ISSN 0012-9593

quatrième série - tome 47 fascicule 5 septembre-octobre 2014

ANNALES SCIENTIFIQUES de L'ÉCOLE NORMALE SUPÉRIEURE

Christophe BREUIL & Fred DIAMOND

Formes modulaires de Hilbert modulo p et valeurs d'extensions entre caractères galoisiens

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Ann. Scient. Éc. Norm. Sup.
4^e série, t. 47, 2014, p. 905 à 974

FORMES MODULAIRES DE HILBERT MODULO p ET VALEURS D'EXTENSIONS ENTRE CARACTÈRES GALOISIENS

PAR CHRISTOPHE BREUIL ET FRED DIAMOND

RÉSUMÉ. – Soit F un corps totalement réel, v une place de F non ramifiée divisant p et $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ une représentation continue irréductible dont la restriction $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ est réductible et suffisamment générique. Si $\overline{\rho}$ est modulaire (et satisfait quelques conditions techniques faibles), nous montrons comment retrouver l'extension correspondante entre les deux caractères de $\mathrm{Gal}(\overline{F_v}/F_v)$ en terme de l'action de $\mathrm{GL}_2(F_v)$ sur la cohomologie modulo p.

ABSTRACT. – Let F be a totally real field, v an unramified place of F dividing p and $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{GL}_2(\overline{\mathbb{F}_p})$ a continuous irreducible representation such that $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ is reducible and sufficiently generic. If $\overline{\rho}$ is modular (and satisfies some weak technical assumptions), we show how to recover the corresponding extension between the two characters of $\mathrm{Gal}(\overline{F_v}/F_v)$ in terms of the action of $\mathrm{GL}_2(F_v)$ on the cohomology mod p.

1. Introduction

Soit F un corps totalement réel, v une place de F divisant p et F_v le complété de F en v ; le H^1 étale modulo p de tours de courbes de Shimura sur F de niveau en v arbitrairement grand fournit des représentations lisses de $\mathrm{GL}_2(F_v)$ sur $\overline{\mathbb{F}_p}$ que l'on aimerait comprendre. Si f est une forme de Hilbert propre pour les opérateurs de Hecke de représentation galoisienne modulo p associée $\overline{\rho}_f$ irréductible, on aimerait par exemple déjà savoir décrire la partie $\overline{\rho}_f$-isotypique de ces représentations de $\mathrm{GL}_2(F_v)$. Ceci est chose faite lorsque $F_v = \mathbb{Q}_p$ (au moins lorsque $F = \mathbb{Q}$ [19], mais cela devrait s'étendre à tout F) mais demeure largement mystérieux lorsque $F_v \neq \mathbb{Q}_p$.

Une des premières tentatives a été de comprendre les représentations de $\mathrm{GL}_2(\mathcal{O}_{F_v})$ apparaissant (à multiplicité près) dans le $\mathrm{GL}_2(\mathcal{O}_{F_v})$-socle de cette partie $\overline{\rho}_f$-isotypique : dans [8], les auteurs donnent une liste conjecturale explicite de ces « poids de Serre » lorsque F_v est

C. B. remercie pour leur soutien le CNRS, l'université Paris-Sud, le projet ThéHopaD ANR-2011-BS01-005 et le Fields Institute. F. D. remercie l'IHÉS pour son soutien durant la phase de recherche, le C.R.M. de Barcelone, l'I.A.S. et le Fields Institute pour leur hospitalité durant la phase de rédaction.

