Séminaire BOURBAKI Novembre 2005
58e année, 2005-2006, no 956, p. 99 à 121

LA CONJECTURE DE MODULARITÉ DE SERRE : LE CAS DE CONDUCTEUR 1 [d'après C. Khare]

par **Jean-Pierre WINTENBERGER**

INTRODUCTION

Soit $\overline{\mathbb{Q}}$ la clôture algébrique de $\mathbb{Q}$ dans $\mathbb{C}$. Notons $G_{\mathbb{Q}}$ le groupe de Galois de $\overline{\mathbb{Q}}/\mathbb{Q}$. Soit p un nombre premier et soit $\overline{\mathbb{F}_p}$ une clôture algébrique du corps à p éléments. Soit $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ une représentation continue (donc à image finie), irréductible et impaire ($\det(\overline{\rho}(c)) = -1$ où c est la conjugaison complexe). Nous appelons une telle représentation galoisienne une représentation de type S.

Soit $\overline{\mathbb{Q}_p}$ une clôture algébrique du corps des nombres p-adiques $\mathbb{Q}_p$. Eichler, Shimura, Deligne, Deligne et Serre ont associé aux formes modulaires (propres) pour les sous-groupes de congruence de $\mathrm{SL}_2(\mathbb{Z})$ des représentations galoisiennes $G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$, dont les réductions modulo p, lorsqu'elles sont irréductibles, sont de type S ([17, 19] ; pour le poids 2, on pourra voir l'appendice de Conrad dans [44]). La conjecture de Serre, qui apparaît pour la première fois en 1972 (p. 9 de [56]), dit que toute représentation de $G_{\mathbb{Q}}$ de type S est *modulaire i.e.* provient comme ceci d'une forme modulaire. Elle est énoncée dans [50] pour les représentations $\overline{\rho}$ non ramifiées en dehors de p (cas de niveau $N = 1$). Dans [52], Serre énonce pour tout niveau N ce que nous appelons la *forme forte* de la conjecture par opposition à sa *forme qualitative.* La forme forte précise le poids k, le niveau N et le caractère ϵ d'une forme primitive pour $\Gamma_1(N)$ dont provient $\overline{\rho}$.

Tate, en réponse à une lettre de Serre, prouve qu'il n'y a pas de représentation $G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_2})$ qui soit irréductible et non ramifiée hors de 2, prouvant ainsi la conjecture pour $p = 2$ et $N = 1$ (1973, [57]). La méthode de Tate, qui repose sur une minoration de discriminant, a été étendue par Serre au cas $p = 3$ et $N = 1$ (p. 710 de [51]), et, sous l'hypothèse de Riemann généralisée, par Brueggeman pour $p = 5$ ([10]).

Grâce aux travaux de Ribet, Mazur, Carayol, Gross, Coleman-Voloch, Edixhoven, Diamond, ..., on sait, pour $p \neq 2$, que la forme qualitative entraîne la forme forte. C'est un grand théorème sur lequel on trouvera d'excellents rapports dans [43, 25, 44] : un cas de ce théorème a permis de déduire Fermat de la conjecture de Taniyama-Weil !

La démonstration, à la suite de Wiles, que toute courbe elliptique sur $\mathbb{Q}$ est modulaire, donne la conjecture de Serre pour les représentations galoisiennes $\overline{\rho}$ provenant

des points d'ordre p des courbes elliptiques sur $\mathbb{Q}$. Pour ce faire, Wiles prouve des énoncés du type suivant (« MR » : modularité des relèvements ; § 3) : si $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ est *géométrique*, *i.e.* vérifie des propriétés de ramification convenables, et si la réduction de ρ est du type S et modulaire, alors ρ est modulaire ([67, 64]). De plus, Wiles utilise un argument de « changement de nombre premier » : pour prouver que la représentation galoisienne $\overline{\rho_5} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_5)$ sur les points d'ordre 5 de la courbe elliptique E est modulaire, Wiles prouve l'existence d'une courbe elliptique E' sur $\mathbb{Q}$ dont la représentation sur les points d'ordre 5 est isomorphe à $\overline{\rho_5}$, et la représentation $\overline{\rho_3}$ sur les points d'ordre 3 est irréductible. La représentation $\overline{\rho_3}$, dont l'image est résoluble, est modulaire d'après Langlands-Tunnell ([66]). Un théorème « MR » entraîne alors que la représentation 3-adique associée à E' est modulaire, donc aussi E' et $\overline{\rho_5}$.

