set of integral matrices with determinant n .

$SL(2, \mathbb{Z})$ acts on $M(n)$ and decomposes it in orbits. We can choose as representing elements for instance the matrices $\alpha_i = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $ad = n$, $d \geq 1$, $0 \leq b < d$. There are $\nu(n) = \sigma_1(n) = \sum_{d|n} d$ of them and we have

$$M(n) = \bigcup_{i=1}^{i=\nu(n)} SL(2, \mathbb{Z}) \alpha_i$$

Hecke operators construct the sum of such Λ' :

Definition. The Hecke operator $T(n) : \mathcal{L} \rightarrow \mathcal{L}$ is the additive operator associating to any lattice Λ the sum of the lattices Λ' s.t. $[\Lambda : \Lambda'] = n$:

$$\begin{aligned} T(n) : \mathcal{L} &\rightarrow \mathcal{L} \\ \Lambda &\mapsto T(n)(\Lambda) = \sum_{[\Lambda:\Lambda']=n} \Lambda' \end{aligned}$$

We have of course

$$T(n) = \sum_{i=1}^{i=\nu(n)} \alpha_i(\Lambda)$$

It is easy to extend the definition of Hecke operators to modular forms. Let us first consider homogeneous functions \tilde{f} of degree $-k$ on the Λ : $\tilde{f}(\alpha\Lambda) = \alpha^{-k} \tilde{f}(\Lambda)$. We define

$$T_k(n) \left(\tilde{f}(\Lambda) \right) = n^{k-1} \sum_{[\Lambda:\Lambda']=n} \tilde{f}(\Lambda')$$

the factor n^{k-1} coming from homogeneity.

Modular functions $f(\tau)$ are related to $\tilde{f}(\Lambda)$ by $f(\tau) = \tilde{f}(\Lambda_\tau) = (\omega_1)^k \tilde{f}(\Lambda)$. Computations yield for the action of Hecke operators on modular forms $f(\tau) \in M_k$, the following explicit formulae:

Proposition. Let $f(\tau) \in M_k$, $f(\tau) = \sum_{n \geq 0} c_n q^n$, be a modular form of weight k . Then $T_k(m)(f(\tau)) \in M_k$, $T_k(m)(f(\tau)) = \sum_{n \geq 0} b_n q^n$ with

$$\begin{cases} b_0 = c_0 \sigma_{k-1}(m) \text{ where } \sigma_j(m) = \sum_{d|m} d^j \\ b_1 = c_m \\ b_n = \sum_{a|(n,m)} a^{k-1} c_{\frac{nm}{a^2}} \text{ for } n > 1 \end{cases}$$

This result shows first that $c_0 = 0 \Rightarrow b_0 = 0$ and therefore if $f(\tau) \in S_k$, $T_k(m)(f(\tau)) \in S_k$. On the other hand, if $m = p$ is prime,

$$\begin{cases} b_0 = c_0 \\ b_1 = c_m \\ b_n = \sum_{a|(n,p)} a^{k-1} c_{\frac{np}{a^2}} = c_{np} \text{ (only the term } a = 1 \text{) for } n > 1 \text{ if } p \nmid n \\ b_n = \sum_{a|(n,p)} a^{k-1} c_{\frac{np}{a^2}} = c_{np} + p^{k-1} c_{\frac{n}{p}} \text{ (the terms } a = 1, a = p \text{) for } n > 1 \text{ if } p \mid n \end{cases}$$

Hecke theorem. On M_k the $T_k(m)$ constitute a commutative algebra \mathcal{T}_k generated by the $T_k(p)$ and we have the product formulae

$$\begin{cases} T_k(p^r)T_k(p) = T_k(p^{r+1}) + p^{k-1}T_k(p^{r-1}) \\ T_k(m)T_k(n) = \sum_{a|(n,m)} a^{k-1}T_k\left(\frac{mn}{a^2}\right) \\ T_k(m)T_k(n) = T_k(mn) \text{ if } (m, n) = 1 \end{cases}$$

But these are precisely the equivalent for operators of the *multiplicative formulae for quadratic Euler products*: $a_p^r a_p = a_p^{r+1} - d_p a_p^{r-1}$.

Now Hecke's key idea is to look at *simultaneous eigenvectors* of the $T_k(m)$, which exist since the algebra \mathcal{T}_k is commutative. These very particular modular forms inherit very particular properties from those of Hecke operators. Their coefficients c_n are *algebraic integers* and satisfy the multiplicative relation $c_{nm} = c_n c_m$ if $(m, n) = 1$, the Dirichlet L -function $L_f(s) = \sum_{n \geq 1} \frac{c_n}{n^s}$ can be expressed as an Euler product, possesses an analytic continuation on \mathcal{H} and satisfies a functional equation.

11.6 Petersson scalar product and Euler product of cusp forms

We can be even more precise when we restrict Hecke operators to the space of *cusp forms* S_k . Let $\tau = \rho + i\sigma$. The measure $\frac{d\rho d\sigma}{\sigma^2}$ on \mathcal{H} is $SL(2, \mathbb{Z})$ -invariant and, if R is a fundamental domain of $SL(2, \mathbb{Z})$,

$$\langle f, g \rangle = \int_R f(\tau) \bar{h}(\tau) \sigma^k \frac{d\rho d\sigma}{\sigma^2}$$

is a scalar product, called Petersson product, on S_k .

Petersson theorem. On S_k the Hecke operators $T_k(n)$ are self-adjoint for the Petersson scalar product $\langle f, g \rangle$.

Petersson theorem implies that S_k possesses an *orthogonal* basis of *simultaneous* eigenvectors of the Hecke operators $T_k(n)$. Let $f(\tau) \in S_k$ be such a simultaneous eigenvector. For every n , $T_k(n)f(\tau) = \lambda(n)f(\tau)$. If $f(\tau) = \sum_{r \geq 1} c_r q^r$ and $T_k(n)f(\tau) = \sum_{r \geq 1} b_r q^r$, we have therefore

$b_r = \lambda(n)c_r$ for $r \geq 1$. But we have seen that $b_1 = c_n$. So $c_n = \lambda(n)c_1$ and $b_r = \lambda(n)c_r = \lambda(n)\lambda(r)c_1$. If we normalize $f(\tau)$ by setting $c_1 = 1$, we get $c_n = \lambda(n)$ and, as the $\lambda(n)$ are eigenvalues of the $T_k(n)$, *the multiplicative properties of Hecke operators become shared by the coefficients of the eigen cusp form $f(\tau)$* :

$$\begin{cases} c_{p^r}c_p = c_{p^{r+1}} + p^{k-1}c_{p^{r-1}} \\ c_m c_n = \sum_{a|(n,m)} a^{k-1} c_{\frac{mn}{a^2}} \\ c_m c_n = c_{mn} \text{ if } (m, n) = 1 \end{cases}$$

and these multiplicative properties imply immediately that the Dirichlet L -function $L_f(s)$ of f can be expressed by a *second order Euler product*:

$$L_f(s) = \prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{c_p}{p^s} + \frac{1}{p^{1-k+2s}}}$$

which is of the standard form

$$\prod_{p \in \mathcal{P}} \frac{1}{1 - \frac{a_p}{p^s} - \frac{d_p}{p^{2s}}}$$

with $a_p = c_p$ and $d_p = -p^{k-1}$. $L_f(s)$ converges for $\Re(s) > \frac{k}{2} + 1$ and has a single simple pole at $s = k$.

11.7 Fricke involution and generalization to the groups $\Gamma_0(N)$

All these results can be generalized to invariance groups *smaller* than $SL(2, \mathbb{Z})$. This corresponds to the introduction of the key concept of *level N* of a modular function, the classical ones being of level 1. The congruence subgroup $\Gamma_0(N)$ of $SL(2, \mathbb{Z})$ is defined by a restriction on the term c :

$$\begin{aligned} \Gamma_0(N) &= \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) : c \equiv 0 \pmod{N} \right\} \\ &= \left\{ \begin{pmatrix} a & b \\ kN & d \end{pmatrix} \in SL(2, \mathbb{Z}) \right\} \end{aligned}$$

We note that $\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$. Of course $\Gamma_0(1) = SL(2, \mathbb{Z})$. Let

$\Gamma_0(1) = \bigcup_j \beta_j \Gamma_0(N)$, $\beta_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix} \in SL(2, \mathbb{Z})$, be a decomposition of $\Gamma_0(1)$ in $\Gamma_0(N)$ -orbits. A fundamental domain R_N of $\Gamma_0(N)$ is $R_N = \bigcup_j \beta_j^{-1}(R)$ where R is a fundamental domain of $SL(2, \mathbb{Z})$,

$\left(\beta_j^{-1} = \begin{pmatrix} d_j & -b_j \\ -c_j & a_j \end{pmatrix}\right)$, and the cusps of R_N are the rational points of the boundary of \mathcal{H} image of the infinite point: $\beta_j^{-1}(\infty) = -\frac{d_j}{c_j} \in \mathbb{Q}$.

1. A modular function of weight k and level N is an $f(\tau)$ satisfying the invariance condition $f(\gamma(\tau)) = (c\tau + d)^k f(\tau) \forall \gamma \in \Gamma_0(N)$.
2. A modular function of weight k and level N is a modular form $f(\tau) \in M_k(N)$ if it is holomorphic not only at infinity but also at the cusps.
3. A modular form of weight k and level N is a cusp form $f(\tau) \in S_k(N)$ if moreover it vanishes at infinity and at the cusps.
4. If $f(\tau) \in M_k(N)$, $f(\tau)$ is N -periodic (since $\gamma = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$ and $f(\gamma(\tau)) = f(\tau + N) = f(\tau)$) and can be developed at infinity in a Fourier series $f(\gamma(\tau)) = \sum_{n \geq 0} c_n q^n$ with nome $q = e^{\frac{2i\pi\tau}{N}}$.