une extension non ramifiée de $\mathbb{Q}_p$, conjecture qui vient d'être presque complètement démontrée indépendamment par Gee-Kisin [25] et Newton [35] (voir le théorème 3.1.1 ci-dessous). Cette liste ne dépend que de la représentation locale $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ (et même seulement de sa restriction à l'inertie). À la suite de [8], des représentations lisses admissibles de $\mathrm{GL}_2(F_v)$ sur $\overline{\mathbb{F}_p}$ avec les $\mathrm{GL}_2(\mathcal{O}_{F_v})$-socles de [8] ont été construites dans [5] de manière purement locale en supposant $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ suffisamment générique. Des résultats récents d'Emerton, Gee et Savitt ([20]) (faisant suite à des résultats partiels dans le cas de variétés de Shimura compactes à l'infini (cf. [3]) et des calculs informatiques de Dembélé (dans le même cadre, cf. [15])) montrent que la partie $\overline{\rho}_f$-isotypique ci-dessus contient l'une des représentations de [5] (lorsque $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ est générique). Mais l'une des nouveautés de [5] est que, dès que $F_v \neq \mathbb{Q}_p$ (avec F_v non ramifiée), alors il y a *énormément* de représentations lisses admissibles de $\mathrm{GL}_2(F_v)$ sur $\overline{\mathbb{F}_p}$ de socle fixé (celui correspondant à $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$). Plus précisément, dans les « diagrammes » que contiennent les représentations de [5] apparaissent de multiples « paramètres » dont le nombre grossit exponentiellement avec le degré $[F_v : \mathbb{Q}_p]$ et dont les valeurs sur la partie $\overline{\rho}_f$-isotypique ci-dessus (supposée contenir l'un de ces diagrammes) sont pour la plupart à ce jour mystérieuses (par exemple, on ignore si toutes sont locales, i.e., ne dépendent que de $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$).

Du côté représentations de Galois, il n'y a pas de paramètres nouveaux qui apparaissent lorsque l'on passe de $\mathrm{Gal}(\overline{\mathbb{Q}_p}/\mathbb{Q}_p)$ à $\mathrm{Gal}(\overline{F_v}/F_v)$, *sauf si la représentation de* $\mathrm{Gal}(\overline{F_v}/F_v)$ *est une extension non scindée entre deux caractères* : on sait en effet que l'espace de ces extensions est génériquement de dimension $[F_v : \mathbb{Q}_p]$ sur $\overline{\mathbb{F}_p}$. Une question naturelle se pose alors : est-ce que parmi les nombreux paramètres qui apparaissent côté $\mathrm{GL}_2(F_v)$ il s'en trouve au moins quelques-uns dont les valeurs déterminent complètement l'extension entre les deux caractères de la représentation de $\mathrm{Gal}(\overline{F_v}/F_v)$ lorsque celle-ci est (générique) réductible non scindée ? Le but de cet article est de montrer que oui : lorsque $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ est générique réductible, nous montrons d'une part que certains des paramètres de [5] côté $\mathrm{GL}_2(F_v)$ sont bien définis (sans aucune conjecture) sur la partie $\overline{\rho}_f$-isotypique du H^1 étale ci-dessus, et d'autre part que leurs valeurs permettent de retrouver effectivement l'extension précise entre les caractères de $\overline{\rho}_f|_{\mathrm{Gal}(\overline{F_v}/F_v)}$. Notons que ce genre de résultat n'a d'équivalent ni modulo $\ell \neq p$ (puisque génériquement il n'y a pas d'extension non scindée entre deux caractères de $\mathrm{Gal}(\overline{F_v}/F_v)$ modulo $\ell \neq p$) ni modulo p pour $\mathrm{Gal}(\overline{\mathbb{Q}_p}/\mathbb{Q}_p)$ (puisque génériquement il y a alors une seule extension non scindée).

Énonçons maintenant plus précisément les résultats principaux de l'article. On suppose donc F_v non ramifiée et on note k_v son corps résiduel. On fixe une représentation continue, irréductible, totalement impaire $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ et on suppose que $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ est réductible générique, c'est-à-dire de la forme (quitte à tordre par un caractère) :

$$\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)} \cong \begin{pmatrix} \mathrm{nr}_v \prod_{\sigma : k_v \hookrightarrow \overline{\mathbb{F}_p}} \omega_\sigma^{r_{v,\sigma}+1} & * \\ 0 & \mathrm{nr}'_v \end{pmatrix}$$

où $r_{v,\sigma} \in \{0, \ldots, p-3\}$ (non tous égaux à 0 ou à $p-3$), ω_σ est le caractère fondamental de Serre associé au plongement σ et $\mathrm{nr}_v, \mathrm{nr}'_v$ des caractères non ramifiés. On peut alors décrire explicitement $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ par le truchement de son module de Fontaine-Laffaille contravariant :