À l'aide de théorèmes « MR » et d'un argument de « changement de nombre premier », Taylor prouve une version potentielle de la conjecture de Serre : une représentation de type S provient d'une forme modulaire de Hilbert après restriction à un corps de nombres totalement réel F ([61, 60] ; § 4). On est alors confronté à un problème de changement de base de F à $\mathbb{Q}$. Dans certains cas, on peut s'assurer que $F/\mathbb{Q}$ est résoluble, et alors le théorème de Langlands et Tunnell permet de prouver la conjecture ([38, 26]). Les théorèmes de Taylor, le théorème de changement de base résoluble d'Arthur-Clozel ([2]), et des arguments de Taylor ([63]) permettent à Dieulefait de prouver que, étant donnée une représentation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ possédant les propriétés d'une représentation galoisienne associée à une forme modulaire, il existe un système compatible (ρ_λ) de représentations ℓ-adiques de $G_{\mathbb{Q}}$ dont fait partie ρ ([23], § 6).

La conjecture de Serre entraîne qu'une représentation $\overline{\rho}$ de type S admet des relèvements $G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$, qui de plus sont géométriques. Ramakrishna, utilisant des techniques de déformations de représentations galoisiennes, le prouve dans de nombreux cas sans supposer $\overline{\rho}$ modulaire ([41]). Khare réalise que ces techniques de déformation et en particulier des résultats de Böckle ([6]), la version potentielle de la conjecture de Serre due à Taylor et les théorèmes de type « MR » pour les corps totalement réels, permettent d'obtenir des relèvements de représentations de type S qui ont des propriétés de ramification plus précises que celles obtenues par Ramakrishna, et qui sont prédites par la forme forte de la conjecture de Serre.

L'existence de relèvements avec ces propriétés de ramification précises et l'existence de systèmes compatibles, permettent d'utiliser la technique de changement de nombre premier de Wiles. Dans [36], la conjecture de Serre pour $N = 1$ est prouvée pour $p = 5, 7$, et pour les poids $k \leq 14, k \neq 10$. Des stratégies sont données pour la ramener dans le cas général à des énoncés « MR ». Utilisant des relèvements de poids 2 avec Nebentypus, Khare parvient à déduire des théorèmes « MR » connus le cas général de niveau (conducteur) 1 ([34] ; pour un « survey » : [35]).

Enfin signalons que ce cercle d'idées a permis de grands progrès dans la preuve de la conjecture d'Artin pour les représentations impaires $G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$ ([12, 62]).

Je remercie Böckle de m'avoir communiqué son preprint [7] qui m'a beaucoup aidé pour la rédaction du § 5. Je remercie Khare pour ses remarques sur une première version du texte.

1. CONJECTURES DE MODULARITÉ DE SERRE ET DE FONTAINE-MAZUR

1.1. Représentations galoisiennes associées aux formes modulaires

Pour N entier ≥ 1, soit $\Gamma_1(N)$ le groupe des matrices :

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ c \equiv 0 \bmod N,\ a \equiv d \equiv 1 \bmod N.$$

Soit k un entier ≥ 1. Soit $S_k(\Gamma_1(N))$ le $\mathbb{C}$-espace vectoriel des formes modulaires paraboliques de poids k et de niveau N. Une forme $f \in S_k(\Gamma_1(N))$ a un développement de Fourier à la pointe $i\infty$:

$$f(z) = \sum_{n=1}^{\infty} a_n q^n,\ q = e^{2\pi i z}.$$

Une forme primitive f est propre pour les opérateurs de Hecke T_n, n entier > 1, et pour les opérateurs diamant $\langle \overline{d} \rangle$, $\overline{d} \in (\mathbb{Z}/N\mathbb{Z})^*$. Elle est normalisée : $a_1 = 1$. Pour $n > 1$, a_n est la valeur propre de T_n. Notons $\epsilon : (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$ le caractère $\langle \overline{d} \rangle(f) = \epsilon(\overline{d}) f$. On a :

$$(1)\ :\ f\left(\frac{az+b}{cz+d}\right) = \epsilon(\overline{d})(cz+d)^k f(z),\ \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ c \equiv 0 \bmod N,$$

$\overline{d}$ étant l'image de d dans $(\mathbb{Z}/N\mathbb{Z})^*$. Les a_n et $\epsilon(\overline{d})$ engendrent un ordre de l'anneau des entiers d'une extension finie E_f de $\mathbb{Q}$ contenue dans $\mathbb{C}$; E_f est le corps des coefficients de f.

Eichler, Shimura, Deligne, et Deligne et Serre ont associé à f et à un plongement ι de E_f dans $\overline{\mathbb{Q}_p}$ une représentation galoisienne : $\rho_{f,\iota} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ qui est caractérisée à conjugaison près par les propriétés suivantes :

– $\rho_{f,\iota}$ est non ramifiée en dehors de $\{p\} \cup S_N$, S_N étant l'ensemble des nombres premiers qui divisent N ;

– pour $\ell \notin \{p\} \cup S_N$, si Frob_ℓ est un élément de $G_{\mathbb{Q}}$ qui relève le Frobenius, on a : $\mathrm{tr}(\rho_{f,\iota}(\mathrm{Frob}_\ell)) = \iota(a_\ell)$.