If $f(\tau) \in S_k(N)$ we can associate to it by Mellin transform a Dirichlet L -function $L_f(s)$. But we must be careful since for $N > 1$ the inversion $\tau \rightarrow -\frac{1}{\tau}$ of matrix $\alpha = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is no longer in $\Gamma_0(N)$. But we can use the transformation $\tau \rightarrow -\frac{1}{N\tau}$ and the operator $w_N(f(\tau)) = N^{-\frac{k}{2}} \tau^{-k} f\left(-\frac{1}{N\tau}\right)$ which leaves stable $M_k(N)$ and $S_k(N)$. As w_N is an *involution*, called Fricke involution, the spaces $M_k(N)$ and $S_k(N)$ split into eigenspaces $M_k^\pm(N)$ and $S_k^\pm(N)$ of the eigenvalues ± 1 . Then, Hecke theorem can be generalized to eigenvectors of the Fricke involution:

Hecke theorem. If $f(\tau) \in S_k^\pm(N)$, its L -function is an entire function and $\Lambda_f(s) = N^{\frac{s}{2}} \frac{1}{(2\pi)^s} \Gamma(s) L_f(s)$ satisfies the functional equation

$$\Lambda_f(s) = \pm(-1)^{k/2} \Lambda_f(k-s)$$

So we have weakened the concept of cusp form in imposing less symmetries, but at the same time we have strengthened it in imposing its vanishing at its cusps and its “parity” relative to Fricke involution w_N . We want now to generalize also Hecke operators and Euler products. The problem is rather subtle since N has prime factors $p \mid N$ and we cannot control easily the relation between w_N and the Hecke operators $T_k(p)$ for $p \mid N$.

11.8 Hecke operators for $\Gamma_0(N)$ and Euler products

For the expansions at infinity, we find essentially the same formulae as before. We get also the same multiplicative recurrence formulae for $p \nmid N$, but for $p \mid N$ we get another formula which is purely multiplicative:

Proposition. If $p \mid N$, $T_k(p^r) = T_k(p)^r$.

Hence the generalization of Hecke theorem:

Generalized Hecke theorem. On $M_k(N)$ the $T_k(m)$ constitute a commutative algebra \mathcal{T}_k generated by the $T_k(p)$ and

$$\begin{cases} T_k(p^r)T_k(p) = T_k(p^{r+1}) + p^{k-1}T_k(p^{r-1}) & \text{if } p \nmid N \\ T_k(p^r) = T_k(p)^r & \text{if } p \mid N \\ T_k(m)T_k(n) = \sum_{a \mid (n,m)} a^{k-1}T_k\left(\frac{mn}{a^2}\right) \\ T_k(m)T_k(n) = T_k(mn) & \text{if } (m, n) = 1 \end{cases}$$

Petersson's theorem can also be generalized, the scalar product being defined now by integration on a fundamental domain R_N of $\Gamma_0(N)$.

Petersson theorem. Hecke operator $T_k(n)$ is self-adjoint on $S_k(N)$ if $(n, N) = 1$.

Let $f(\tau) \in S_k(N)$ be a cusp form of weight k and level N which is a common eigenvector of *all* the $T_k(n)$. Due to Hecke theorem, its Dirichlet L -function $L_f(s)$ can be expressed by a second order Euler product but with a first order part corresponding to the primes p dividing N :

$$L_f(s) = \prod_{\substack{p \in \mathcal{P} \\ p \mid N}} \frac{1}{1 - \frac{c_p}{p^s}} \prod_{\substack{p \in \mathcal{P} \\ p \nmid N}} \frac{1}{1 - \frac{c_p}{p^s} + \frac{1}{p^{1-k+2s}}}$$

$L_f(s)$ converges for $\Re(s) > \frac{k}{2} + 1$ and has a single simple pole at $s = k$.

11.9 New forms and Atkin-Lehner theorem

As we have already noticed, the main difficulty encountered with these generalizations of the case $SL(2, \mathbb{Z})$ to the case $\Gamma_0(N)$ concerns the control of the relation between w_N and $T_k(p)$ when $p \mid N$. It has been solved by Atkin and Lehner with the concept of *newform*. Among the cusp forms of level N , some come from a cusp form of sublevel N/r . They are called "old" forms. $S_k(N)$ is the orthogonal sum of the subspaces of old and new (i.e. non old) forms: $S_k(N) = S_k^{\text{old}}(N) \oplus S_k^{\text{new}}(N)$.

If $f(\tau) \in S_k^{\text{new}}(N)$ is a *new* form, everything is fine: $f(\tau)$ possesses at the same time an Euler product and a functional equation.

12 L -functions of elliptic curves and Eichler-Shimura theory

12.1 Encoding geometrico-arithmetic information into L -functions

ζ and L -functions of eigenforms encode a lot of deep arithmetic information. As was emphasized by Anthony Knapp,

“This is part of a general pattern in algebraic number theory and algebraic geometry, that L functions are used to *encode information prime by prime* and that properties of these L functions are expected to yield deep insights into the original problem being studied.”

We have just seen how to associate, via a Mellin transformation, to each eigen cusp form an L -function satisfying a functional equation and expressible as an Euler product by means of Hecke operators.

We will now see how to associate to an elliptic curve E defined over \mathbb{Q} a L -function L_E which counts the number of integer points of $E \bmod \mathbb{Z}$. Such L_E

“*encode geometric information*, and deep properties of the elliptic curve come out (partly conjecturally) as a consequence of properties of these functions”.(Knapp)

And as for the zeta function:

“It is expected that *deep arithmetic information is encoded* in the behavior of $L(s, E)$ *beyond the region of convergence*”.

It is then very natural to try to compare these two kinds of L -functions. The situation is perfectly described by A. Knapp:

“We have two kinds of L functions, the kind from cusp forms that we understand very well and the kind from elliptic curves that contains a great deal of information.”

Of course it would be a “miracle” that two L functions belonging to these two completely different classes will be the same. But it is precisely such an astonishing result which was proved by Eichler et Shimura for a set of cusp forms which can be *parametrized* by what are called modular curves $X_0(N)$ and are called for that *modular elliptic curves*. As Knapp explains:

“Two miracles occur in this construction. The first miracle is that $X_0(N)$, E , and the mapping can be defined compatibly over \mathbb{Q} . (...) The second miracle is that the L function of E matches the L function of the cusp form f .”

12.2 The L -function of an elliptic curve E

Let E be an elliptic curve defined over \mathbb{Q} . It has an infinity of points over \mathbb{C} (but can have no points on \mathbb{Q}). But if we reduce $E \bmod p$, its reduction E_p will necessarily have a finite number of points $N_p = \#E_p(\mathbb{F}_p)$ over the finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. The most evident arithmetic data on E consists therefore in combining these local data N_p depending on primes p in an *Euler product*. We must be cautious here since for p dividing the discriminant Δ of E , the reduction is “bad”, i.e. E_p is singular.

The numbers N_p can be computed using the *Frobenius morphism* φ of degree p which fixes the points of E modulo p that is the points of $E_p(\mathbb{F}_p)$. The Frobenius morphism is defined on the algebraic closure $\overline{\mathbb{F}}_p$ of \mathbb{F}_p as $\text{Frob}_p : x \rightarrow x^p$; it is the generator of the Galois group $\text{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$. It acts on E by acting on the coordinates of the points of E .

We have the equivalence

$$x \in E_p(\mathbb{F}_p) \Leftrightarrow x \in \text{Ker}(1 - \varphi)$$

and as $\#\text{Ker}(1 - \varphi) = \deg(1 - \varphi)$ we have

$$N_p = \#E_p(\mathbb{F}_p) = \deg(1 - \varphi)$$

For technical reasons on which we will return, it is better to use the difference $a_p = p + 1 - N_p$. The good choice of an Euler product is the following which defines the L -function $L_E(s)$ of the elliptic curve E :

$$L_E(s) = \prod_{p|\Delta} \frac{1}{1 - \frac{a_p}{p^s}} \prod_{p \nmid \Delta} \frac{1}{1 - \frac{a_p}{p^s} + \frac{1}{p^{1-2s}}}$$

We note the similarity with the L -function of a modular form of level Δ and weight $k = 0$.

As $1 \leq N_p \leq 2p + 1$ (we count the point at infinity), $|a_p| \leq p$, and $L_E(s)$ converges for $\Re(s) > 2$. In fact a theorem due to Hasse asserts that $|a_p| \leq 2\sqrt{p}$ and therefore $L_E(s)$ converges for $\Re(s) > 3/2$.

12.3 The modular curve $X_0(N)$ and its Jacobian $J_0(N)$

Let $\overline{\mathcal{H}}$ be the completion of \mathcal{H} at infinity and at the cusps, $\overline{\mathcal{H}} = \mathcal{H} \cup \mathbb{Q} \cup \{i\infty\}$. The modular group $SL_2(\mathbb{Z})$ and the congruence subgroups

$\Gamma_0(N)$ act on $\overline{\mathcal{H}}$. Let Y be the quotient $\overline{\mathcal{H}}/\Gamma_0(N)$. One can construct a well behaved compactification of Y , $X_0(N)$. $X_0(N)$ is called the *modular curve* of level N and classifies pairs (Λ, C) of a lattice Λ and a cyclic subgroup C of order N . For the lattice Λ_τ ($\tau \in \mathcal{H}$), C_τ is simply the cyclic subgroup generated by $1/N$.