$$\prod_{\sigma:k_v\hookrightarrow\overline{\mathbb{F}_p}}\left(M^\sigma=\overline{\mathbb{F}_p}e^\sigma\oplus\overline{\mathbb{F}_p}f^\sigma,\mathrm{Fil}^{r_{v,\sigma}+1}M^\sigma=\overline{\mathbb{F}_p}f^\sigma\right)$$

avec $\begin{cases}\varphi(e^\sigma) &=\alpha_{v,\sigma}e^{\sigma\circ\varphi^{-1}}\\ \varphi_{r_{v,\sigma}+1}(f^\sigma)&=\beta_{v,\sigma}(f^{\sigma\circ\varphi^{-1}}+x_{v,\sigma}e^{\sigma\circ\varphi^{-1}})\end{cases}$ où $\alpha_{v,\sigma},\beta_{v,\sigma}\in\overline{\mathbb{F}_p}^\times$ et $x_{v,\sigma}\in\overline{\mathbb{F}_p}$.

Maintenant, on suppose $\overline{\rho}$ modulaire, c'est-à-dire :

$$\pi_D(\overline{\rho})\overset{\text{déf}}{=}\mathrm{Hom}_{\overline{\mathbb{F}_p}[\mathrm{Gal}(\overline{\mathbb{Q}}/F)]}\left(\overline{\rho},\varinjlim_U H^1_{\text{ét}}(X_{U,\overline{\mathbb{Q}}},\overline{\mathbb{F}_p})(1)\right)\neq 0$$

où $(X_U)_U$ est une tour de courbes de Shimura sur F associée à une algèbre de quaternions D sur F déployée en une seule des places infinies de F ainsi qu'aux places divisant p (U parcourant les sous-groupes ouverts compacts de $(D\otimes_{\mathbb{Z}}\widehat{\mathbb{Z}})^\times$). Sous quelques hypothèses techniques sur $\overline{\rho}$ (que nous n'avons pas cherché à optimiser, cf. début du § 3.3), on peut utiliser l'action de $(D\otimes_{\mathbb{Z}}\widehat{\mathbb{Z}})^\times$ sur $\pi_D(\overline{\rho})$ aux places différentes de v pour définir un « facteur local » $\pi_{D,v}(\overline{\rho})$ en v qui est une représentation lisse admissible de $(D\otimes_F F_v)^\times\cong\mathrm{GL}_2(F_v)$ sur $\overline{\mathbb{F}_p}$ (mais dont on ignore si elle ne dépend que de $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$). Notons que l'on ne dispose pas ici *a priori* d'une factorisation de $\pi_D(\overline{\rho})$ « à la Flath » (bien que cela soit conjecturé, cf. [8, Conj. 4.7] et le § 3.1), d'où la nécessité de définir soigneusement ce facteur local $\pi_{D,v}(\overline{\rho})$.

Si J est un sous-ensemble des plongements de k_v dans $\overline{\mathbb{F}_p}$, on définit la « frontière de J » $F(J)$ comme l'ensemble des plongements $\sigma:k_v\hookrightarrow\overline{\mathbb{F}_p}$ tels que ou bien $\sigma\in J$ et $\sigma\circ\varphi^{-1}\notin J$, ou bien $\sigma\notin J$ et $\sigma\circ\varphi^{-1}\in J$ où φ est le Frobenius usuel $x\mapsto x^p$ sur k_v. Notons $\mathrm{I}(\mathcal{O}_{F_v})$ le sous-groupe de $\mathrm{GL}_2(\mathcal{O}_{F_v})$ des matrices triangulaires supérieures modulo p et $\mathrm{I}_1(\mathcal{O}_{F_v})\subset\mathrm{I}(\mathcal{O}_{F_v})$ celui des matrices unipotentes supérieures modulo p. Le premier théorème associe à $\pi_{D,v}(\overline{\rho})$ certains invariants $x(J)$ de $\overline{\mathbb{F}_p}^\times$ qui apparaissent naturellement dans les diagrammes de [5, § 13] (bien qu'ils n'y soient pas explicités).

