FERMAT, WILES ET GL(2)

par

Guy Henniart

PREMIÈRE PARTIE : FERMAT

Dans cette première partie, nous indiquons comment Andrew Wiles a démontré le « grand théorème » de Fermat. Le groupe GL(2) n'y intervient pas explicitement, mais il est partout sous-jacent, comme nous le verrons dans la seconde partie. Pour les références, voir « Modular Forms and Fermat's last theorem » édité par G. Cornell, J. Silverman & G. Stevens, Springer (1997) ou, en français, l'« Invitation aux mathématiques de Fermat-Wiles » par Y. Hellegouarch chez Masson en 1997. Le cours récent de P. Colmez à l'École polytechnique est également une bonne référence. Pour une introduction aux courbes elliptiques et formes modulaires, le cours d'arithmétique de Serre reste une merveille.

1. Le « grand théorème » de Fermat

Ce « grand théorème » (Last Theorem en anglais !) est un énoncé de Fermat, qui est resté une conjecture pendant 350 ans, jusqu'à sa démonstration par Andrew Wiles – avec la collaboration de Richard Taylor – en 1993. L'énoncé dit que si r est un entier, $r \geqslant 3$, toute solution en entiers a, b et c de l'équation $a^r + b^r = c^r$ est triviale, c'est-à-dire que a, b ou c est nul. Le cas où r vaut 4 a été traité par Fermat lui-même, celui où r vaut 3 par Euler.

La démonstration de Wiles procède par l'absurde : on suppose qu'on a une solution non triviale, $abc \neq 0$, et on suppose aussi, ce qui est loisible, que r est un nombre premier, $r \geqslant 5$, et que a, b et c sont premiers entre eux (notez que le cas où r est le double d'un nombre premier impair est dû à Terjanian, par une méthode relativement élémentaire, ce qui n'est pas du tout le cas de celle de Wiles).

La première étape consiste à utiliser une astuce d'Yves Hellegouarch, dont l'importance pour le problème a été mise au jour par Gerhard Frey. On associe à la solution non triviale $a^r + b^r = c^r$ de l'équation de Fermat une *courbe elliptique* E_{ab} d'équation

$$Y^2 = X(X - a^r)(X + b^r).$$

Une courbe elliptique E sur le corps des nombres rationnels est donnée par une équation $Y^2 = f(X)$ où f est un polynôme unitaire de degré 3 à coefficients rationnels, n'ayant pas de racine complexe double, de sorte que son discriminant $\mathrm{disc}(f)$ n'est pas nul : dans le cas de E_{ab}, où $f(X) = X(X - a^r)(X + b^r)$, le discriminant vaut $(abc)^{2r}$. L'origine de l'étude des courbes elliptiques remonte également à Fermat et à sa méthode de descente infinie.

Le point majeur de la démarche de Frey est que la courbe elliptique E_{ab} aurait de tellement bonnes propriétés qu'elle ne peut pas exister. Ce lien entre l'équation de départ et une courbe elliptique est absolument crucial : auparavant il n'était pas clair que le « grand théorème de Fermat » fût vrai ; après les travaux de Kummer, on disposait de certains critères, et les ordinateurs avaient permis de traiter les cas où r vaut au plus 125000, ce qui est petit ! Faire intervenir les courbes elliptiques faisait entrer dans un grand faisceau de conjectures, très cohérentes et étayées de très nombreux cas particuliers. C'est en prouvant une de ces conjectures majeures, la conjecture de Taniyama-Weil, que Wiles a démontré l'énoncé de Fermat. Expliquons cela.

2. Courbes elliptiques

Il vaut mieux considérer une courbe elliptique E sur $\mathbb{Q}$ comme une courbe dans le plan projectif, autrement dit remplacer l'équation $Y^2 = f(X)$ par l'équation homogène de degré 3 : $Y^2Z = Z^3 f(X/Z)$;

les points rationnels de E sont alors les points du plan projectif sur $\mathbb{Q}$ qui vérifient cette équation, c'est-à-dire les solutions (x, y, z) en nombres rationnels non tous nuls, à multiplication près par un même nombre rationnel non nul : aux solutions (x, y) de l'équation initiale, donnant les solutions projectives $(x, y, 1)$, on a simplement ajouté un « point à l'infini » de coordonnées homogènes $(0, 1, 0)$. Bien sûr, on peut de cette façon considérer les points de E dans le plan projectif sur tout corps contenant $\mathbb{Q}$, par exemple le corps des nombres réels, le corps des nombres complexes ou le corps des nombres p-adiques pour un nombre premier p.