Soit $\chi_p : G_{\mathbb{Q}} \to \mathbb{Z}_p^*$ le caractère cyclotomique. Le déterminant de $\rho_{f,\iota}$ est $\epsilon \chi_p^{k-1}$, où nous avons identifié de la manière naturelle $(\mathbb{Z}/N\mathbb{Z})^*$ avec le groupe de Galois de l'extension cyclotomique $\mathbb{Q}(\mu_N)/\mathbb{Q}$.

La représentation $\rho_{f,\iota}$ est impaire. En effet, il n'existe pas de forme parabolique non nulle de poids k et de caractère ϵ si l'on n'a pas $\epsilon(-1)(-1)^{k-1} = -1$, comme on le voit en considérant la matrice $-\mathrm{id}$ dans (1).

La représentation $\rho_{f,\iota}$ est irréductible ([19, 42]).

Elle est *géométrique* : elle est non ramifiée en dehors d'un ensemble fini de nombres premiers et sa restriction au groupe de décomposition D_p est potentiellement semi-stable, au sens de la théorie de Fontaine ([29]). Ceci résulte, si $k \neq 1$, de ce que $\rho_{f,\iota}$ apparaît dans la cohomologie étale d'une variété algébrique et des théorèmes de comparaison p-adiques ([65]). Pour $k = 1$, $\overline{\rho}$ a une image finie et est donc aussi géométrique.

1.2. La conjecture de Fontaine et Mazur ([31])

Conjecture 1.1. — *Soit $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ une représentation p-adique impaire, irréductible et géométrique. Alors, il existe f, ι comme ci-dessus et un entier j tels que ρ soit isomorphe à $\rho_{f,\iota}$ tordue par χ_p^j.*

La forme f, (N, k, ϵ) et j sont bien déterminés par ρ (f à conjugaison galoisienne près). En effet, les poids de Hodge-Tate de ρ sont $(j, j + k - 1)$. Après torsion par χ_p^{-j}, on se ramène au cas $j = 0$. Soit, pour nombre premier ℓ, r_ℓ la représentation F-semi-simple du groupe de Weil-Deligne WD_ℓ à valeurs dans $\mathrm{GL}_2(\overline{\mathbb{Q}_p})$ qui est associée à la restriction de ρ au groupe de décomposition D_ℓ. Pour $\ell \neq p$, r_ℓ a été définie par Grothendieck et Deligne (§ 8 de [18]). Pour $\ell = p$, elle a été définie par Fontaine à partir de l'action de D_p sur le module de Dieudonné filtré associé à la représentation potentiellement semi-stable $\rho_{|D_p}$ ([27]).

À r_ℓ est associée la partie ℓ primaire N_ℓ du conducteur. On a $N_\ell = 1$ si et seulement si, soit $\ell \neq p$ et ρ est non ramifiée en ℓ, soit $\ell = p$ et $\rho_{|D_p}$ est cristalline. Le conducteur N est le produit des N_ℓ. Le caractère ϵ est défini par la formule $\det(\rho) = \epsilon\chi_p^{k-1}$.

Enfin f est déterminée par ρ par la formule $\mathrm{tr}(\rho(\mathrm{Frob}_\ell)) = \iota(a_\ell)$ pour ℓ premier à N et p, puisque l'on a pris soin de choisir f primitive. On peut aussi dire que la correspondance de Langlands locale associe à r_ℓ une représentation π_ℓ de $\mathrm{GL}_2(\mathbb{Q}_\ell)$. La représentation automorphe associée à f a pour composantes locales les π_ℓ, π_∞ étant la représentation de $\mathrm{GL}_2(\mathbb{R})$ correspondant aux formes modulaires de poids k.

1.3. La conjecture de Serre

Soient f et ι comme au 1.1. Notons $\overline{\mathbb{Z}_p}$ l'anneau des entiers de $\overline{\mathbb{Q}_p}$ et fixons un isomorphisme du corps résiduel de $\overline{\mathbb{Z}_p}$ avec $\overline{\mathbb{F}_p}$.

On peut conjuguer $\rho_{f,\iota}$ de sorte que $\rho_{f,\iota}$ soit à valeurs dans $\mathrm{GL}_2(\overline{\mathbb{Z}_p})$. Par l'homomorphisme de réduction $\overline{\mathbb{Z}_p} \to \overline{\mathbb{F}_p}$, on en déduit une représentation $G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$. La semi-simplifiée de cette représentation est bien déterminée à isomorphisme près. On la note $\overline{\rho_{f,\iota}}$. Elle est bien sûr impaire.