Barry Mazur proved a beautiful theorem on the *genus* g of the modular curve $X_0(N)$. For low genus he got:

genus g	level N
0	1, ..., 10, 12, 13, 16, 18, 25
1	11, 14, 15, 17, 19, 20, 21, 24, 27, 32, 36, 49
2	22, 23, 26, 28, 29, 31, 37, 50

Let g be the genus of the modular curve $X_0(N)$ and let (c_1, \dots, c_{2g}) be a \mathbb{Z} -basis of its integral homology $H_1(X_0(N), \mathbb{Z})$. Let $(\omega_1, \dots, \omega_g)$ be the dual \mathbb{C} -basis of the cohomology group $H^1(X_0(N), \mathbb{Z})$ and (f_1, \dots, f_g) the associated basis of $S_2(N)$. One defines a map Φ from the modular curve $X_0(N)$ on \mathbb{C}^g by

$$\Phi(\tau) = \left\{ \int_{\tau_0}^{\tau} f_j(z) dz \right\}_{j=1, \dots, g}$$

where τ_0 is a base point on $X_0(N)$. $\Phi(\tau)$ is well defined modulo the lattice $\Lambda(X_0(N))$ generated over \mathbb{Z} by the $2g$ points of \mathbb{C}^g

$$u_k = \left\{ \int_{c_k} f_j(z) dz \right\}_{j=1, \dots, g}$$

The Jacobian $J_0(N)$ is the quotient $\mathbb{C}^g/\Lambda(X_0(N))$.

12.4 Modular elliptic curves

The great success of Eichler and Shimura was to look at the possibility of expressing $L_E(s)$ as an $L_f(s)$ for a certain modular form f . The good objects are $\Gamma_0(N)$ -invariant holomorphic differentials $f(z)dz$ on the modular curve $X_0(N)$. But for that f must be a cusp form of level N and weight 2 for $\Gamma_0(N)$. Let therefore $f \in S_2(N)$. We integrate the differential $f(z)dz$ and get the function on \mathcal{H}

$$F(\tau) = \int_{\tau_0}^{\tau} f(z) dz$$

where τ_0 is a base point in \mathcal{H} . Let now $\gamma \in \Gamma_0(N)$. Since $f(z)dz$ is

$\Gamma_0(N)$ -invariant, we have:

$$\begin{aligned} F(\gamma(\tau)) &= \int_{\tau_0}^{\gamma(\tau)} f(z)dz = \int_{\tau_0}^{\gamma(\tau_0)} f(z)dz + \int_{\gamma(\tau_0)}^{\gamma(\tau)} f(z)dz \\ &= \int_{\tau_0}^{\gamma(\tau_0)} f(z)dz + \int_{\tau_0}^{\tau} f(z)dz \\ &= F(\tau) + \Phi_f(\gamma) \text{ with } \Phi_f(\gamma) = \int_{\tau_0}^{\gamma(\tau_0)} f(z)dz \end{aligned}$$

Φ_f is a map $\Phi_f : \Gamma_0(N) \rightarrow \mathbb{C}$ and we see that if its image $\Phi_f(\Gamma_0(N))$ is a lattice Λ in \mathbb{C} then the primitive $F(\tau)$ becomes a map

$$F : X_0(N) \rightarrow E = \mathbb{C}/\Lambda$$

which yields a parametrization of the elliptic curve E by the modular curve $X_0(N)$. In that case E is called a modular elliptic curve.

12.5 The Eichler-Shimura construction

Eichler-Shimura rather technical construction shows that if f is a *new-form* (in the sense of Atkin and Lehner) then

1. Λ is a lattice in \mathbb{C} ;
2. $X_0(N)$, E and F are defined over \mathbb{Q} in a *compatible* way;
3. and the L -functions of the elliptic curve E and the cusp form f are equal: $L_E(s) = L_f(s)$.

The construction is mediated by the Jacobian curve $J_0(N)$ of the modular curve $X_0(N)$, the elliptic curve E being a *quotient* of its Jacobian.

13 From Taniyama-Shimura-Weil to Fermat: Ribet theorem.

We have just seen that in the case of a modular elliptic curve E parametrized by a modular curve $X_0(N)$ the Dirichlet L -function $L_E(s)$ encoding the arithmetic properties of E shares all the good automorphy properties of the L -function $L_f(s)$ s.t. $L_E(s) = L_f(s)$. The Taniyama-Shimura-Weil conjecture says essentially that every elliptic curve is modular.

Taniyama-Shimura-Weil conjecture. Every elliptic curve is isogenous (that is a covering of finite degree) with a modular elliptic curve coming from an $X_0(N)$ and a $f \in S_2^{\text{new}}(N)$ by the Eichler-Shimura construction.

A result due to Carayol says that in that case the level N must be equal to the *conductor* N_E of E .

We arrive now at the turning point of the whole Odyssey.

Theorem. Taniyama-Shimura-Weil conjecture implies Fermat theorem.

Let $a^l + b^l + c^l = 0$ be an hypothetic solution of Fermat theorem for a prime $l \geq 5$ and a, b, c relatively prime non vanishing integers. We consider the associated Frey elliptic curve E of equation

$$y^2 = x(x - a^l)(x + c^l)$$

We know that the discriminant is $\Delta = 16(abc)^{2l}$ and it can be shown that the conductor is $N = \prod_{p|abc} p$. Ribet proved that these values *forbid* E to be modular.

The idea is to show that the level N can be reduced to the case $N = 2$ and then to use the fact that $S_2(2) = 0$ which shows that a parametrization associated to a modular form f cannot exist. The *reduction to level 2* is a consequent of a theorem of Ribet.

Ribet theorem. Let E be an elliptic curve defined over \mathbb{Q} having discriminant Δ with prime decomposition $\Delta = \prod_{p|\Delta} p^{\delta_p}$ and conductor

$N = \prod_{p|\Delta} p^{f_p}$. If E is a modular elliptic curve of level N associated to a

cuspidal form $f \in S_2(N)$, if l is a prime dividing the power δ_p of p in Δ and if $f_p = 1$ (that is if $p \parallel N$ in the sense $p \mid N$ but $p^2 \nmid N$) then *modulo* l the modular parametrization can be reduced to level $N' = N/p \pmod{l}$ in the sense that there exists a cuspidal form $f' \in S_2(N')$ s.t. the coefficients of f and f' are equal modulo l : $c_n \equiv c'_n \pmod{l} \forall n \geq 1$.

Let us apply Ribet theorem to the Frey curve. As a, b, c are relatively primes, for $p \neq 2$ we have $\delta_p = 2l$ and $f_p = 1$ and we can apply the theorem. For $p = 2$ the situation is different since $\delta_p = 4 + 2l$ and $l \nmid \delta_p$ and the reduction of levels leads to $N' = 2$. So there exists $f' \in S_2(2)$ such that $c_n \equiv c'_n \pmod{l} \forall n \geq 1$. We apply the lemma:

Lemma. $S_2(2) = 0$.

Indeed, in the $N = 2$ case, the modular curve $X_0(2)$ is of genus $g = 0$ (it is a sphere) and there exist therefore *no* non trivial holomorphic differential ω on $X_0(2)$ (the differential dz has a pole at infinity). As an $f \in S_2(2)$ corresponds to an ω , $S_2(2) = 0$. As $S_2(2) = 0$, we get for $n = 1$ the congruence $(c_1 = 1) \equiv (c'_1 = 0) \pmod{l}$ which is clearly impossible and $TSW \Rightarrow$ Fermat. So under the *TSW* conjecture, the proof of Fermat theorem amounts to the *topological obstruction* that a torus of genus 1 cannot be parametrized by a sphere of genus 0.

14 Encoding information in Galois representations

To prove the Taniyama-Shimura-Weil conjecture, Andrew Wiles used deep works of Jean-Pierre Serre and Barry Mazur on a specific class Galois representations naturally associated to elliptic curves. We meet here another extraordinary example of encoding informations of a theory into another theory. The arithmetic informations we will focus on are associated to *torsion points* of elliptic curves.

14.1 Torsion points and Galois representations

Let E be an elliptic curve and consider its Jacobian J which is a complex torus $J = \mathbb{C}/\Lambda$. The torsion points of order N of E correspond in J to those of the smaller lattice $\frac{1}{N}\Lambda$ that is those satisfying $Nx = 0$. Their set T_N in J is the kernel of the scale magnification $x \rightarrow Nx$, $T_N = \text{Ker}(x \rightarrow Nx)$, and is isomorphic to $\frac{\mathbb{Z}}{N\mathbb{Z}} \times \frac{\mathbb{Z}}{N\mathbb{Z}}$. So, through the isomorphism $E \simeq J$, the torsion points of E constitute a group $E_N \simeq \frac{\mathbb{Z}}{N\mathbb{Z}} \times \frac{\mathbb{Z}}{N\mathbb{Z}}$. If $\{\omega_1, \omega_2\}$ is a \mathbb{Z} -basis of Λ , $\{\omega_1/N, \omega_2/N\}$ is a \mathbb{Z} -basis of Λ/N and if x_i corresponds to ω_i/N by the isomorphism, $\{x_1, x_2\}$ is a \mathbb{Z} -basis of E_N .