L'avantage de considérer une courbe projective est qu'on dispose alors sur la courbe elliptique E d'une structure de groupe abélien : l'élément neutre O du groupe est le point à l'infini, et la loi de groupe est telle que trois points P, Q et R sont de somme O exactement quand ce sont les trois points d'intersection, comptés avec la multiplicité appropriée, de E avec une droite du plan projectif. L'opposé d'un point (x, y, z) est le point $(x, -y, z)$, de sorte que les points d'ordre 2 sont les points où y est nul.

Le théorème de Mordell dit que le groupe abélien $E(\mathbb{Q})$ des points rationnels de E est de type fini : il est isomorphe à $\mathbb{Z}^r$ fois un groupe abélien fini, où r est un entier positif appelé le *rang* de E sur $\mathbb{Q}$.

Le groupe abélien $E(\mathbb{C})$, lui, est isomorphe au quotient du groupe additif $\mathbb{C}$ par un réseau de $\mathbb{C}$.

On peut toujours faire un léger changement de variables dans une équation définissant une courbe elliptique E sur $\mathbb{Q}$ pour que les coefficients de l'équation deviennent entiers ; réduisant ces coefficients modulo un nombre premier p, on obtient alors l'équation d'une courbe sur le corps $\mathbb{Z}/p\mathbb{Z}$. Si p ne divise pas le discriminant de f, on obtient une courbe elliptique sur ce corps et on définit un entier $a_p = a_p(E)$ par la formule

$$a_p = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}).$$

Ici $E(\mathbb{Z}/p\mathbb{Z})$ est l'ensemble des points dans le plan projectif sur $\mathbb{Z}/p\mathbb{Z}$ qui donnent des solutions de l'équation homogène réduite modulo p ; c'est un groupe abélien fini. Si E était une droite dans le plan projectif, son nombre de points sur $\mathbb{Z}/p\mathbb{Z}$ serait $p+1$, de sorte que l'entier a_p mesure combien E est loin d'une droite.

Comme on l'a dit, on peut se permettre certains changements de variable dans l'équation f donnant la courbe elliptique E, tout en conservant au fond la « même » courbe elliptique sur $\mathbb{Q}$. En fait il existe une équation à coefficients entiers de discriminant minimal qu'on note $\mathrm{disc}(E)$; pour E_{ab} ce discriminant minimal vaut $2^{-8}(abc)^{2r}$, du moins pourvu qu'on ait choisi b pair et a congru à -1 modulo 4, ce qui est toujours possible.

Pour p premier ne divisant pas ce discriminant minimal, l'équation réduite modulo p donne donc une courbe elliptique $E \mod p$ sur $\mathbb{Z}/p\mathbb{Z}$, et un entier $a_p(E)$; on sait que $|a_p(E)| \leqslant 2\sqrt{p}$. Pour un mauvais nombre premier p – ce sont les diviseurs de $\mathrm{disc}(E)$ – l'équation réduite modulo p ne définit plus une courbe elliptique ; on obtient une courbe singulière qui possède un point double (réduction multiplicative) ou un point de rebroussement (réduction additive). Dans le cas de réduction additive, on pose $a_p(E) = 0$ et dans le cas de réduction multiplicative on pose $a_p(E) = 1$ ou $a_p(E) = -1$ selon que les tangentes en le point double ont leur pente dans $\mathbb{Z}/p\mathbb{Z}$ ou pas ; dans tous les cas a_p vaut $p + 1$ moins le nombre de points non singuliers dans la courbe sur $\mathbb{Z}/p\mathbb{Z}$.

A l'aide des entiers $a_p = a_p(E)$ on forme une série de Dirichlet, sur le modèle de la fonction Zêta de Riemann définie par

$$\zeta(s) = \prod_{p \text{ premier}} (1 - p^{-s})^{-1} = \sum_{n \geqslant 1} n^{-s}$$

pour les nombres complexes s vérifiant $\mathrm{Re}(s) > 1$.

On obtient ainsi la fonction L de Hasse-Weil de E

$$L(E, s) = \prod_{p \text{ mauvais}} (1 - a_p\, p^{-s})^{-1} \prod_{p \text{ bon}} (1 - a_p\, p^{-s} + p^{1-2s})^{-1},$$

qu'on écrit $\sum_{n \geqslant 1} a_n\, n^{-s}$ où les a_n sont des entiers, a_1 valant 1. La borne $|a_p| \leqslant 2\sqrt{p}$ implique qu'une telle série de Dirichlet converge absolument pour s vérifiant $\mathrm{Re}(s) > 3/2$. Comme on va le voir, $L(E, s)$ se prolonge en une fonction holomorphe dans le plan complexe tout entier. L'un des problèmes du millénaire, dont la solution est récompensée par un million de dollars, est le suivant.

