* Deformation of a modular form and modular Galois representations

Haruzo Hida
UCLA, Los Angeles, CA 90095-1555, U.S.A.

UCLA student seminar, September 27, 2022.

*I would like to explain the motivation of my starting p-adic deformation theory of modular forms and modular Galois representations and the main points of the theory. All modular forms in this talk are elliptic either classical or p-adic. We write $X_?(N)$ $(? = 0, 1)$ for compactified modular curves of the level N groups $\Gamma_?(N) \subset \mathrm{SL}_2(\mathbb{Z})$, and $Y_?(N) = X_?(N) - \{cusps\}$. Here are references:

[H13] Inventiones Math. **194** (2013), 1–40 (Image paper)
[H86a] Annals Sci. Ec Norm. Sup. **19** (1986) (Iwasawa module paper)
[H86b] Inventiones **85** (1986) (Galois representation paper)
[GME] Geometric Modular Forms and Elliptic Curves, World Scientific, 2012.
[EMI] Elementary Modular Iwasawa Theory, World Scientific, 2022.

§**0. A fundamental question in 1980.** Consider a homogeneous space $X := \mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A})/\mathbb{A}^\times$ with a Haar measure $d\mu$ and an L^2 Hilbert space

$$L^2_{cusp} := \{f : X \to \mathbb{C} | \int_X |f|^2 d\mu < \infty, \int_{\mathbb{A}} f(\left(\begin{smallmatrix} 1 & u \\ 0 & 1 \end{smallmatrix}\right) h) du = 0 \ \forall h\}.$$

By Harish-Chandra, the representation $\Pi : \mathrm{GL}_2(\mathbb{A}) \to \mathrm{Aut}_{\mathbb{C}}(L^2_{cusp})$ given by $\Pi(f)(h) = f(hg)$ is a discrete direct sum of irreducible representations. Towards the end of 1980, I felt that having a discrete spectrum is unfortunate. One cannot easily come close to a harder irreducible constituent $\pi \subset \Pi$ from another somehow simpler π_0 just by discreteness.

Question. *Is there any other topology, keeping the information of π and π_0 to good extent, to have a continuous spectrum?*

§**1. Why not try p-adic topology?** We need a p-adic replacement of L^2_{cusp}: $S(\mathbb{C}) = \mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A})/\mathbb{R}_+^\times \mathrm{SO}_2(\mathbb{R})$ is the complex points of the Shimura curve (a pro-curve made of modular curves: $\Gamma\backslash\mathfrak{H}$) defined over $\mathbb{Z}$. Consider a holomorphic cusp form $f : \mathfrak{H} \to \mathbb{C}$ of weight k on Γ. Writing $j(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right), \tau) = c\tau + d$, f satisfies $f(\gamma(\tau)) = f(\tau)j(\gamma,\tau)^k$ for $\gamma \in \Gamma$. The automorphic factor $g \mapsto j(g,\tau)^k$ is a 1-cocycle having values in nowhere vanishing holomorphic functions on $\mathfrak{H}$ as $j(gh,\tau) = j(g,h(\tau))j(g,\tau)$. Thus $\Gamma\backslash(\mathfrak{H}\times\mathbb{C})$ by the action $(\tau,u) \mapsto (\gamma(\tau), j(\gamma,\tau)^k u)$ gives a cuspidal line bundles $\omega^k_{cusp} \subset \omega^k$ on $\Gamma\backslash\mathfrak{H}$, which descends to $\mathbb{Z}$ canonically, as $\omega = \varphi_*\Omega^1_{\mathbb{E}/S}$ for the universal elliptic curve $\mathbb{E} \xrightarrow{\varphi} S_{/\mathbb{Z}}$. Similarly, writing $L(n;A) = AX^n + AX^{n-1}Y + \cdots + AY^n$ and letting $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathfrak{H} \times L(n;A)$ by $(\tau, P(X,Y)) \mapsto (\gamma(\tau), P((X,Y)^t\gamma^{-1}))$, we have the sheaf $\mathcal{L}(n;A)$ of locally constant sections from $\Gamma\backslash\mathfrak{H}$ to $\Gamma\backslash(\mathfrak{H}\times L(n;A))$. We replace L^2 by the p-adic completion of

$$\begin{cases} V_k := \varinjlim_\alpha \widehat{H^0(X_1(Np^\alpha)_{/\mathbb{Z}_p}, \omega^k_{cusp})} & \text{coherent cohomology,} \\ TJ_n := \varprojlim_\alpha H^1_{et,!}(Y_1(Np^\alpha), \mathcal{L}(n;\mathbb{Z}_p)) & \text{pro-étale cohomology.} \end{cases}$$

Here $H^1_! = \mathrm{Im}(H^1_c \to H^1)$.