Suppose now that E is defined over a field k . We can consider the extension $k(E_N)$ of the base field k defined by the adjunction of the coordinates of the N -torsion points which are algebraic over k (look at the formulae of division on E). If the characteristic of k doesn't divide N (we will have to use fields of characteristic l when we will work modulo l), $k(E_N)$ is an algebraic Galois extension of k and the elements $\sigma \in \text{Gal}(k(E_N)/k)$ act on $k(E_N)$. In the \mathbb{Z} -basis $\{x_1, x_2\}$ of E_N any such automorphism σ of $k(E_N)$ over k is represented by a 2×2 matrix and we get therefore a representation, called a *Galois representation*,

$$\rho : G = \text{Gal}(k(E_N)/k) \rightarrow GL_2\left(\frac{\mathbb{Z}}{N\mathbb{Z}}\right)$$

More generally, if K is an extension of k containing $k(E_N)$, we get a representation $\rho : \text{Gal}(K/k) \rightarrow GL_2\left(\frac{\mathbb{Z}}{N\mathbb{Z}}\right)$. In particular, for the case $k = \mathbb{Q}$ and $K = \overline{\mathbb{Q}}$ we get a Galois representation

$$\rho : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2\left(\frac{\mathbb{Z}}{N\mathbb{Z}}\right)$$

and in the case $N = p$ a prime, we get a Galois representation

$$\bar{\rho}_{E,p} : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{F}_p)$$

It is “continuous” in the sense it factorizes through the Galois group $\text{Gal}(K/\mathbb{Q})$ of a *finite* algebraic Galois extension K/\mathbb{Q} .

The Galois group $G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is one of the deepest object of Arithmetics and a lot of work have been devoted to its comprehension. There exist deep links between the Galois representations $\bar{\rho}_{E,p}$ of $G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and the L -functions. It is due to the remarkable following theorem:

Theorem. Let E an elliptic curve defined over \mathbb{Q} . Its Galois representation $\bar{\rho}_{E,p}$ satisfies the following properties:

1. Trace $\bar{\rho}_{E,p}(\text{Frob}_q) \equiv q + 1 - \#E(\mathbb{F}_q) = a_q \pmod{p}$ (this is the reason why we use a_q instead of $\#E(\mathbb{F}_q)$ for counting the points of $E \pmod{q}$).
2. $\text{Det } \bar{\rho}_{E,p} = \bar{\varepsilon}_p$ where $\bar{\varepsilon}_p : G \rightarrow \mathbb{F}_p^\times$ is the cyclotomic character giving the action of G on the p th roots of unity.

14.2 Modular representations and Deligne theorem

We have encode a lot of arithmetic informations on elliptic curves in Galois representations $\bar{\rho}_{E,p} : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{F}_p)$. Now, due to a fundamental work of Pierre Deligne, one can also associate such Galois representations to *modular forms*. Hence the strategic idea of proving *TSW* conjecture by proving that the $\bar{\rho}_{E,p}$ are modular.

Let $S_k(N, \varepsilon)$ be the space of cusp forms of weight k , level N and character ε . Hecke operators $T_k(p)$ act on $S_k(N, \varepsilon)$ and commute between them. Let $\lambda(p)$ be the eigenvalues of a common eigenform f of the $T_k(p)$, let R be the ring generated by the $\lambda(p)$ and the $\varepsilon(p)$ and let $\sim : R \rightarrow \mathbb{F}_p$ be a morphism of R into the finite field \mathbb{F}_p .

Deligne theorem. Under these hypotheses there exists a representation $\rho_f : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{F}_p)$ associated with f which is continuous, semi-simple and unramified outside lN s.t.

1. Trace $\rho_f(\text{Frob}_p) = \tilde{a}_p \forall p \nmid lN$.
2. $\text{Det } \rho_f(\text{Frob}_p) = p^{k-1} \widetilde{\varepsilon}(p)$.
3. $\text{Det } \rho_f(c) = -1$ where c is the complex conjugation (the character $\text{Det } \rho_f$ is odd).

15 Wiles side story

Andrew Wiles gave three lectures at the Isaac Newton Institute in Cambridge the 21-23 June 1993 presenting his proof of the Taniyama-Shimura-

Weil conjecture. The proof was not complete as it stood but was completed in a joint work with Richard Taylor and sent by Wiles to colleagues (including Faltings) on October 6, 1994. An excellent introduction to Wiles work is the text of Rubin and Silverberg. We will use it as our main reference.

15.1 Semi-stable modular lifting conjecture for $p = 3, 5$

Let E be an elliptic curve defined over \mathbb{Q} and consider the Galois representations yielded by its torsion points:

$$\bar{\rho}_{E,p} : G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{F}_p)$$

We have seen that $\text{Trace } \bar{\rho}_{E,p}(\text{Frob}_q) \equiv q + 1 - \#E(\mathbb{F}_q) = a_q \pmod{p}$ for almost every prime q and $\text{Det } \bar{\rho}_{E,p} = \bar{\varepsilon}_p : G \rightarrow \mathbb{F}_p^\times$ (the cyclotomic character giving the action of G on the p th roots of unity).

A first key idea of Wiles is to weaken *TSW* by considering it modulo p and to relativize it to a *single* prime p . The transformed conjecture is called the “semi-stable modular lifting conjecture”:

Semi-stable modular lifting conjecture (SSML). Suppose that E is *semi-stable* (that is if E is singular mod p then $E \pmod{p}$ is a node and not a cusp) and that there exists a prime $p \geq 3$ s.t.

- (a) $\bar{\rho}_{E,p}$ is irreducible,
- (b) E is modular but only mod \mathfrak{p} (where the ideal \mathfrak{p} lifts p in the ring of integers \mathcal{O}_f of the extension $\mathbb{Q}(a_n)$ of \mathbb{Q} by the algebraic integers a_n), i.e. there exists an eigenform $f \in S_2(N)$, $f = \sum_{n \geq 1} a_n q^n$, satisfying $a_q \equiv q + 1 - \#E(\mathbb{F}_q) \pmod{\mathfrak{p}}$ (very approximative equality) for almost every prime q ,

then E is really modular, i.e. there exists an eigenform $f \in S_2(N)$, $f = \sum_{n \geq 1} a_n q^n$, satisfying $a_q = q + 1 - \#E(\mathbb{F}_q)$ (exact equality) for almost every prime q .

Wiles proof is based on the fact that the semi-stable modular lifting conjecture for the first two primes $p = 3, 5$ is sufficient to prove the *semi-stable TSW* conjecture, which is itself sufficient for *FLT*. The key reason is that the group $PGL_2(\mathbb{F}_3)$ is isomorphic to the symmetric group S_4 of permutations of 4 elements and that for this extremely special dihedral case there exists a result of modularity.

Theorem. Semi-stable modular lifting conjecture for $p = 3, 5 \Rightarrow$ semi-stable *TSW* \Rightarrow *FLT*.

Sketch of the proof. Let E be defined over \mathbb{Q} and semi-stable and suppose that the semi-stable modular lifting conjecture is true for $p = 3$. Suppose first that the Galois representation $\bar{\rho}_{E,3}$ is *irreducible* (hypothesis (a)). Then E will be modular via the semi-stable modular lifting conjecture if hypothesis (b) is verified. For proving (b) one relies upon a fundamental theorem of Langlands and Tunnell concerning Galois representation ρ of $G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $GL_2(\mathbb{C})$ (and not in $GL_2(\mathbb{F}_p)$). This theorem is one of the few results constructing an eigen cusp form from a Galois representation. To formulate it we need to define the congruence group $\Gamma_1(N) = \left\{ \gamma \in SL_2(\mathbb{Z}) \mid \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}$.

Langlands-Tunnell theorem. Let $\rho : G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{C})$ be a continuous irreducible representation with odd determinant $\text{Det } \rho(c) = -1$ ($c = \text{complex conjugation}$). Suppose that the image $\rho(G)$ is a subgroup of S_4 (fundamental hypothesis of dihedrality). Then there exist a level N and an eigenform $g \in S_1(\Gamma_1(N))$, $g = \sum_{n \geq 1} b_n q^n$, s.t. for almost every prime q one has $b_q = \text{Trace } \rho(\text{Frob}_q)$.

To construct ρ in our case, we consider $\bar{\rho}_{E,3} : G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{F}_3)$. It is irreducible by hypothesis. We use the key fact that $GL_2(\mathbb{F}_3)$ can be embedded in $GL_2(\mathbb{C})$ through a well suited morphism ψ which factorizes through $GL_2(\mathbb{Z}\sqrt{-2})$ and satisfies

$$\begin{cases} \text{Trace}(\psi(g)) = \text{Trace}(g) \pmod{1 + \sqrt{-2}} \\ \text{Det}(\psi(g)) = \text{Det}(g) \pmod{3} \end{cases}$$

If $\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ are generators of $GL_2(\mathbb{F}_3)$, we define explicitly ψ by $\psi\left(\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}\right) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$ and $\psi\left(\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\right) = \begin{pmatrix} \sqrt{-2} & 1 \\ 1 & 0 \end{pmatrix}$. One shows that $\rho = \psi \circ \bar{\rho}_{E,3} : G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{C})$ is irreducible with odd determinant $\text{Det } \rho(c) = -1$ and that $\text{Im}(\rho) \subseteq PGL_2(\mathbb{F}_3) \simeq S_4$. One can therefore apply Langlands-Tunnell. There exist a level N and an eigenform $g \in S_1(\Gamma_1(N))$, $g = \sum_{n \geq 1} b_n q^n$, s.t. for almost every prime q one has $b_q = \text{Trace } \rho(\text{Frob}_q)$. From g , one constructs then an eigenform $f \in S_2(N) = S_2(\Gamma_0(N))$ s.t. $\forall n \ a_n \equiv b_n \pmod{\mathfrak{p}}$, where \mathfrak{p} is the prime ideal of $\bar{\mathbb{Q}}$ containing $1 + \sqrt{-2}$. The congruences show that the eigenform f satisfies (b) for the ideal $\mathfrak{p}' = \mathfrak{p} \cap \mathcal{O}_f$ and therefore E is modular.