***Conjecture* (Birch et Swinnerton-Dyer).** *En $s = 1$, $L(E, s)$ a un zéro d'ordre le rang de E sur $\mathbb{Q}$.*

La conjecture de Taniyama-Weil, dont l'origine remonte aux années 1950, donne plus encore que l'holomorphie de $L(E,s)$. Elle a été prouvée par Wiles et Taylor si tous les diviseurs premiers de $\mathrm{disc}(E)$ sont de réduction multiplicative, et ensuite le cas général a été obtenu par Breuil, Conrad, Diamond et Taylor.

Dans l'énoncé, $\mathfrak{H}$ désigne le demi-plan de Poincaré, formé des nombres complexes de partie imaginaire strictement positive, et pour z dans $\mathfrak{H}$ on pose $q = \exp(2\pi i z)$.

***Théorème* (« Conjecture de Taniyama-Weil »).** *Soit E une courbe elliptique sur $\mathbb{Q}$ et $L(E,s) = \sum_{n\geqslant 1} a_n n^{-s}$ sa fonction L. Alors $\sum_{n\geqslant 1} a_n q^n$ est le développement en q d'une forme modulaire de poids 2 pour $\Gamma_0(N)$, pour un certain entier $N \geqslant 1$.*

Ce que sont ces formes modulaires sera expliqué au prochain paragraphe. En tout cas ce théorème, à cause des propriétés connues des formes modulaires, donne que $L(E,s)$ a bien un prolongement holomorphe à $\mathbb{C}$ tout entier, et même que la fonction

$$\Lambda(E,s) = (2\pi)^{-s}\Gamma(s)L(E,s),$$

où Γ est la fonction gamma d'Euler, a une équation fonctionnelle

$$\Lambda(E,s) = \pm N^{(s-1/2)}\Lambda(E,2-s).$$

On peut d'ailleurs préciser cet entier N qui figure dans le théorème et l'équation fonctionnelle ci-dessus : c'est le *conducteur* N_E de la courbe elliptique, un diviseur de $\mathrm{disc}(E)$ qui peut se calculer à partir d'une équation pour E. Plus précisément l'exposant dans N_E d'un diviseur premier p de $\mathrm{disc}(E)$, c'est-à-dire un nombre premier de mauvaise réduction, peut se calculer en regardant les points de E sur le corps $\mathbb{Q}_p$ des nombres p-adiques ; cet exposant vaut entre 1 et 8, et il vaut 1 si la réduction est multiplicative. Pour E_{ab} on trouve le conducteur $N_{ab} = \prod_{p|abc} p$.

3. Formes modulaires

Les formes modulaires sont apparues dès le dix-neuvième siècle et sont reliées à beaucoup de problèmes d'arithmétique – voir le cours d'arithmétique de Serre.

Ce sont des fonctions sur le demi-plan de Poincaré $\mathfrak{H}$, à valeurs complexes, qui sont holomorphes sur $\mathfrak{H}$ et vérifient un certain nombre d'autres propriétés.

Remarquons que le groupe G^+ des matrices de $M(2,\mathbb{R})$ de déterminant >0 agit sur $\mathfrak{H}$ par la loi d'action $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z = (az+b)/(cz+d)$; en particulier le sous-groupe $\mathrm{SL}(2,\mathbb{Z})$ de G^+ agit ainsi sur $\mathfrak{H}$. Pour un entier $N \geqslant 1$, on note $\Gamma_0(N)$ le sous-groupe de $\mathrm{SL}(2,\mathbb{Z})$ formé des matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ où c est multiple de N.

Si k est un entier positif, une forme modulaire de poids k pour $\Gamma = \Gamma_0(N)$ est une fonction holomorphe de $\mathfrak{H}$ dans $\mathbb{C}$ vérifiant certaines équations fonctionnelles et des conditions de croissance. Les équations fonctionnelles imposent que pour $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ dans Γ on ait l'égalité

$$(cz+d)^{-k} f\Big(\frac{az+b}{cz+d}\Big) = f(z);$$

pour $k=2$, cela s'exprime aussi en disant que $f(z)dz$ est une forme différentielle sur $\mathfrak{H}$ invariante par l'action de Γ. En particulier, prenant $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, on trouve $f(z+1) = f(z)$, de sorte que f possède un développement en $q = \exp(2\pi i z)$,

$$f(z) = \sum_{n\in\mathbb{Z}} a_n q^n.$$

On impose alors la condition $a_m = 0$ pour $m<0$, ainsi que les conditions analogues pour les fonctions $(cz+d)^{-k} f\big((az+b)/(cz+d)\big)$, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ parcourant $\mathrm{SL}(2,\mathbb{Z})$. On obtient alors un espace vectoriel $M_k(N)$ de dimension finie sur $\mathbb{C}$. Si on impose en outre la condition $a_0 = 0$ – et de même pour les fonctions $(cz+d)^{-k} f\big((az+b)/(cz+d)\big)$, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ dans $\mathrm{SL}(2,\mathbb{Z})$ – on obtient le sous-espace vectoriel $S_k(N)$ des formes *paraboliques*.