THÉORÈME 1.1. – *Soit J tel que $F(J)\cap\{\sigma,x_{v,\sigma}=0\}=\varnothing$. Il existe à scalaire près un unique vecteur $v\in\pi_{D,v}(\overline{\rho})^{\mathrm{I}_1(\mathcal{O}_{F_v})}$ non nul sur lequel $\mathrm{I}(\mathcal{O}_{F_v})$ agisse par le caractère $\prod_{\sigma\in J}\sigma^{p-1}\prod_{\sigma\notin J}\sigma^{r_{v,\sigma}}\otimes\prod_{\sigma\in J}\sigma^{r_{v,\sigma}}\prod_{\sigma\notin J}\sigma^{p-1}$ de $\mathrm{I}(\mathcal{O}_{F_v})/\mathrm{I}_1(\mathcal{O}_{F_v})\cong k_v^\times\times k_v^\times$ et un unique élément $x(J)\in\overline{\mathbb{F}_p}^\times$ tel que l'on ait l'égalité dans $\pi_{D,v}(\overline{\rho})$:*

$$\sum_{s\in k_v}\left(\prod_{\sigma\in J}\sigma(s)^{p-1-r_{v,\sigma}}\right)\begin{pmatrix}[s]&1\\1&0\end{pmatrix}\begin{pmatrix}0&1\\p&0\end{pmatrix}v=x(J)\sum_{s\in k_v}\left(\prod_{\sigma\notin J}\sigma(s)^{p-1-r_{v,\sigma}}\right)\begin{pmatrix}[s]&1\\1&0\end{pmatrix}v$$

où $[s]$ est le représentant multiplicatif de s dans $\mathcal{O}_{F_v}$.

Lorsque $\{\sigma,x_{v,\sigma}=0\}=\varnothing$, les invariants $x(J)$ ci-dessus sont les *seuls* invariants de [5], mais il en apparaît bien d'autres lorsque $\{\sigma,x_{v,\sigma}=0\}\neq\varnothing$. Les résultats de [5] montrent par ailleurs que l'on peut construire des représentations lisses admissibles de $\mathrm{GL}_2(F_v)$ sur $\overline{\mathbb{F}_p}$ avec le $\mathrm{GL}_2(\mathcal{O}_{F_v})$-socle correspondant à $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ et des valeurs *presque quelconques* de ces invariants $x(J)$ (voir § 2.6), de sorte que les valeurs prises par les scalaires $x(J)$ du théorème 1.1 ne peuvent pas du tout être prédites *a priori*. Le deuxième théorème donne ces valeurs précises.

THÉORÈME 1.2. – *Soit J tel que $F(J)\cap\{\sigma, x_{v,\sigma}=0\}=\varnothing$, on a :*

$$x(J)=-\Big(\prod_{\sigma\in J}\alpha_{v,\sigma}\prod_{\sigma\notin J}\beta_{v,\sigma}\Big)^{-1}\frac{\prod\limits_{\substack{\sigma\in J\\ \sigma\circ\varphi^{-1}\notin J}}x_{v,\sigma}(r_{v,\sigma}+1)}{\prod\limits_{\substack{\sigma\notin J\\ \sigma\circ\varphi^{-1}\in J}}x_{v,\sigma}(r_{v,\sigma}+1)}\in\overline{\mathbb{F}_p}^{\times}.$$

En particulier, on voit que ces valeurs sont *locales*, i.e., ne dépendent que de $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$, ce qui n'était pas évident *a priori*. Notons que les scalaires $\alpha_{v,\sigma}$, $\beta_{v,\sigma}$ et $x_{v,\sigma}$ ne sont pas définis de manière unique (comme le lecteur peut immédiatement le voir en faisant un changement de base sur M^σ qui respecte les structures), mais on peut vérifier directement que les scalaires $x(J)$ du théorème 1.2 ne dépendent pas des choix faits. En particulier, on peut supposer tous les $\alpha_{v,\sigma}$ (resp. $\beta_{v,\sigma}$) égaux à 1 sauf un, donné alors par $\mathrm{nr}'_v(p^{-1})$ (resp. $\mathrm{nr}_v(p^{-1})$) et, quitte à faire un changement de base, on peut également supposer que l'un des $x_{v,\sigma}$ vaut 1 (du moins s'il existe un $x_{v,\sigma}$ non nul, mais dans le cas contraire $\overline{\rho}|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ est scindée et les invariants $x(J)$ donnés par les théorèmes ci-dessus se limitent à $\mathrm{nr}_v(p)$ et $\mathrm{nr}'_v(p)$). Le lecteur pourra alors vérifier, en prenant par exemple des J de la forme $\{\sigma,\sigma\circ\varphi,\ldots,\sigma\circ\varphi^j\}$ pour σ et j convenables, que l'on retrouve facilement les valeurs de tous les autres $x_{v,\sigma}$ non nuls à partir des valeurs des $x(J)$ du théorème 1.2 et des $r_{v,\sigma}$ (cf. la remarque 2.1.2 (iii)).

