SÉMINAIRE DARBOUX, MAY 4 2011

Seminaire Darboux, May 4, 2011

I want to thank the organizers for inviting me to speak in English; not only does has this saved me a great deal of time in preparing the talk, but it also allows me to say twice as much as I otherwise would in the same amount of time.

After taking some time to look at the various ways in which automorphic forms have recommended themselves to physicists, I decided that it would be most useful to me to explain my own perspective as a number theorist. In the wake of the Langlands program, I am more accustomed to thinking about automorphic forms as symptoms of automorphic representations, which on the one hand are spaces of automorphic forms on which certain Lie groups, or adelic groups, act irreducibly, but more crucially from my perspective, are in duality (Langlands correspondence) with number-theoretic structures, specifically with representations of Galois groups satisfying certain properties.

I have been asked to start with automorphic forms and move gradually to the Galois representations. For number theorists, the Galois representations provide the most reliable parametrization of automorphic representations, whereas the automorphic representations provide the only means for obtaining information in general about the diophantine problems.

We start with a semi-simple Lie group G and a maximal compact subgroup $K \subset G$ (or open subgroup thereof) and let $X = G/K$ be the corresponding Riemannian symmetric space. Let $\Gamma \subset G$ be a discrete subgroup of cofinite volume with respect to the invariant Haar measure on G, and let $\mathcal{I}$ be an ideal of finite codimension in the algebra $Z(\mathfrak{g})$ of biinvariant differential operators on G (the center of the enveloping algebra $U(\mathfrak{g})$ of $\mathfrak{g} = Lie(G)$). Then an automorphic form of type Γ for $\mathcal{I}$ is a function $f : \Gamma\backslash X \to \mathbb{C}$ that (a) is annihilated by $\mathcal{I}$ and (b) is of moderate growth along a natural compactification of $\Gamma\backslash X$. But very quickly, we replace f by its pullback $F : \Gamma\backslash G \to \mathbb{C}$, forget f, and let $A_{\Gamma,\mathcal{I}}$ be the space of F obtained in this way; these are the automorphic forms, and the group G acts on $A_{\Gamma,\mathcal{I}}$ (because of the biinvariance of $\mathcal{I}$) by

$$r(h)F(g) = F(gh).$$

In general, this ruins the condition that F comes from a function f on $\Gamma\backslash X$, so we let $A_{\Gamma,\mathcal{I}}$ be the smallest space of $F : \Gamma\backslash G \to \mathbb{C}$ generated by the pullbacks of f as above.

Inside $A_{\Gamma,\mathcal{I}}$ we have $A^2_{\Gamma,\mathcal{I}} \supset A^0_{\Gamma,\mathcal{I}}$, the square-integrable forms and cusp forms respectively, which we complete with respect to the L_2 norm to obtain Hilbert space representations of G. Then Langlands proved that

Theorem. *The space $A^0_{\Gamma,\mathcal{I}}$ is a countable Hilbert sum of irreducible representations of G, each with finite multiplicity. The space $A^2_{\Gamma,\mathcal{I}}$ equals the direct (Hilbert) sum*

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

of $A^0_{\Gamma,\mathcal{I}}$ *(plus a countable residual spectrum) and a finite collection of continuous (Hilbert) integrals of irreducible representations parametrized by finite-dimensional linear spaces, all obtained by Eisenstein series from cusp forms on Levi factors of parabolic subgroups of* G*.*

(The first part of this theorem is due to Gelfand and Piatetski-Shapiro in general, and is a consequence of reduction theory for locally symmetric spaces.)

Strictly speaking, we are assuming that G is the group of real points of a semisimple algebraic group over $\mathbb{Q}$ and that Γ is a congruence subgroup contained in the $\mathbb{Q}$-rational points. In the example of Eisenstein series for $SL(2,\mathbb{Z})$ described by Boris, $G = SL(2,\mathbb{R})$ and $\Gamma = SL(2,\mathbb{Z})$. Usually one takes $\mathcal{I}$ of codimension 1 in $Z(\mathfrak{g})$; in other words, one asks $Z(\mathfrak{g})$ to act by a fixed scalar character. This is always the case (Schur's lemma) when F belongs to an irreducible constituent of $A^2_{\Gamma,\mathcal{I}}$ for some $\mathcal{I}$ (then F actually belongs to $A^2_{\Gamma,\mathcal{I}'}$ for a codimension 1 ideal $\mathcal{I}' \supset \mathcal{I}$).

The irreducible constituents of $A^0_{\Gamma,\mathcal{I}}$ can be thought of as the atoms of the theory, the Eisenstein series as sums of these atoms; the Eisenstein series for $SL(2,\mathbb{Z})$ is obtained as the sum of two cusp forms for $GL(1)$. But to continue in a sensible way, we need to replace Lie groups by adelic groups.

In the first place, it is better to work with reductive groups like $GL(n)$ rather than semisimple groups like $SL(n)$, and indeed the theory of automorphic forms on $GL(n)$ is in some ways much simpler than the theory for general groups.