Exemples.

– Pour $N=1$, on a $S_k(1) = 0$ pour $k \leqslant 11$ et $S_{12}(1) = \mathbb{C}\Delta$ où $\Delta = q\prod_{n\geqslant 1}(1-q^n)^{24}$.

– Pour $k=2$, on a $S_2(N) = 0$ pour $N \leqslant 10$ et $S_2(11) = \mathbb{C} f_{11}$ avec $f_{11} = q\sum_{n\geqslant 1}(1-q^n)^2(1-q^{11n})^2$.

Sur l'espace $S_k(N)$ agissent les opérateurs de Hecke T_p pour p premier ne divisant pas N (il existe aussi de tels opérateurs pour les autres nombres premiers, ils ne nous serviront pas directement ici) ;

pour $f = \sum_{n\geqslant 1} a_n q^n$ on a $T_p f = \sum_{n\geqslant 1} b_n q^n$ avec $b_n = a_{np} + p^{k-1} a_{n/p}$ si p divise n et $b_n = a_{np}$ sinon. On sait que $S_k(N)$ possède une base de vecteurs propres pour les T_p. Certaines de ces formes propres sont « anciennes » au sens où elles proviennent de formes de $S_k(M)$ pour un diviseur strict M de N ; on appelle « nouvelles » les formes propres qui ne sont pas anciennes. Si $f = \sum_{n\geqslant 1} a_n q^n$ est une forme propre, la valeur propre α_p correspondant à T_p vérifie $a_p = \alpha_p a_1$; une forme propre est dite *normalisée* si on a $a_1 = 1$.

La démonstration de Wiles associe à une courbe elliptique E sur $\mathbb{Q}$ une forme modulaire parabolique de poids 2 pour $\Gamma_0(N_E)$ qui est une forme propre nouvelle et normalisée, son coefficient en un nombre premier p de bonne réduction étant $a_p(E)$. Déjà dans les années 1960–70, Shimura savait procéder en sens inverse. Si l'on part d'une forme modulaire parabolique $f = \sum_{n\geqslant 1} a_n q^n$, de poids 2 pour $\Gamma_0(N)$, qui est propre pour les T_p, p ne divisant pas N, qui est en outre nouvelle et normalisée, et qui enfin a des coefficients a_n rationnels – ils sont alors entiers – alors il existe une courbe elliptique E_f sur $\mathbb{Q}$ telle que $L(E_F, s) = L(f, s)$.

Par exemple, pour $N = 11$ et $f = f_{11}$ dans l'exemple plus haut, on obtient $E_{f_{11}}$ en ajoutant à $\Gamma_0(11)\backslash\mathfrak{H}$ deux points à l'infini ; la courbe correspondante a pour équation

$$Y^2 + Y = X^3 - X^2 - 10X - 20.$$

Noter qu'il reste difficile de trouver l'équation de E_f à partir de celle de f ; noter aussi que f ne détermine pas E_f à isomorphisme près, mais, grâce à un théorème de Faltings, f détermine la « classe d'isogénie » de E_f.

4. Congruences de formes modulaires

La preuve par Wiles de l'énoncé de Fermat repose alors sur des considérations de congruence entre formes modulaires, basées sur un théorème antérieur de Ribet. Plus précisément, quelque sept ans avant la preuve de Wiles, les considérations de G. Frey avaient amené Serre à formuler des conjectures de congruence précises qui formaient un pas vers le grand théorème de Fermat. Ribet a alors prouvé ces conjectures.

***Théorème* (K. Ribet).** *Soit r un nombre premier. Soit $f = \sum_{n\geqslant 1} a_n q^n$ une forme parabolique de poids 2 pour $\Gamma_0(N)$ qui est nouvelle, propre et normalisée. Soit M l'entier obtenu en enlevant de N tous les nombres premiers tels que E_f ait mauvaise réduction multiplicative en p et que $v_p(\mathrm{disc}(E_f))$ soit* multiple de r. *Alors il existe une forme $g = \sum_{n\geqslant 1} b_n q^n$ parabolique de poids 2 pour $\Gamma_0(M)$, telle que pour p premier ne divisant pas Nr on ait*

$$b_p \equiv a_p \mod r.$$

Appliquant cela à la forme modulaire f_{ab} associée à la courbe elliptique E_{ab}, on a $N_{ab} = \prod_{p|abc} p$ et $\mathrm{disc}(E_{ab}) = 2^{-8}\prod_{p|abc} p^{2r}$. On trouve alors $M = 2$. Mais $S_2(2) = 0$ donne la contradiction voulue !