§**1. Why taking the limit for $Y_1(Np^n)$.** Taking limit with respect to p^n is an analogue from Kummer and Iwasawa, as $\mathbb{Z}_p^\times = \varprojlim_\alpha \Gamma_0(p^n)/\Gamma_1(p^n)$, the modules V and TJ_n are Iwasawa modules over $\mathbb{Z}_p[[\Gamma]] \cong \mathbb{Z}_p[[T]]$ for $\Gamma = \mathbb{Z}_p^\times/\{\text{torsion}\} = \gamma^{\mathbb{Z}_p}$. I wanted to make a GL(2)-version of Iwasawa theory. Let

$$\widehat{\Gamma}_0(N) := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\widehat{\mathbb{Z}}) \middle| c \in N\widehat{\mathbb{Z}} \right\},$$
$$\widehat{\Gamma}_1(N) := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \widehat{\Gamma}_0(N) \middle| a \equiv 1 \mod N\widehat{\mathbb{Z}} \right\}.$$

Then $Y_?(N) = S/\widehat{\Gamma}_?(N) \cong \Gamma_?(N)\backslash\mathfrak{H}$. For each $\pi \subset \Pi$, there is a minimal N with $H^0(\widehat{\Gamma}_1(N), \pi) = \mathbb{C}f_\pi$. If f is holomorphic, we call π is holomorphic, and $f = \sum_{n=1}^\infty a(n,f)\mathbf{e}(n\tau)$. Decomposing

$$\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(\mathbb{Z}) | ad - bc = n, a \equiv 1 \mod N, N|c\} = \bigsqcup_i \alpha_i \Gamma_1(N),$$

define the Hecke operator $T(n)$ by $f|T(n)(g) = \sum_i f(g\alpha_i^{(\infty)})$, we get $f|T(n) = a(n,f)f$ if $a(1,f) = 1$ as $H^0(\widehat{\Gamma}_1(N), \pi) = \mathbb{C}f_\pi$, and $T(n)$ determines f and hence π. Therefore, it is sufficient to take limit with respect to $\Gamma_1(Np^\alpha)$.

§3. **Weight reduction to the constant sheaf.** Consider a morphism $I : L(n; \mathbb{Z}/p^\alpha\mathbb{Z}) \to \mathbb{Z}/p^\alpha\mathbb{Z}$ sending $P(X,Y) \mapsto P(1,0)$. Note $\Gamma_1(p^\alpha) \ni \gamma \equiv \left(\begin{smallmatrix} 1 & u \\ 0 & 1 \end{smallmatrix}\right)$ mod p^α. Thus

$$I(\gamma P(X,Y)) = P((1,0)\left(\begin{smallmatrix} 1 & 0 \\ -u & 1 \end{smallmatrix}\right)) = I(P(X,Y)),$$

and hence $I : \mathcal{L}(n; \mathbb{Z}/p^\alpha\mathbb{Z}) \to \mathbb{Z}/p^n\mathbb{Z}$ is an étale sheaf morphism over $Y_0(p^\alpha)$. Passing to the limit, we have a Galois and Hecke equivariant map: $TJ_n \to TJ_0$. We have well defined Hecke operators $T(n)$ and $T(l) = U(l)$ for $l|Np$ on $H^1_{et.!}$, and I is also Hecke equivariant. Define for $k = n+2$ and $\Lambda := \mathbb{Z}_p[[\Gamma]] = \mathbb{Z}_p[[T]]$ (by $\langle\gamma\rangle \mapsto 1+T$ for the fixed generator γ of Γ)

$$\mathbf{h}_k(Np^\infty; \mathbb{Z}_p) = \Lambda[[T(n) : 0 < n \in \mathbb{Z}]] \subset \mathrm{End}_\Lambda(TJ_n).$$

Independence Theorem: [EMI, Theorem 4.2.17]. We have a canonical isomorphism $i : \mathbf{h}_k(Np^\infty; \mathbb{Z}_p) \cong \mathbf{h}_2(Np^\infty; \mathbb{Z}_p)$ such that $i(T(n)) = T(n)$, $i(U(l)) = U(l)$ and $i(\langle z\rangle) = z^n\langle z\rangle$ for $z \in \mathbb{Z}_p^\times$, where $\langle z\rangle \in \varprojlim \Gamma_0(p^\alpha)/\Gamma_1(p^\alpha) \cong \mathbb{Z}_p^\times$ and $z^n \in \mathbb{Z}_p^\times \hookrightarrow \mathbb{Q}_p$.

Hereafter we just write $\mathbf{h} := \mathbf{h}_2(Np^\infty; \mathbb{Z}_p)$.