Le travail récent d'Emerton, Gee et Savitt ([20], voir aussi [3]) montre que, sous nos hypothèses, la $\mathrm{GL}_2(F_v)$-représentation $\pi_{D,v}(\overline{\rho})$ contient un des diagrammes construits dans [5]; notons-le $D_v(\overline{\rho})$. Une conjecture naturelle supportée par le théorème 1.2 ci-dessus est que $D_v(\overline{\rho})$ est entièrement *local*, c'est-à-dire ne dépend que de la restriction de $\overline{\rho}$ à $\mathrm{Gal}(\overline{F_v}/F_v)$ (c'est donc le cas si $\{\sigma, x_{v,\sigma}=0\}=\varnothing$). Si l'on est optimiste on peut même penser que toute la $\mathrm{GL}_2(F_v)$-représentation $\pi_{D,v}(\overline{\rho})$ pourrait elle-même être locale. Bien que nous espérions que ces énoncés soient vrais, nous avons néanmoins choisi de ne pas les présenter sous forme de conjectures. Une raison est que, en dehors des résultats de cet article et de l'article de Hu [29], nous ne savons pour l'instant rien de plus sur les diagrammes $D_v(\overline{\rho})$, qui demeurent donc en général largement mystérieux (sans parler des $\mathrm{GL}_2(F_v)$-représentations $\pi_{D,v}(\overline{\rho})$).

Disons quelques mots sur les preuves des théorèmes 1.1 et 1.2. Le cœur du théorème 1.1 est de montrer que le poids de Serre $\bigotimes_{\sigma:k_v\hookrightarrow\overline{\mathbb{F}_p}}(\mathrm{Sym}^{r_{v,\sigma}}\,\overline{\mathbb{F}_p}^2)^\sigma$ (voir § 2.1 pour les notations) apparaît *avec multiplicité un* dans le $\mathrm{GL}_2(\mathcal{O}_{F_v})$-socle de $\pi_{D,v}(\overline{\rho})$ (le fait qu'il apparaisse était essentiellement déjà connu). Cela se démontre en utilisant les techniques de multiplicité un issues de la méthode de Taylor-Wiles comme inauguré par Fujiwara ([23]) et l'un d'entre nous ([16]). Un deuxième ingrédient essentiel est que la représentation de $\mathrm{GL}_2(\mathcal{O}_{F_v})$ sur $\overline{\mathbb{F}_p}$:

$$\mathrm{ind}_{\mathrm{I}(\mathcal{O}_{F_v})}^{\mathrm{GL}_2(\mathcal{O}_{F_v})}\Big(\prod_{\sigma\in J}\sigma^{p-1}\prod_{\sigma\notin J}\sigma^{r_{v,\sigma}}\otimes\prod_{\sigma\in J}\sigma^{r_{v,\sigma}}\prod_{\sigma\notin J}\sigma^{p-1}\Big)$$

n'a qu'un seul de ses constituants qui apparaît dans ce $\mathrm{GL}_2(\mathcal{O}_{F_v})$-socle : à savoir le poids de Serre $\bigotimes_\sigma(\mathrm{Sym}^{r_{v,\sigma}}\,\overline{\mathbb{F}_p}^2)^\sigma$ ci-dessus. Cela se déduit par exemple directement de [25] et d'un calcul facile (mais peut aussi se démontrer de manière plus élémentaire sans utiliser [25]). Une fois ces deux ingrédients disponibles, l'existence de $x(J)$ se ramène essentiellement à de la théorie des représentations (cf. proposition 2.6.1).