Pour ne pas rester dans le domaine de l'absurde, voici un exemple de congruence donnée par le résultat de Ribet. On considère les courbes elliptiques E_1, d'équation

$$Y^2 = X^3 - X^2 - 77X + 330,$$

de discriminant $\Delta_1 = 2^4 \cdot 3^{10} \cdot 11$ et de conducteur $N_1 = 2^2 \cdot 3 \cdot 11$, et la courbe elliptique E_2, d'équation

$$Y^2 = X^3 + X^2 + 3X - 1,$$

de conducteur $N_2 = 2^2 \cdot 11$.

Si l'on applique le résultat de Ribet à f_{E_1}, pour $r = 5$, on trouve $M = 2^2 \cdot 11$ et en fait on peut prendre $g = f_{E_2}$. Voici un petit tableau illustratif.

p	2	3	5	7	11	13	17
$a_p(E_1)$	0	-1	2	2	-1	6	-4
$a_p(E_2)$	0	1	-3	2	-1	-4	6

DEUXIÈME PARTIE : GL(2)

Dans ce qui précède, le groupe GL(2) intervient à peine. Il est pourtant partout sous-jacent, et de manière fondamentale, tant du côté des formes modulaires que du côté des courbes elliptiques.

Comme on le verra, une forme modulaire parabolique de poids k pour $\Gamma_0(N)$ nouvelle, propre et normalisée, donne naissance à une représentation unitaire irréductible de $\mathrm{GL}(2,\mathbb{R})$, qui reflète le poids de f, et pour chaque nombre premier p à une représentation lisse irréductible unitarisable de $\mathrm{GL}(2,\mathbb{Q}_p)$, qui pour p ne divisant pas N reflète le coefficient a_p de $f = \sum_{n\geqslant 1} a_n q^n$.

Par ailleurs, notons $\overline{\mathbb{Q}}$ le corps des nombres algébriques ; c'est la clôture algébrique de $\mathbb{Q}$ dans $\mathbb{C}$. Une courbe elliptique E sur $\mathbb{Q}$ donne pour chaque nombre premier ℓ une représentation de $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ – le groupe des automorphismes du corps $\overline{\mathbb{Q}}$ – dans $\mathrm{GL}(2,\mathbb{Z}_\ell)$, et ainsi pour tout nombre premier p, distinct ou non de ℓ, une représentation dans $\mathrm{GL}(2,\mathbb{Z}_\ell)$ de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ où $\overline{\mathbb{Q}}_p$ est une clôture algébrique de $\mathbb{Q}_p$.

Le lien entre E et f_E, f et E_f de la première partie est reflété par la correspondance de Langlands qui pour ℓ différent de p relie les représentations lisses irréductibles de $\mathrm{GL}(2,\mathbb{Q}_p)$ aux représentations de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ dans $\mathrm{GL}(2,\overline{\mathbb{Q}}_\ell)$. Énoncée dans les années 1960, cette correspondance, un cas très particulier de l'immense faisceau de conjectures dû à Langlands, a été prouvée en 1978 par Ph. Kutzko. Noter que dans le cas où ℓ et p sont égaux, il y a bien un lien aussi, mais il n'a été élucidé que récemment, et il faut élargir considérablement le cadre des représentations lisses de $\mathrm{GL}(2,\mathbb{Q}_p)$.

5. Le côté galoisien : modules de Tate

Partons d'une courbe elliptique E sur $\mathbb{Q}$. Comme on l'a dit le groupe abélien $E(\mathbb{C})$ est isomorphe au quotient de $\mathbb{C}$ par un réseau Λ. Pour m entier, $m \geqslant 1$, notons $E[m]$ l'ensemble des points P de $E(\mathbb{C})$ vérifiant $mP = 0$. Alors le groupe $E[m]$ est isomorphe à $(\mathbb{Z}/m\mathbb{Z})^2$. Comme l'addition dans E est donnée par des formules polynomiales à coefficients rationnels, les coordonnées des points de $E[m]$ sont

algébriques et $E[m]$ est inclus dans $E(\overline{\mathbb{Q}})$. En fait le groupe de Galois $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, via son action sur $\overline{\mathbb{Q}}$, opère sur $E[m]$, ce qui donne une représentation[(1)] $G_{\mathbb{Q}} \to \mathrm{GL}(2, \mathbb{Z}/m\mathbb{Z})$.

On peut prendre pour m un nombre premier ℓ ou une puissance ℓ^k de ce nombre premier, et bien sûr on retrouve la représentation modulo ℓ^k en réduisant modulo ℓ^k la représentation modulo ℓ^{k+1}. Passant à la limite, on obtient une représentation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}(2, \mathbb{Z}_\ell)$; par réduction modulo ℓ, on a la représentation $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}(2, \mathbb{Z}/\ell\mathbb{Z})$ donnée par l'action sur les points d'ordre ℓ.