§**4. Finite level Hecke algebras.** Define

$$\mathbf{h}_k(Np^\alpha;\mathbb{Z}_p) = \mathbb{Z}_p[T(n)|0 < n \in \mathbb{Z}] \subset \mathrm{End}_{\mathbb{Q}_p}(H^1_!(Y_1(Np^\alpha),\mathcal{L}(n;\mathbb{Q}_p))).$$

Then by definition

$$\mathbf{h} = \varprojlim_\alpha \mathbf{h}_k(Np^\alpha;\mathbb{Z}_p) \quad \text{for all } 2 \leq k \in \mathbb{Z}.$$

By the Eichler–Shimura Hecke equivariant isomorphism:

$$H^1_!(X_1(Np^\alpha),\mathcal{L}(n;\mathbb{C})) \cong H^0(X_1(Np^\alpha)_{/\mathbb{C}},\omega^k_{cusp}) \oplus \overline{H^0(X_1(Np^\alpha)_{/\mathbb{C}},\omega^k_{cusp})},$$

alternatively,

$$\mathbf{h}_k(Np^\alpha;\mathbb{Z}_p) := \mathbb{Z}_p[T(n)|0 < n \in \mathbb{Z}] \subset \mathrm{End}_{\mathbb{Z}_p}(H^0(X_1(Np^\alpha)_{/\mathbb{Z}_p},\omega^k_{cusp})).$$

Thus $\mathbf{h}$ acts on V_k.

§**5.** q-**Expansion.** The J-invariant has its q-expansion in $q^{-1}\mathbb{Z}[[q]]$ starting with q^{-1}; so, $J^{-1} = q + \cdots \in \mathbb{Z}[[q]]^\times$ gives rise to a parameter at ∞ of $X_0(1) = \mathbf{P}^1(J)_{/\mathbb{Z}}$. Since $X_1(N)$ is étale over $\infty \in \mathbf{P}^1(J)$, $\widehat{\mathcal{O}}_{X_1(N),\infty/\mathbb{Z}} = \mathbb{Z}[[q]]$. Thus $f \in H^0(X_1(Np^\alpha)_{/\mathbb{Z}_p}, \omega^k_{cusp})$ has q-expansion $f(q) = \sum_{n=1}^\infty a(n,f)q^n \in \mathbb{Z}_p[[q]]$. By the q-expansion principle, we have $V_k \hookrightarrow \mathbb{Z}_p[[q]]$ as $\mathbb{Z}_p[[q]]$ is p-adically complete. Serre considered the space of p-adic modular forms as a p-adic completion of $\sum_k H^0(X_1(N)_{/\mathbb{Z}_p}, \omega^k_{cusp}) \subset \mathbb{Z}_p[[q]]$ in 1973. Soon after, Katz generalized this to the p-adic completion $V(Np^\alpha)$ of $\sum_k H^0(X_1(N)_{/\mathbb{Q}_p}, \omega^k_{cusp}) \cap \mathbb{Z}_p[[q]]$ inside $\mathbb{Z}_p[[q]]$, and via his notion of geometric modular forms, he remarked

$$V(Np^\alpha) = V(N) \quad \text{inside } \mathbb{Z}_p[[q]].$$

Duality Theorem: [EMI, §3.2.5]. We have a perfect duality $\langle\cdot,\cdot\rangle : \mathbf{h}_k(Np^\alpha;\mathbb{Z}_p) \times H^0(X_1(Np^\alpha)_{/\mathbb{Z}_p}, \omega^k_{cusp}) \to \mathbb{Z}_p$ given by $\langle h, f\rangle = a(1, f|h)$ for $\alpha = 1, \ldots, \infty$, where we put $H^0(X_1(Np^\infty)_{/\mathbb{Z}_p}, \omega^k_{cusp}) := V_k$. So, $V_k = V_2 \overset{\text{Katz}}{=} V(N) \subset \mathbb{Z}_p[[q]]$ by Independence theorem.

§6. **Ordinary part.** The algebra $\mathbf{h}$ is a little too big to have an exact structure theorem. We consider the ordinary projection $e := \lim_{n\to\infty} U(p)^{n!}$ inside $\mathbf{h}$. Write $\mathbf{h}_k^{ord} = e\mathbf{h}_k$ and $\mathbf{h}^{ord} = e\mathbf{h}$.

Why e cuts down $V(N)$ and $\mathbf{h}$ to a reasonable size?