Suppose now that the representation $\bar{\rho}_{E,3}$ is *reducible*. If the representation $\bar{\rho}_{E,5}$ is also reducible then E is modular. Indeed, the group of points of E over $\bar{\mathbb{Q}}$ contains a cyclic subgroup of order $15 = 3 \cdot 5$ which

is G -stable. But the pairs (E, C) are classified by the *rational* points of the modular curve $X_0(15)$. But $X_0(15)$ has only 4 rational points and it can be shown that they all correspond to modular curves.

We can therefore suppose that $\bar{\rho}_{E,5}$ is *irreducible*. In that case, Wiles method is to construct *another* auxiliary elliptic curve E' defined over \mathbb{Q} and semi-stable s.t.

1. $\bar{\rho}_{E',5} = \bar{\rho}_{E,5}$, and
2. $\bar{\rho}_{E',3}$ is *irreducible*.

Let us suppose that E' is constructed. According to the case explained before, E' is modular. Let $f \in S_2(N)$, $f = \sum_{n \geq 1} a_n q^n$, be the associated eigenform. For almost every prime q we have $a_q = q + 1 - \#E'(\mathbb{F}_q)$. But $q + 1 - \#E'(\mathbb{F}_q) \equiv \text{Trace } \bar{\rho}_{E',5}(\text{Frob}_q) \pmod{5}$. And, as $\bar{\rho}_{E',5} = \bar{\rho}_{E,5}$, we have the congruence

$$\text{Trace } \bar{\rho}_{E',5} \left(\text{Frob}_q \right) = \text{Trace } \bar{\rho}_{E,5} \left(\text{Frob}_q \right) \equiv q + 1 - \#E(\mathbb{F}_q) \pmod{5}$$

and f satisfies therefore the condition (b) of the semi-stable modular lifting conjecture for $p = 5$. We conclude that E is *modular*.

15.2 The construction of the auxiliary elliptic curve

At this point, the main difficulty is to construct the auxiliary elliptic curve E' . The starting point is that elliptic curves E' satisfying $\bar{\rho}_{E',p} = \bar{\rho}_{E,p}$ are classified by the rational points of the Riemann surface $X(p)$ (defined over \mathbb{Q}) $X(p) = \mathcal{H}/\Gamma(p)$ where

$$\Gamma(p) = \{\gamma \in SL_2(\mathbb{Z}) \mid \gamma \equiv \text{Id} \pmod{p}\}$$

is the subgroup of integral matrices of $SL_2(\mathbb{Z})$ which are congruent to the identity matrix modulo p . We will use again a *topological* argument, namely that $X(p)$ is of genus $g = 0$ for $p \leq 5$. But when $g = 0$, if there exists a rational point (which is the case here with $E' = E$) then there exist an *infinite number* of rational points. One then shows:

Proposition. For an *infinite number* of rational points of $X(5)$ $\bar{\rho}_{E',3}$ is *irreducible*.

One uses the fact that if E' is a *generic* point (and therefore not rational) of $X(5)$ then its Galois group given by its p -torsion points is “big” in the sense that the image of $G = \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ in $GL_2(\mathbb{F}_p)$ is maximal (that is equal to $GL_2(\mathbb{F}_p)$). But a theorem due to Hilbert, called the *irreducibility theorem*, says that “many” specializations of a generic point have the same Galois group and we can conclude.

One shows next that E' can be chosen semi-stable. If the prime $q \neq 5$ semi-stability reads on $E'[5]$ and as $E'[5] = E[5]$ and E is semi-stable at q by hypothesis, E' is also semi-stable at q . For $q = 5$ one choose an E' which is “close” to E for the p -adic metric and use the fact that semi-stability is an *open* property. As E is semi-stable at 5 by hypothesis, E' is also semi-stable at 5.

15.3 Lifting to p -adic representations: from characteristic p to characteristic 0

Up to now, we have considered only representations of $G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into $GL_2(\mathbb{Z}/N\mathbb{Z})$ induced by the N -torsion points $E[N] \simeq \frac{\mathbb{Z}}{N\mathbb{Z}} \times \frac{\mathbb{Z}}{N\mathbb{Z}}$ of elliptic curves. We will now look at *all* the representations associated to the successive powers p^n of a prime p . Taking their projective limit, we get a continuous representation in the algebra \mathbb{Z}_p of *p -adic integers*

$$\rho_{E,p} : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{Z}_p)$$

which satisfies the properties:

1. $\text{Det } \rho_{E,p} = \varepsilon_p$ (where ε_p is the cyclotomic character $\varepsilon_p : G \rightarrow \mathbb{Z}_p^\times$),
2. for almost every prime q , $\text{Trace } \rho_{E,p}(\text{Frob}_q) = q + 1 - \#E(\mathbb{F}_q)$ (exact equality).

(of course, through the quotient $\mathbb{Z}_p \rightarrow \mathbb{F}_p$, $\rho_{E,p}$ returns $\bar{\rho}_{E,p}$).

Once again, we will say that a p -adic representation

$$\rho : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(\mathbb{Z}_p)$$

is *modular* if there exists an eigenform $f \in S_2(N)$, $f = \sum_{n \geq 1} a_n q^n$, s.t.

$\text{Trace } \rho(\text{Frob}_q) = a_q$ for almost every prime q in a well suited extension of \mathbb{Z}_p (for instance a completion $\mathcal{O}_{f,\mathfrak{p}}$ for $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$). The semi-stable modular lifting conjecture says essentially that, given E defined over \mathbb{Q} and semi-stable and $p \geq 3$, if $\bar{\rho}_{E,p}$ is irreducible and modular then $\rho_{E,p}$ is modular. We see that this is a problem of *lifting* the modularity property from the prime field \mathbb{F}_p of characteristic p to the p -adic algebra \mathbb{Z}_p which is the ring of integers of the local field \mathbb{Q}_p of characteristic 0.

We generalize the lifting problem to the case where we have a finite algebraic extension K/k of $k = \mathbb{F}_p$ or of \bar{k} and a \mathbb{Z}_p -algebra A Noetherian, local, complete with residual field k . We start from a representations $\bar{\rho} : G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow GL_2(k)$ and we look for liftings $\rho : G \rightarrow GL_2(A)$

making the following diagram commutative ($i \circ \rho = \bar{\rho} \otimes_k \bar{k}$):

$$\begin{array}{ccc} & & GL_2(A) \\ & \nearrow \rho & \downarrow i \\ G & \xrightarrow{\bar{\rho} \otimes_k \bar{k}} & GL_2(\bar{k}) \end{array}$$

where $i : A \rightarrow \bar{k}$ is a morphism and $\bar{\rho} \otimes_k \bar{k}$ extends the field of scalars from k to \bar{k} .

We use now the deep analogy between arithmetics and geometry linking finite fields \mathbb{F}_p and p -adic fields \mathbb{Q}_p : k is like a “point” and A like a “germ of deformation” and therefore a lifting $\bar{\rho} \rightarrow \rho$ is like to lift the value of a function at a point to a germ of function near the point.

15.4 Deformation data and Barry Mazur conjectures

We consider liftings satisfying constraints called “deformation data” by Barry Mazur. A deformation data is a pair $\mathfrak{D} = (\Sigma, t)$ where Σ is a finite set of primes q outside of which representations are *unramified* (which means that if $q \notin \Sigma$ then $\rho(I_q) = 1$, where I_q is the inertia group of q , that is the subgroup of $\text{Gal}(K/\mathbb{Q})$ constituted by the σ which fix \mathfrak{q} and induce the identity on $\mathcal{O}_K/\mathfrak{q}$) and t is a set of properties of representations ρ at p (to be “ordinary”, to be “flat”, etc.). Once again, a representation $\bar{\rho} : G \rightarrow GL_2(k)$ is called *\mathfrak{D} -modular* if there exists an eigenform $f \in S_2(N)$ and a prime ideal \mathfrak{p} over p ($\mathfrak{p} \mid p$) in \mathcal{O}_f s.t. the representation $\rho_{f,\mathfrak{p}}$ associated to f by the Eichler-Shimura construction is a \mathfrak{D} -lifting of $\bar{\rho}$.

Mazur conjecture 1. Let $\bar{\rho} : G \rightarrow GL_2(k)$ be absolutely irreducible (that is $\bar{\rho} \otimes_k \bar{k}$ is irreducible) and \mathfrak{D} -modular, then every \mathfrak{D} -lifting of $\bar{\rho}$ to the integer ring $\mathcal{O} = A$ of a finite extension of \mathbb{Q}_p with residual field k is modular.

Wiles theorem. Mazur 1 \Rightarrow Semi-stable modular lifting conjecture.

Indeed let E be an elliptic curve defined over \mathbb{Q} and semi-stable which satisfies the conditions (a) and (b) of the semi-stable modular lifting conjecture for p and let $\bar{\rho}$ be the representation $\bar{\rho} = \bar{\rho}_{E,p}$. According to hypothesis (a) $\bar{\rho}$ is irreducible. One shows that it is also *absolutely* irreducible. The hypothesis (b) means that $\bar{\rho}$ is modular. Let $\mathfrak{D} = (\Sigma, t)$ be the deformation data defined by

$$\Sigma = \{p\} \cup \{q \mid E \text{ has bad reduction at } q\}$$

and t means “ordinary” if E has ordinary or bad reduction at q (ordinary reduction means good reduction and $E[q]$ has a subgroup of order

q which is I_q -stable) and “flat” if E has supersingular reduction at q (supersingular reduction means good reduction and $E[q]$ has no subgroup of order q which is I_q -stable). One shows that $\rho_{E,p}$ is a \mathfrak{D} -lifting of $\bar{\rho}$ and that $\bar{\rho}$ is \mathfrak{D} -modular. Mazur 1 implies that $\rho_{E,p}$ is modular and therefore E is modular.