La représentation ρ permet de retrouver les entiers $a_p = a_p(E)$ pour p premier ne divisant pas $N_E\ell$: pour un tel p, on dispose d'une matrice A_p dans $\rho(G_{\mathbb{Q}})$ dont la trace est l'entier a_p – vu comme un élément de $\mathbb{Z}_\ell$ – et le déterminant p. Ainsi ρ code les entiers a_p, donc la fonction $L(E, s)$, et de même $\overline{\rho}$ code les a_p modulo ℓ.

C'est via ces représentations ℓ-adiques que les congruences interviennent pour la courbe elliptique E.

La méthode de Wiles, résumée de façon outrageusement simplifiée, est la suivante. On part de E, on choisit un nombre premier ℓ d'où une représentation ℓ-adique ρ comme plus haut, et sa réduction $\overline{\rho}$ modulo ℓ. On considère l'ensemble des représentations ℓ-adiques de $G_{\mathbb{Q}}$ qui proviennent de courbes elliptiques et donnent $\overline{\rho}$ en réduction, et le sous-ensemble formé de celles qui proviennent de formes modulaires de poids 2. Il s'agit de prouver l'égalité de ces deux ensembles. Le point crucial est que, dans certaines circonstances bien particulières, Wiles a été capable de prouver l'égalité pourvu que le second ensemble ne soit pas vide. Mais on ne sait pas cela, a priori, pour la représentation $\overline{\rho}$ de départ. Cependant on peut utiliser les cas existants, et la possibilité de changer ℓ, pour augmenter progressivement les circonstances où on a bien égalité. Mais si E est donnée, il n'est pas clair comment la relier aux « cas existants ». Le point de départ est de prendre $\ell = 3$; comme $\mathrm{GL}(2, \mathbb{Z}/3\mathbb{Z})$ a des représentations complexes fidèles de dimension 2 (voir le premier texte de ce volume), la représentation $\overline{\rho}$ peut se voir comme représentation complexe de $G_{\mathbb{Q}}$, et un théorème prouvé par Langlands et Tunnell dans les années 1970,

[(1)] On utilise le mot « représentation » dans un sens plus large qu'aux exposés précédents.

fournit une forme modulaire dont les coefficients b_p proviennent de $\overline{\rho}$. Mais cette forme modulaire est de poids 1...

6. Le côté galoisien : représentations locales

Dans la situation précédente, la représentation ℓ-adique ρ de $G_{\mathbb{Q}}$ attachée à E contient bien plus d'information que la donnée des coefficients a_p. Par exemple, fixons un nombre premier p quelconque. Les équations donnant les coordonnées des points de $E[\ell^k]$ étant à coefficients rationnels, on peut aussi les considérer comme des équations à coefficients p-adiques. Choisissant une clôture algébrique $\overline{\mathbb{Q}}_p$ de $\mathbb{Q}_p$, cela donne une action du groupe $G_{\mathbb{Q}_p}$, le groupe des $\mathbb{Q}_p$-automorphismes de $\overline{\mathbb{Q}}_p$, sur les points de $E(\overline{\mathbb{Q}}_p)$ vérifiant $\ell^k P = O$; on en tire une représentation $\rho_p : G_{\mathbb{Q}_p} \to \mathrm{GL}(2, \mathbb{Z}_\ell)$ qui est bien définie à isomorphisme près et peut s'obtenir à partir de ρ par un plongement adéquat de $G_{\mathbb{Q}_p}$ dans $G_{\mathbb{Q}}$. Pour p ne divisant pas $N_E \ell$, $\rho_p(G_{\mathbb{Q}_p})$ est engendré topologiquement par la matrice A_p du paragraphe précédent. Si p divise N_E mais est distinct de ℓ, $\rho_p(G_{\mathbb{Q}_p})$ est plus compliqué mais pas trop car $G_{\mathbb{Q}_p}$ contient un gros sous-groupe distingué, le groupe de ramification sauvage, qui est un pro-p-groupe et dont l'image par ρ_p est finie. Si $p = \ell$, alors $\rho_p(G_{\mathbb{Q}_p})$ peut être très compliquée quand l'image du groupe de ramification sauvage est grosse.

7. Le côté automorphe

Partons cette fois d'une forme modulaire $f = \sum_{n \geqslant 1} a_n q^n$ de poids k pour $\Gamma_0(N)$, nouvelle, propre et normalisée. On peut lui associer une fonction φ_f sur G^+ en posant

$$\varphi_f(g) = f(g(i))(ci+z)^{-k}(\det g)^{-k/2} \quad \text{pour} \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{dans} \quad G^+.$$

On obtient ainsi une « forme automorphe » pour G^+ au sens où

- $\varphi(\gamma g) = \varphi(g)$ pour γ dans $\Gamma_0(N)$,
- $\varphi\left(g\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)\right) = e^{-ik\theta}\varphi(g)$ pour $\theta \in \mathbb{R}$,

et où φ vérifie des conditions de croissance et de parabolicité traduisant celles de f. L'avantage de passer de fonctions sur $\mathfrak{H}$ à des

fonctions sur le groupe G^+ est que l'on peut faire agir le groupe G^+ ! Par translations à droite, la fonction φ_f engendre en fait une représentation unitaire irréductible de G^+ ; pour $k \geqslant 2$ c'est la « série discrète » de poids k – voir le texte de M. Andler.