For whatever $\alpha > 0$, we have

$$\Gamma_0(p^\alpha)\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)\Gamma_0(p^\alpha) = \Gamma_0(p^\alpha)\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)\Gamma_0(p^{\alpha-1}) = \bigsqcup_{u=0}^{p-1}\Gamma_0(p^\alpha)\left(\begin{smallmatrix}1&u\\0&p\end{smallmatrix}\right)$$

independent of α; so, a power of $U(p)$ reduces the level of $f \in S_k(\Gamma_1(p^\alpha))$ down to $\Gamma_1(p)$ as long as the Neben character is trivial (contraction property of $U(p)$).

Control theorem:

$$\mathbf{h}^{ord}/(\langle\gamma\rangle - \gamma^n)\mathbf{h}^{ord} \cong \mathbf{h}_k^{ord}(Np; \mathbb{Z}_p) \quad (k = n + 2 \geq 2).$$

This is, for example, proven as [EMI, Theorem 4.1.29], and hence $\mathbf{h}^{ord}$ is Λ-free of rank equal to $\mathbf{h}_2(Np; \mathbb{Z}_p)$ by ring theory.

§7. Galois representations.

Similarly TJ_0 is also Λ-free of rank equal to $\mathrm{rank}_{\mathbb{Z}_p} eH^1_!(X_1(Np), \mathbb{Z}_p)$, which is a Galois module (Tate module of the Jacobian of $X_1(Np)$). By Eichler–Shimura, Frob_l satisfies $X^2 - T(l)X + l\langle l\rangle = 0$ on TJ_0 [GME, Theorem 4.2.2]. Indeed, $TJ_0^{ord} := eTJ_0$ fits into the following connected-étale exact sequence of $\mathbf{h}$-modules [H13, Lemma 4.2]: $0 \to \mathbf{h}^{ord} \to TJ_0^{ord} \xrightarrow{\mathrm{red}} \mathrm{Hom}_\Lambda(\mathbf{h}^{ord}, \Lambda) \to 0$. This follows by a property of reduction modulo p of $X_1(Np^\alpha)$. Take a local ring $\mathbb{T}$ of $\mathbf{h}^{ord}$. If $\mathbb{T} \cong \mathrm{Hom}_\Lambda(\mathbb{T}, \Lambda)$ (i.e., $\mathbb{T}$ is Gorenstein), we have $TJ_0^{ord} \otimes_{\mathbf{h}^{ord}} \mathbb{T} \cong \mathbb{T}^2$ and we get a Galois representation unramified outside Np: $\rho_{\mathbb{T}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{T})$, and if $\mathbb{T}$ is not Gorenstein, localize $\mathbb{T}$ into $\mathbb{T}_P$ at a prime P of $\mathbb{T}$ to make $\mathbb{T}_P$ Gorenstein, we have $\rho_{\mathbb{T}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{T}_P)$ such that

$$\det(1 - \rho_{\mathbb{T}}(\mathrm{Frob}_l)X) = 1 - T(l)|_{\mathbb{T}}X - l\langle l\rangle X^2 \quad \text{for all primes } l \nmid Np,$$

where $\langle l\rangle$ is the image of l in $\varprojlim_\alpha \Gamma_0(Np^\alpha)/\Gamma_1(Np^\alpha) \cong \mathbb{Z}_p^\times \times (\mathbb{Z}/N\mathbb{Z})^\times$.

§8. Λ-adic forms.

Pick $\lambda \in \mathrm{Hom}_\Lambda(\mathbf{h}^{ord}, \Lambda)$. Consider a formal expansion $F_\lambda := \sum_{n=1}^{\infty} \lambda(T(n))q^n \in \Lambda[[q]]$. By the duality and control theorem,

$$\begin{aligned} F_\lambda \bmod (\langle\gamma\rangle - \gamma^n) = F_\lambda(\gamma^n - 1) \\ &= \sum_{n=1}^{\infty} \lambda(T(n))(\gamma^n - 1)q^n \\ &\in H^0_{cusp}(X_1(Np)_{/\mathbb{Z}_p}, \omega^k_{cusp}) =: S_k(\Gamma_0(Np); \mathbb{Z}_p). \end{aligned}$$

Thus F_λ is a Λ-adic form, and

$$\mathrm{Hom}_{\Lambda\text{-alg}}(\mathbf{h}^{ord}, \Lambda) \otimes_\Lambda \Lambda/(\langle\gamma\rangle - \gamma^n) \cong S_k(\Gamma_0(Np); \mathbb{Z}_p).$$

If the local ring $\mathbb{T}$ in §7 is equal to Λ and λ is a Λ-algebra homomorphism, then

$$\rho_{\mathbb{T}} \bmod (\langle\gamma\rangle - \gamma^n) : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Z}_p[\psi])$$

is the Galois representation of the Hecke eigenform $F_\lambda(\gamma^n - 1)$.