In a second stage, one reformulates the first Mazur conjecture in terms of *universal deformations* for $(\mathfrak{D}, \mathcal{O})$ -deformations

$$\begin{array}{ccc} & & GL_2(A) \\ & \nearrow^{\rho} & \downarrow i \\ G & \xrightarrow{\bar{\rho}} & GL_2(k) \end{array}$$

where A is a local, Noetherian, complete \mathcal{O} -algebra of residual field k , \mathcal{O} being the ring of integers of a finite extension of \mathbb{Q}_p .

Mazur-Ramakrishna theorem. There exists a *universal* $(\mathfrak{D}, \mathcal{O})$ -lifting $\rho_R : G \rightarrow GL_2(R)$ of $\bar{\rho}$, that is for every $(\mathfrak{D}, \mathcal{O})$ -lifting $\rho : G \rightarrow GL_2(A)$ there exists one and only one morphism of algebras $\varphi_\rho : R \rightarrow A$ s.t. the following diagram is commutative:

$$\begin{array}{ccc} & & GL_2(A) \\ & \nearrow^{\rho} & \downarrow \varphi_\rho^* \\ G & \xrightarrow{\rho_R} & GL_2(R) \\ \parallel & \nearrow^{\rho_R} & \downarrow i \\ G & \xrightarrow{\bar{\rho}} & GL_2(k) \end{array}$$

But if $\bar{\rho}$ is \mathfrak{D} -modular with an eigenform f and a prime ideal \mathfrak{p} of \mathcal{O}_f s.t. $\rho_{f,\mathfrak{p}}$ is a \mathfrak{D} -lifting of $\bar{\rho}$ and $\rho_{f,\mathfrak{p}} \otimes \mathcal{O}_f$ is a $(\mathfrak{D}, \mathcal{O})$ -lifting of $\bar{\rho}$ then there exists also a *modular* universal deformation in the following sense:

- T1 The algebra A is a generalized Hecke algebra T .
- T2 There exists a level N divisible only by “bad” primes $q \in \Sigma$ and a morphism $j : T(N) \rightarrow T$ from the Hecke algebra $T(N)$ acting on $S_2(N)$ to T s.t. T is generated over \mathcal{O} by the images $j(T_q)$ of the Hecke operators T_q for $q \notin \Sigma$.
- T3 There exists a $(\mathfrak{D}, \mathcal{O})$ -lifting of $\bar{\rho}$, $\rho_T : G \rightarrow GL_2(T)$, s.t.

$$\text{Trace } \rho_T \left(\text{Frob}_q \right) = j(T_q) \forall q \notin \Sigma.$$

- T4 If ρ is a modular $(\mathfrak{D}, \mathcal{O})$ -lifting of $\bar{\rho}$ to an A , then there exists one and only one \mathcal{O} -morphism $\psi_\rho : T \rightarrow A$ s.t. the following diagram

is commutative:

$$\begin{array}{ccc} & & GL_2(T) \\ & \nearrow^{\rho_T} & \downarrow \psi_\rho^* \\ G & \xrightarrow{\rho} & GL_2(R) \end{array}$$

As ρ_T is a $(\mathfrak{D}, \mathcal{O})$ -lifting of $\bar{\rho}$, Mazur-Ramakrishna theorem implies that there exists one and only one morphism of algebras $\varphi : R \rightarrow T$ s.t. $\rho_T = \varphi \circ \rho_R$. The map φ is *surjective* since

$$\forall q \notin \Sigma, \varphi \left(\text{Trace } \rho_R \left(\text{Frob}_q \right) \right) = \text{Trace } \rho_T \left(\text{Frob}_q \right) = j(T_q)$$

and the $j(T_q)$ generate T by (2).

Mazur introduced a second conjecture saying intuitively that parametrizations of ordinary liftings and modular liftings are equivalent or that “universal” is equivalent to “modular universal”, which is clearly a translation of the *TSW* conjecture in the context of universal deformations.

Mazur conjecture 2. $\varphi : R \rightarrow T$ is an *isomorphism*.

Theorem. Mazur conjecture 2 implies Mazur conjecture 1.

Sketch of the proof. Let $\bar{\rho} : G \rightarrow GL_2(k)$ be absolutely irreducible and \mathfrak{D} -modular. If ρ is a \mathfrak{D} -lifting of $\bar{\rho}$ to A , we want to show that ρ is modular. We first extend ρ and $\bar{\rho}$ to \mathcal{O} and ρ becomes a $(\mathfrak{D}, \mathcal{O})$ -lifting. Let $\psi_\rho : R \rightarrow A$ be the morphism of algebras asserted by Mazur-Ramakrishna theorem. If $\varphi : R \rightarrow T$ is an *isomorphism* we can consider the *inverse* map $\varphi^{-1} : T \rightarrow R$ and the composed map $\psi = \psi_\rho \circ \varphi^{-1} : T \rightarrow A$

$$\psi : T \xrightarrow{\varphi^{-1}} R \xrightarrow{\psi_\rho} A$$

We deduce from (T3) that $\psi(T_q) = \text{Trace } \rho(\text{Frob}_q)$ for almost every prime q . Shimura results imply then the existence of an eigenform $f \in S_2(N)$, $f = \sum_{n \geq 1} a_n q^n$, s.t. $a_q = \text{Trace } \rho(\text{Frob}_q) = \psi(T_q)$ for almost every prime q . But this implies that the representation ρ is modular.

15.5 Gorenstein rings and “cotangent spaces”

The problem is now to prove that $\varphi : R \rightarrow T$ is an *isomorphism*. The idea is to *bound* the order of “tangent spaces” at a prime ideal of R in the following sense. If $\bar{\rho}$ is \mathfrak{D} -modular, there exists an eigenform $f \in S_2(N)$ and a prime ideal $\mathfrak{p} \mid p$ of \mathcal{O}_f such that $\rho_{f,\mathfrak{p}}$ is a \mathfrak{D} -lifting of $\bar{\rho}$. If $\mathcal{O}_f \subset \mathcal{O}$, $\rho_{f,\mathfrak{p}} \otimes_{\mathcal{O}_f} \mathcal{O}$ is a $(\mathfrak{D}, \mathcal{O})$ -lifting of $\bar{\rho}$. As the Galois representation $\rho_{f,\mathfrak{p}} \otimes_{\mathcal{O}_f} \mathcal{O}$ is modular by construction, (T4) implies that there exists one and only one morphism $\pi : T \rightarrow \mathcal{O}$ s.t. $\pi \circ \rho_T = \rho_{f,\mathfrak{p}} \otimes_{\mathcal{O}_f} \mathcal{O}$. Let $\mathfrak{p}_T = \text{Ker}(\pi)$

and $\mathfrak{p}_R = \text{Ker}(\pi \circ \varphi) = \varphi^{-1}(\mathfrak{p}_T)$. (T2) and the fact that for almost every prime q $\text{Trace } \rho_{f,\mathfrak{p}}(\text{Frob}_q) = a_q$ imply that for almost every prime q , $\pi(T_q) = a_q$.

At this point Wiles uses a special property of the Hecke algebra T , namely to be a *Gorenstein ring*. This result due to by Barry Mazur means that there exists a (non canonical) isomorphism of T -modules between T and $\text{Hom}_{\mathcal{O}}(T, \mathcal{O})$. The morphism $\pi : T \rightarrow \mathcal{O}$ corresponds therefore to an element ξ of T and, via π itself, to an element $\pi(\xi)$ of the ring \mathcal{O} :

$$\begin{array}{ccc} \text{Hom}_{\mathcal{O}}(T, \mathcal{O}) & \xrightarrow{\sim} & T & \xrightarrow{\pi} & \mathcal{O} \\ & & \pi & \mapsto & \xi & \mapsto & \pi(\xi) \end{array}$$

Let η be the ideal $(\pi(\xi))$ of \mathcal{O} (η is well defined independently of the isomorphism $\text{Hom}_{\mathcal{O}}(T, \mathcal{O}) \simeq T$). Wiles gave a sufficient condition for $\varphi : R \rightarrow T$ to be an isomorphism in terms of order of the “cotangent space” $\mathfrak{p}_R/\mathfrak{p}_R^2$.

Theorem (Wiles). If $\# \left(\frac{\mathfrak{p}_R}{\mathfrak{p}_R^2} \right) \leq \# \left(\frac{\mathcal{O}}{\eta} \right)$ is *finite*, then $\varphi : R \rightarrow T$ is an isomorphism.

As φ is onto, $\# \left(\frac{\mathfrak{p}_R}{\mathfrak{p}_R^2} \right) \geq \# \left(\frac{\mathfrak{p}_T}{\mathfrak{p}_T^2} \right)$. Wiles shows $\# \left(\frac{\mathfrak{p}_T}{\mathfrak{p}_T^2} \right) \geq \# \left(\frac{\mathcal{O}}{\eta} \right)$ and therefore, if $\# \left(\frac{\mathfrak{p}_R}{\mathfrak{p}_R^2} \right) \leq \# \left(\frac{\mathcal{O}}{\eta} \right)$ we get the equalities

$$\# \left(\frac{\mathfrak{p}_R}{\mathfrak{p}_R^2} \right) = \# \left(\frac{\mathfrak{p}_T}{\mathfrak{p}_T^2} \right) = \# \left(\frac{\mathcal{O}}{\eta} \right)$$

and φ induces an isomorphism of the “tangent spaces” of R and T at the corresponding “points” \mathfrak{p}_R and \mathfrak{p}_T . Due to the properties of T (namely to be a complete intersection over \mathcal{O}), this “tangent isomorphism” implies that φ is an isomorphism.