Pour obtenir une interprétation en termes de représentations de groupe, des opérateurs de Hecke T_p, il faut faire intervenir les groupes $\mathrm{GL}(2, \mathbb{Q}_p)$ en sus de G^+ ou $\mathrm{GL}(2, \mathbb{R})$.

Plus précisément on construit l'anneau des adèles $\mathbb{A}$, produit *restreint* de $\mathbb{R}$ et des $\mathbb{Q}_p$: cela signifie que $\mathbb{A}$ est la partie du produit formée des familles (x_∞, x_p) telles que x_p soit dans $\mathbb{Z}_p$ sauf pour un nombre fini de nombres premiers p. L'anneau $\mathbb{A}$ a une topologie naturelle pour laquelle $\mathbb{R} \prod_p \mathbb{Z}_p$, avec sa topologie produit, est un sous-anneau ouvert. Comme $\mathbb{R}$ et les $\mathbb{Q}_p$ contiennent $\mathbb{Q}$, l'anneau $\mathbb{A}$ contient une copie diagonale de $\mathbb{Q}$ qui est un sous-groupe discret de $\mathbb{A}$. Le quotient $\mathbb{A}/\mathbb{Q}$ est compact, et est vraiment l'équivalent « arithmétique » de $\mathbb{R}/\mathbb{Z}$. De manière analogue, le groupe $\mathbb{A}^\times$ des éléments inversibles de $\mathbb{A}$ a une topologie naturelle, il contient $\mathbb{Q}^\times$ comme sous-groupe discret, et le quotient est extension de $\mathbb{R}$ par un groupe compact.

On considère ici le groupe $\mathrm{GL}(2, \mathbb{A})$, produit restreint de $\mathrm{GL}(2, \mathbb{R})$ et des $\mathrm{GL}(2, \mathbb{Q}_p)$: la restriction est par rapport aux sous-groupes $\mathrm{GL}(2, \mathbb{Z}_p)$. Il contient $\mathrm{GL}(2, \mathbb{Q})$ comme sous-groupe discret.

Rappelons l'entier N fixé au début du paragraphe. On lui associe, pour chaque nombre premier p, un sous-groupe K_p de $\mathrm{GL}(2, \mathbb{Z}_p)$ formé des matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ telles que c soit multiple de N dans $\mathbb{Z}_p$ – on a $K_p = \mathrm{GL}(2, \mathbb{Z}_p)$ si p ne divise pas N. Tout élément g de $\mathrm{GL}(2, \mathbb{A})$ peut alors s'écrire

$$g = \gamma \; g_\infty \prod_p k_p,$$

avec $\gamma \in \mathrm{GL}(2, \mathbb{Q})$, $g_\infty \in G^+$ et $k_p \in K_p$ pour p premier.

Remarquant qu'on a

$$\mathrm{GL}(2, \mathbb{Q}) \cap \Big(G^+ \prod_p K_p \Big) = \Gamma_0(N),$$

on voit que pour $g \in \mathrm{GL}(2, \mathbb{A})$ écrit comme précédemment, la formule $\Phi_f(g) = \varphi_f(g_\infty)$ donne une fonction bien définie sur $\mathrm{GL}(2, \mathbb{A})$, invariante à gauche par $\mathrm{GL}(2, \mathbb{Q})$: on a $\Phi_f(\gamma g) = \Phi_f(g)$ pour γ dans $\mathrm{GL}(2, \mathbb{Q})$. Cela donne en fait une fonction de carré intégrable sur

$\mathrm{GL}(2,\mathbb{Q})\mathbb{A}^\times \backslash \mathrm{GL}(2,\mathbb{A})$, où on voit $\mathbb{A}^\times$ comme le centre de $\mathrm{GL}(2,\mathbb{A})$ formé des matrices scalaires.

Via l'action de $\mathrm{GL}(2,\mathbb{A})$ par translations à droite, la fonction φ_f engendre une sous-représentation unitaire irréductible Π_f de $L^2(\mathrm{GL}(2,\mathbb{Q})\mathbb{A}^\times \backslash \mathrm{GL}(2,\mathbb{A}))$; la fonction φ_f en donne un vecteur fixé par $\prod_{p \text{ premier}} K_p$.