On dit qu'une représentation irréductible $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ est *modulaire* si elle est isomorphe à $\overline{\rho_{f,\iota}}$, pour f et ι comme ci-dessus. La forme qualitative de la conjecture de Serre est :

CONJECTURE 1.2. — *Soit $\overline{\rho} : G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ une représentation de type S, i.e. irréductible et impaire. Alors $\overline{\rho}$ est modulaire.*

C'est un fait central dans le sujet qu'une représentation $\overline{\rho}$ peut provenir de formes f de niveaux et de poids différents. La forme forte de la conjecture définit $(k(\overline{\rho}), N(\overline{\rho}), \epsilon(\overline{\rho}))$ minimal, en un sens précisé, tel que $\overline{\rho}$ provienne de f de poids $k(\overline{\rho})$, de niveau $N(\overline{\rho})$. Si l'on exclut certaines représentations diédrales pour $p = 2$ ou 3, on peut de plus imposer que le caractère de f soit $\epsilon(\overline{\rho})$.

On peut formuler une version de la forme forte de la conjecture, en terme de formes modulaires modulo p de Katz, qui de plus prédit quand $\overline{\rho}$ provient d'une forme modulaire de Katz de poids 1. Si $p \neq 2$, la version qualitative de la conjecture entraîne la version forte (sous l'une ou l'autre de ses deux formes). On renvoie pour ceci au rapport d'Edixhoven ([25]).

Le niveau $N(\overline{\rho})$ est défini par la formule usuelle pour le conducteur d'une représentation, sauf que l'on ne tient compte que de la ramification en dehors de p. En particulier, on a $N(\overline{\rho}) = 1$ si et seulement si $\overline{\rho}$ est non ramifiée en dehors de p.

Le poids $k(\overline{\rho})$ ne dépend que de l'action de la ramification en p. Notons $I_p \subset G_{\mathbb{Q}}$ le sous-groupe d'inertie pour une valuation p-adique de $\overline{\mathbb{Q}}$. La définition de $k(\overline{\rho})$ repose sur les propriétés que l'on connaît de l'action de I_p dans les représentations p-adiques associées aux formes modulaires ([52, 24]). Rappelons en quelques propriétés, pour $p \neq 2$.

Soit $\overline{\chi_p} : G_{\mathbb{Q}} \to \mathbb{F}_p^*$ la réduction modulo p du caractère cyclotomique, et notons encore $\overline{\chi_p}$ sa restriction à I_p. Soit Ψ un caractère fondamental de niveau 2 de I_p : on peut prendre pour Ψ la restriction à I_p du caractère de Kummer pour l'extension $K(p^{\frac{1}{(p^2-1)}})/K$, K étant l'extension quadratique non ramifiée de $\mathbb{Q}_p$ dans $\overline{\mathbb{Q}_p}$.

On a : $2 \leq k(\overline{\rho}) \leq p^2 - 1$. Il existe j entier tel que l'on ait : $2 \leq k(\overline{\chi_p}^j \otimes \overline{\rho}) \leq p+1$. On a $2 \leq k(\overline{\rho}) \leq p+1$ si et seulement si l'on est dans l'un des deux cas suivants :

– il existe une droite D dans l'espace V de $\overline{\rho}$ telle que I_p opère trivialement sur V/D. L'action de I_p s'écrit alors :

$$\begin{pmatrix} \overline{\chi_p}^b & \eta \\ 0 & 1 \end{pmatrix}.$$

avec $1 \leq b \leq p-1$. Si $b \neq 1$, on a : $k(\overline{\rho}) = 1 + b$. Si $b = 1$, η est un 1-cocycle de $Z^1(I_p, \overline{\mathbb{F}_p}(\overline{\chi_p}))$. Sa classe de cohomologie $c(\eta)$ provient par la théorie de Kummer d'un élément de $\mathbb{Q}_{p,\mathrm{nr}}^* \otimes \overline{\mathbb{F}_p}$, $\mathbb{Q}_{p,\mathrm{nr}}$ désignant l'extension maximale non ramifiée de $\mathbb{Q}_p$. Soit $v : \mathbb{Q}_{p,\mathrm{nr}}^* \otimes \overline{\mathbb{F}_p} \to \overline{\mathbb{F}_p}$ l'homomorphisme défini par la valuation. On a $k(\overline{\rho}) = 2$ si $v(c(\eta)) = 0$ (cas peu ramifié) et $k(\overline{\rho}) = p+1$ sinon (cas très ramifié). Dans le cas peu ramifié, la restriction de $\overline{\rho}$ à D_p provient d'un schéma en groupes fini et plat sur $\mathbb{Z}_p$.