The last difficulty in the proof of the *TSW* conjecture is then to *bound* the order $\# \left(\frac{\mathcal{O}}{\eta} \right)$. The new idea is to give a *cohomological* interpretation of “tangent spaces” in terms of *Selmer groups*. It is the most technical and difficult part of the proof.

15.6 Wiles own description of his proof

In his reference paper, Wiles summarizes the story of his proof. We will quote this passage in its integrality.

“The following is an account of the origins of this work and of the more specialized developments of the 1980’s that affected it. I began working on these problems in the late summer of 1986 immediately on learning of Ribet’s result.

For several years I had been working on the Iwasawa conjecture for totally real fields and some applications of it. In the process, I had been using and developing results on l -adic representations associated to Hilbert modular forms. It was therefore natural for me to consider the problem of modularity from the point of view of l -adic representations. I began with the assumption that the reduction of a given ordinary l -adic representation was reducible and tried to prove under this hypothesis that the representation itself would have to be modular. I hoped rather naively that in this situation I could apply the techniques of Iwasawa theory. Even more optimistically I hoped that the case $l = 2$ would be tractable as this would suffice for the study of the curves used by Frey. From now on and in the main text, we write p for l because of the connections with Iwasawa theory.

“After several months studying the 2-adic representation, I made the first real breakthrough in realizing that I could use the 3-adic representation instead: the Langlands-Tunnell theorem meant that ρ_3 , the mod 3 representation of any given elliptic curve over \mathbb{Q} , would necessarily be modular. This enabled me to try inductively to prove that the $GL_2(\mathbb{Z}/3^n\mathbb{Z})$ representation would be modular for each n . At this time I considered only the ordinary case. This led quickly to the study of $H^i(\text{Gal}(F_\infty/\mathbb{Q}), W_f)$ for $i = 1$ and 2 , where F_∞ is the splitting field of the \mathfrak{m} -adic torsion on the Jacobian of a suitable modular curve, \mathfrak{m} being the maximal ideal of a Hecke ring associated to ρ_3 and W_f the module associated to a modular form f described in Chapter 1. More specifically, I needed to compare this cohomology with the cohomology of $\text{Gal}(\mathbb{Q}_\Sigma/\mathbb{Q})$ acting on the same module.

“I tried to apply some ideas from Iwasawa theory to this problem. In my solution to the Iwasawa conjecture for totally real fields, I had introduced a new technique in order to deal with the trivial zeroes. It involved replacing the standard Iwasawa theory method of considering the fields in the cyclotomic \mathbb{Z}_p -extension by a similar analysis based on a choice of infinitely many distinct primes $q_i \equiv 1 \pmod{p^{n_i}}$ with $n_i \rightarrow \infty$ as $i \rightarrow \infty$. Some aspects of this method suggested that an alternative to the standard technique of Iwasawa theory, which seemed problematic in the study of W_f , might be to make a comparison between the cohomology groups as Σ varies but with the field \mathbb{Q} fixed. The new principle said roughly that the unramified cohomology classes are trapped by the tamely

ramified ones. After reading the paper [Gre1]⁵, I realized that the duality theorems in Galois cohomology of Poitou and Tate would be useful for this. The crucial extract from this latter theory is in Section 2 of Chapter 1.

“In order to put ideas into practice I developed in a naive form the techniques of the first two sections of Chapter 2. This drew in particular on a detailed study of all the congruences between f and other modular forms of differing levels, a theory that had been initiated by Hida and Ribet. The outcome was that I could estimate the first cohomology group well under two assumptions, first that a certain subgroup of the second cohomology group vanished and second that the form f was chosen at the minimal level for \mathfrak{m} . These assumptions were much too restrictive to be really effective but at least they pointed in the right direction. Some of these arguments are to be found in the second section of Chapter 1 and some form the first weak approximation to the argument in Chapter 3. At that time, however, I used auxiliary primes $q \equiv -1 \pmod{p}$ when varying Σ as the geometric techniques I worked with did not apply in general for primes $q \equiv 1 \pmod{p}$. (This was for much the same reason that the reduction of level argument in [Ri1]⁶ is much more difficult when $q \equiv 1 \pmod{p}$.) In all this work I used the more general assumption that ρ_p was modular rather than the assumption that $p = -3$.

“In the late 1980’s, I translated these ideas into ring-theoretic language. A few years previously Hida had constructed some explicit one-parameter families of Galois representations. In an attempt to understand this, Mazur had been developing the language of deformations of Galois representations. Moreover, Mazur realized that the universal deformation rings he found should be given by Hecke rings, at least in certain special cases. This critical conjecture refined the expectation that all ordinary liftings of modular representations should be modular. In making the translation to this ring-theoretic language I realized that the vanishing assumption on the subgroup of H^2 which I had needed should be replaced by the stronger condition that the Hecke rings were complete intersections. This fitted well with their being deformation rings where one could estimate the number of generators and relations and so made the original assumption more plausible.

⁵A paper of R. Greenberg on Iwasawa theory for p -adic representations.

⁶1990 Ribet’s paper in the bibliography.

“To be of use, the deformation theory required some development. Apart from some special examples examined by Boston and Mazur there had been little work on it. I checked that one could make the appropriate adjustments to the theory in order to describe deformation theories at the minimal level. In the fall of 1989, I set Ramakrishna, then a student of mine at Princeton, the task of proving the existence of a deformation theory associated to representations arising from finite flat group schemes over \mathbb{Z}_p . This was needed in order to remove the restriction to the ordinary case. These developments are described in the first section of Chapter 1 although the work of Ramakrishna was not completed until the fall of 1991. For a long time the ring-theoretic version of the problem, although more natural, did not look any simpler. The usual methods of Iwasawa theory when translated into the ring-theoretic language seemed to require unknown principles of base change. One needed to know the exact relations between the Hecke rings for different fields in the cyclotomic \mathbb{Z}_p -extension of \mathbb{Q} , and not just the relations up to torsion.

“The turning point in this and indeed in the whole proof came in the spring of 1991. In searching for a clue from commutative algebra I had been particularly struck some years earlier by a paper of Kunz [Ku2]. I had already needed to verify that the Hecke rings were Gorenstein in order to compute the congruences developed in Chapter 2. This property had first been proved by Mazur in the case of prime level and his argument had already been extended by other authors as the need arose. Kunz’s paper suggested the use of an invariant (the η -invariant of the appendix) which I saw could be used to test for isomorphisms between Gorenstein rings. A different invariant (the $\mathfrak{p}/\mathfrak{p}^2$ -invariant of the appendix) I had already observed could be used to test for isomorphisms between complete intersections. It was only on reading Section 6 of [Ti2]⁷ that I learned that it followed from Tate’s account of Grothendieck duality theory for complete intersections that these two invariants were equal for such rings. Not long afterwards I realized that, unlike though it seemed at first, the equality of these invariants was actually a criterion for a Gorenstein ring to be a complete intersection. These arguments are given in the appendix.

“The impact of this result on the main problem was enor-

⁷Tilouine’s paper on Iwasawa’s theory and Hecke algebras.

mous. Firstly, the relationship between the Hecke rings and the deformation rings could be tested just using these two invariants. In particular I could provide the inductive argument of section 3 of Chapter 2 to show that if all liftings with restricted ramification are modular then all liftings are modular. This I had been trying to do for a long time but without success until the breakthrough in commutative algebra. Secondly, by means of a calculation of Hida summarized in [Hi2] the main problem could be transformed into a problem about class numbers of a type well-known in Iwasawa theory. In particular, I could check this in the ordinary CM case⁸ using the recent theorems of Rubin and Kolyvagin. This is the content of Chapter 4. Thirdly, it meant that for the first time it could be verified that infinitely many j -invariants were modular. Finally, it meant that I could focus on the minimal level where the estimates given by me earlier Galois cohomology calculations looked more promising. Here I was also using the work of Ribet and others on Serre’s conjecture (the same work of Ribet that had linked Fermat’s Last Theorem to modular forms in the first place) to know that there was a minimal level.

“The class number problem was of a type well-known in Iwasawa theory and in the ordinary case had already been conjectured by Coates and Schmidt. However, the traditional methods of Iwasawa theory did not seem quite sufficient in this case and, as explained earlier, when translated into the ring theoretic language seemed to require unknown principles of base change. So instead I developed further the idea of using auxiliary primes to replace the change of field that is used in Iwasawa theory. The Galois cohomology estimates described in Chapter 3 were now much stronger, although at that time I was still using primes $q \equiv -1 \pmod{p}$ for the argument. The main difficulty was that although I knew how the η -invariant changed as one passed to an auxiliary level from the results of Chapter 2, I did not know how to estimate the change in the $\mathfrak{p}/\mathfrak{p}^2$ -invariant precisely. However, the method did give the right bound for the generalised class group, or Selmer group as it often called in this context, under the additional assumption that the minimal Hecke ring was a complete intersection.

“I had earlier realized that ideally what I needed in this

⁸CM means “complex multiplication”.

method of auxiliary primes was a replacement for the power series ring construction one obtains in the more natural approach based on Iwasawa theory. In this more usual setting, the projective limit of the Hecke rings for the varying fields in a cyclotomic tower would be expected to be a power series ring, at least if one assumed the vanishing of the η -invariant. However, in the setting with auxiliary primes where one would change the level but not the field, the natural limiting process did not appear to be helpful, with the exception of the closely related and very important construction of Hida [Hi1]. This method of Hida often gave one step towards a power series ring in the ordinary case. There were also tenuous hints of a patching argument in Iwasawa theory ([Scho]⁹, [Wi4, §10]¹⁰), but I searched without success for the key.