Utilisant le fait que $\mathrm{GL}(2,\mathbb{A})$ est presque le produit de $\mathrm{GL}(2,\mathbb{R})$ et des $\mathrm{GL}(2,\mathbb{Q}_p)$, on peut « décomposer » Π_f en produit tensoriel de représentations π_∞ et π_p pour p premier, où π_∞ est une représentation unitaire irréductible de $\mathrm{GL}(2,\mathbb{R})$ (de la série discrète si $k \geqslant 2$) et π_p une représentation lisse irréductible de $\mathrm{GL}(2,\mathbb{Q}_p)$ unitarisable au sens où son espace porte un produit scalaire invariant. De plus par construction, pour p premier ne divisant pas N, π_p est non ramifiée, φ_f fournissant un vecteur non nul fixé par $\mathrm{GL}(2,\mathbb{Z}_p)$, et le coefficient a_p de la forme modulaire f de départ, n'est autre que la valeur propre, sur la droite formée de tels vecteurs, de l'opérateur sur π_p attaché à la double classe $\mathrm{GL}(2,\mathbb{Z}_p)\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)\mathrm{GL}(2,\mathbb{Z}_p)$. En revanche, pour p divisant N, π_p peut être compliquée, par exemple l'une des représentations cuspidales montrées par C. Blondel.

8. La correspondance de Langlands

Une idée fondamentale de Langlands est que le lien entre une courbe elliptique E et la forme modulaire f_E doit être vu comme le lien entre la représentation ℓ-adique ρ attachée à E (pour un nombre premier ℓ fixé, ou pour tous les ℓ à la fois), et la représentation automorphe Π_f.

Nous allons expliquer précisément ce lien, mais mentionnons que le cadre de l'heuristique de Langlands est bien plus général. En particulier on espère relier, via les représentations galoisiennes ou des objets plus généraux, les représentations « automorphes » de $G(\mathbb{A})$ où G est un groupe classique sur $\mathbb{Q}$ – ou plus généralement sur un corps de nombres – et des objets géométriques, variétés ou « motifs » sur $\mathbb{Q}$ ou les corps de nombres. Si l'on part d'une forme modulaire parabolique f d'un poids quelconque k pour $\Gamma_0(N)$, on sait lui associer des représentations ℓ-adiques, et même, si $k = 1$, une représentation de $G_\mathbb{Q}$ dans $\mathrm{GL}(2,\mathbb{C})$ d'image finie. On peut parfois aller en sens inverse,

et, comme on l'a vu au §5, c'est un résultat particulier de ce genre, en poids 1, dû à Langlands et Tunnell, qui forme l'un des points de départ de la preuve de Wiles !

Revenons au lien précis entre les représentations ρ et Π_f attachées à la forme modulaire f de départ. L'idée est qu'il doit y avoir une recette, au moins pour chaque nombre premier p distinct de ℓ, pour passer de ρ_p, la « restriction » de ρ à $G_{\mathbb{Q}_p}$ à π_p la « composante » en p de Π_f et réciproquement. Noter que cela n'a rien d'évident : si ρ détermine Π_f et réciproquement, il n'y a pas de raison a priori que le lien entre ρ_p et π_p soit indépendant des autres composantes. Cependant, pour p ne divisant pas $N\ell$, la recette est très simple :

– ρ_p est donnée par la matrice A_p déjà rencontrée, à conjugaison près ; autrement dit, ρ_p est déterminée par la trace a_p de cette matrice et son déterminant p.

– de même π_p est la représentation lisse irréductible non ramifiée de $\mathrm{GL}(2,\mathbb{Q}_p)$ dont le caractère central prend en p la valeur p^{-1} et sur laquelle l'opérateur de Hecke associé à $\mathrm{GL}(2,\mathbb{Z}_p)\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)\mathrm{GL}(2,\mathbb{Z}_p)$ a pour trace a_p.

Il n'est nullement clair que cela se généralise si p divise N, $p \neq \ell$ mais, après des efforts de nombreux auteurs, cela a été prouvé par Ph. Kutzko, nous l'avons dit.

Pour terminer, regardons à nouveau le cas où $p = \ell$. On dispose toujours des représentations ρ_ℓ de $G_{\mathbb{Q}_\ell}$ et π_ℓ de $\mathrm{GL}(2,\mathbb{Q}_\ell)$, et d'une recette donnant π_ℓ à partir de ρ_ℓ (Fontaine), mais on ne peut reconstituer ρ_ℓ à partir de π_ℓ car ρ_ℓ est intrinsèquement plus riche. Cependant, si l'on est prêt à considérer des représentations – non lisses – de $\mathrm{GL}(2,\mathbb{Q}_p)$ sur des espaces de Banach p-adiques, on retrouve une « correspondance de Langlands » (travaux de Breuil, Colmez et bien d'autres).

G. Henniart, CNRS, Laboratoire de Mathématique d'Orsay, UMR 8628, Bâtiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex • *E-mail :* guy.henniart@math.u-psud.fr