“Then, in August, 1991, I learned of a new construction of Flach [Fl] and quickly became convinced that an extension of his method was more plausible. Flach’s approach seemed to be the first step towards the construction of an Euler system, an approach which would give the precise upper bound for the size of the Selmer group if it could be completed. By the fall of 1992, I believed I had achieved this and begun then to consider the remaining case where the mod 3 representation was assumed reducible. For several months I tried simply to repeat the methods using deformation rings and Hecke rings. Then unexpectedly in May 1993, on reading of a construction of twisted forms of modular curves in a paper of Mazur [Ma3], I made a crucial and surprising breakthrough: I found the argument using families of elliptic curves with a common ρ_5 which is given in Chapter 5. Believing now that the proof was complete, I sketched the whole theory in three lectures in Cambridge, England on June 21-23. However, it became clear to me in the fall of 1993 that the construction of the Euler system used to extend Flach’s method was incomplete and possibly flawed.

“Chapter 3 follows the original approach I had taken to the problem of bounding the Selmer group but had abandoned on learning of Flach’s paper. Darmon encouraged me in February, 1994, to explain the reduction to the complete intersection property, as it gave a quick way to exhibit infinite families of modular j -invariants. In presenting it in a lecture at Princeton, I made, almost unconsciously, critical

⁹Schoof’s paper on the minus class groups of abelian number fields.

¹⁰Wiles’ paper on the Iwasawa conjecture for totally real fields.

switch to the special primes used in Chapter 3 as auxiliary primes. I had only observed the existence and importance of these primes in the fall of 1992 while trying to extend Flach’s work. Previously, I had only used primes $q \equiv -1 \pmod{p}$ as auxiliary primes. In hindsight this change was crucial because of a development due to de Shalit. As explained before, I had realized earlier that Hida’s theory often provided one step towards a power series ring at least in the ordinary case. At the Cambridge conference de Shalit had explained to me that for primes $q \equiv 1 \pmod{p}$ he had obtained a version of Hida’s results. But except for explaining the complete intersection argument in the lecture at Princeton, I still did not give any thought to my initial approach, which I had put aside since the summer of 1991, since I continued to believe that the Euler system approach was the correct one.

“Meanwhile in January, 1994, R. Taylor had joined me in the attempt to repair the Euler system argument. Then in the spring of 1994, frustrated in the efforts to repair the Euler system argument, I began to work with Taylor on an attempt to devise a new argument using $p = 2$. The attempt to use $p = 2$ reached an impasse at the end of August. As Taylor was still not convinced that the Euler system argument was irreparable, I decided in September to take one last look at my attempt to generalise Flach, if only to formulate more precisely the obstruction. In doing this I came suddenly to a marvelous revelation: I saw in a flash on September 19th, 1994, that de Shalit’s theory, if generalised, could be used together with duality to glue the Hecke rings at suitable auxiliary levels into a power series ring. I had unexpectedly found the missing key to my old abandoned approach. It was the old idea of picking q_i ’s with $q_i \equiv 1 \pmod{p^{n_i}}$ and $n_i \rightarrow \infty$ as $i \rightarrow \infty$ that I used to achieve the limiting process. The switch to the special primes of Chapter 3 had made all this possible.

“After I communicated the argument to Taylor, we spent the next few days making sure of the details. the full argument, together with the deduction of the complete intersection property, is given in [TW]¹¹.

“In conclusion the key breakthrough in the proof had been the realization in the spring of 1991 that the two invariants introduced in the appendix could be used to relate the defor-

¹¹Taylor-Wiles’ paper.

mation rings and the Hecke rings. In effect the η -invariant could be used to count Galois representations. The last step after the June, 1993, announcement, though elusive, was but the conclusion of a long process whose purpose was to replace, in the ring-theoretic setting, the methods based on Iwasawa theory by methods based on the use of auxiliary primes.

“One improvement that I have not included but which might be used to simplify some of Chapter 2 is the observation of Lenstra that the criterion for Gorenstein rings to be complete intersections can be extended to more general rings which are finite and free as \mathbb{Z}_p -modules. Faltings has pointed out an improvement, also not included, which simplifies the argument in Chapter 3 and [TW]. This is however explained in the appendix to [TW].”

16 Conclusion: the categorical complexity of a proof

We have tried to present briefly the elements of Wiles’ proof of the Taniyama-Shimura-Weil conjecture and we emphasized the fact that its main steps consist in translating parts of theories into another theories in order to make explicit and tractable some pieces of information. To say that the proof is not “direct and elementary” but “indirect and complex” is to say that the amount of such translational steps is very high. The problem would be now to try to formalize this type of complexity.

I think that there would be a possibility working in the framework of category theory. Indeed “translations” are in general *functors* from one category into another and we can say that a “direct and elementary” proof is a sequence of deductive steps (in the sense of proof theory) inside a single category, while an “indirect and complex” proof is a proof using also many functorial changes of category.

References

- [1] Carayol, H., “Sur les représentations galoisiennes modulo l attachées aux formes modulaires”, *Duke Math. J.* 59 (1989), 785-801.
- [2] Darmon, H., Diamond, F., Taylor, R.L., “Fermat’s Last Theorem”, in *Current Developments in Mathematics*, 1995, International Press. COMPLETE.

- [3] Deligne, P., “Formes modulaires et représentations l -adiques”, *Séminaire Bourbaki*, 1968-1969, Exposé 355, *Lect. Notes in Maths.* 179 (1971), 139-172.
- [4] Deligne, P., Serre, J-P., “Formes modulaires de poids 1”, *Ann. Sci. ENS*, 7 (1974), 507-530.
- [5] Diamond, F., “On deformations rings and Hecke rings”, *Ann. of Maths.*, COMPLETE.
- [6] Diamond, F., “The Taylor-Wiles construction and multiplicity one”, *Invent. Math.*, COMPLETE.
- [7] Dieudonné, J., *Panorama des Mathématiques pures. Le choix bourbachique*, Paris, Gauthier-Villars, 1977.
- [8] Frey, G., “Links between stable elliptic curves and certain Diophantine equations”, *Ann. Univ. Saraviensis, Ser. Math.*, 1 (1986), 1-40.
- [9] Frey, G., “Links between solutions of $A - B = C$ and elliptic curves”, *Number Theory, Ulm1987, Proceedings* (H.P. Schlickewei, E. Wirsing, eds), *Lecture Notes in Mathematics*, 1380, Springer-Verlag, New York, 31-62, 1989.
- [10] Grothendieck, A., Serre, J-P., *Correspondance Grothendieck-Serre*, (P. Colmez, J-P. Serre eds), Société Mathématique de France, Paris, 2001.
- [11] Hellegouarch, Y., “Points d’ordre $2p^h$ sur les courbes elliptiques”, *Acta. Arith.*, 26 (1974/75), 253-263.
- [12] Hellegouarch, Y., *Invitation to the Mathematics of Fermat-Wiles*. San Diego, Academic Press, 2002.
- [13] Knapp, A., *Elliptic curves*, Princeton University Press, 1992.
- [14] Mazur, B., “Modular curves and the Eisenstein ideal”, *Publ. Math. IHES*, 47 (1977), 33-186.
- [15] Mazur, B., “Deforming Galois representations”, *Galois groups over \mathbb{Q}* (Y. Ihara, K. Ribet, J.-P. Serre, eds.), *Math. Sci. Res. Inst. Publ.*, 16, Springer-Verlag, New York, 385-437, 1989.
- [16] Oesterlé, J., “Nouvelles approches du théorème de Fermat”, *Séminaire Bourbaki* 694(1987-1988), *Astérisque* 161/162, 165-186, 1988.

- [17] Oesterlé, J., “Travaux de Wiles (et Taylor)”, II, *Séminaire Bourbaki, Astérisque*, 237 (1996), 333-355. COMPLETE
- [18] Murty, R. M., “Selberg’s conjectures and Artin L -functions”, *Bulletin of the AMS*, 31, 1, 1-14, 1994.
- [19] Murty, V.K., “Modular elliptic curves”, *Seminar on Fermat’s Last Theorem*, Canadian Math. Soc. Conf. Proc., 17, 1995. COMPLETE.
- [20] Ribet, K., “Galois Representations and Modular Forms”, *Bulletin of the AMS*, Oct. 1995, 375-402.
- [21] Ribet, K., “On modular representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms”, *Invent. Math.*, 100, 431-476, 1990.
- [22] Rubin, K., Silverberg, A., “A report on Wiles’ Cambridge lectures”, *Bulletin of the AMS*, 31, 1, 15-38, 1994.
- [23] Serre, J-P., “Propriétés galoisiennes des points d’ordre fini des courbes elliptiques”, *Invent. Math.*, 15 (1972), 259-331.
- [24] Serre, J-P., “Sur les représentations modulaires de degré 2 de $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ ”, *Duke Math. J.*, 54, 179-230, 1987.
- [25] Serre, J-P., “Travaux de Wiles (et Taylor)”, I, *Séminaire Bourbaki, Astérisque* 237 (1996), 319-332).
- [26] Silverman, J., Tate, J., *Rational Points on Elliptic Curves*, New York, Springer-Verlag, 1992.
- [27] Vojta, P., “Diophantine Approximations and Value Distribution Theory”, *Lec. Notes in Math.*, 1239, Springer, 1989.
- [28] Taylor, R., Wiles, A., “Ring-theoretic properties of certain Hecke algebras”, *Ann. of Math. (2)* 141, 553–572, 1995.
- [29] Wiles, A., “Modular elliptic curves and Fermat’s Last Theorem”, *Ann. of Math. (2)* 141, 443–551, 1995.