In the second place, once we have forgotten about X, we can just let $A^0_{\Gamma,\mathcal{I}}$ be the space of all functions on G with the growth conditions, annihilated by $\mathcal{I}$, and which generate finite-dimensional representations of the maximal compact subgroup K. Then G still acts on the Hilbert space completion, and the finiteness theorems of Gelfand-Piatetski-Shapiro and Langlands remain true. An *automorphic representation of* G is then an irreducible G-summand of $A^0_{\Gamma,\mathcal{I}}$ for some $\Gamma,\mathcal{I}$. But these are not the true automorphic representations, because:

Finally, the spaces $A_{\Gamma,\mathcal{I}}$, $A^2_{\Gamma,\mathcal{I}}$, and $A^0_{\Gamma,\mathcal{I}}$ have additional symmetries that are not visible in the previous picture. The point is that for any pair of groups $\Gamma' \subset \Gamma$, there are inclusions $A_{\Gamma,\mathcal{I}} \subset A_{\Gamma',\mathcal{I}}$ but there are also twisted inclusions: if $\alpha \in G$ has the property that $\alpha\Gamma'\alpha^{-1} \subset \Gamma$, then we can map

$$A_{\Gamma,\mathcal{I}} \subset A_{\alpha\Gamma'\alpha^{-1},\mathcal{I}} \xrightarrow{\sim} A_{\Gamma',\mathcal{I}} \to A_{\Gamma,\mathcal{I}}$$

where the second map is just conjugation by α^{-1} and the last map is the average (trace map) over cosets of Γ' in Γ. This composite is an example of a *Hecke operator.* The properties of algebraic groups guarantee that the collection of Hecke operators contains a distinguished subset of so-called *spherical Hecke operators* which generate a very large commutative algebra that nevertheless has a distinguished set of generators with a simple structure. The Hecke operators then provide a finer decomposition of $A^0_{\Gamma,\mathcal{I}}$ than that provided by the action of G.

There is a better way to explain all this. We change notation. Now G is a reductive algebraic group over $\mathbb{Q}$, $G_\infty = G(\mathbb{R})$ is its group of points with coordinates in $\mathbb{R}$, and for every prime number p we let $G_p = G(\mathbb{Q}_p)$. Then $G(\mathbb{Q}) \subset G_\infty \times \prod_p G_p$ is a discrete subgroup of the infinite product, but unfortunately this product is much too big – it is not locally compact, hence does not have a well-behaved measure theory nor theory of L_2-functions. Inside the infinite product is a locally compact subgroup $G(\mathbf{A})$ in which $G(\mathbb{Q})$ is still discrete. Then the coset space

$$G(\mathbb{Q})\backslash G(\mathbf{A})$$

is non-canonically equivalent to a finite union $\Gamma_i\backslash G_\infty$ of locally symmetric spaces of the kind considered above (also by reduction theory). But the version $G(\mathbb{Q})\backslash G(\mathbf{A})$ is much better, because it has an action of the locally compact group $G(\mathbf{A})$. We can define $A_{G,\mathcal{I}}$ to be the space of functions $F : G(\mathbb{Q})\backslash G(\mathbf{A}) \to \mathbb{C}$ in the kernel of $\mathcal{I}$, satisfying the moderate growth condition, and such that the space of functions $r(k)F$, for $k \in K \subset G_\infty$, is finite dimensional. This is the most general kind of automorphic form, and an automorphic representation is an irreducible $G(\mathbf{A})$-subquotient of $A_{G,\mathcal{I}}$. One defines the subspaces $A^2_{G,\mathcal{I}}$ and $A^0_{G,\mathcal{I}}$; the latter is a (Hilbert) direct sum of irreducible subrepresentations

$$A^0_{G,\mathcal{I}} = \oplus_\pi m(\pi)\pi$$

where π runs over all irreducible admissible representations of $G(\mathbf{A})$ and the π for which $m(\pi) > 0$ are the *cuspidal automorphic representations*, which are the atomic objects of the theory. I will not define “admissible”.

The Hecke operators are replaced by the representations of the local factors G_p of $G(\mathbf{A})$. This is done according to a few very simple rules. The first is that

Rule 1. *Any irreducible admissible representation of $G(\mathbf{A})$ is a (restricted) tensor product*

$$\pi_\infty \otimes {\bigotimes_p}' \pi_p$$

where π_∞ is an irreducible Harish-Chandra module for $(\mathfrak{g}, K)$ and each π_p is an irreducible admissible representation of G_p.

Rule 2. *In the above factorization, for almost all p, π_p is spherical: it has a vector fixed under a (hyperspecial) maximal compact subgroup K_p.*

If $G_p = GL(n, \mathbb{Q}_p)$, then the hyperspecial maximal compact subgroups are $K_p = GL(n, \mathbb{Z}_p)$ and its conjugates. The space of K_p-fixed vectors in π_p is of dimension 1 and is a module for the convolution algebra of compactly supported K_p-biinvariant functions on G_p; this is the *local Hecke algebra* $\mathcal{H}_p$ at p and is a commutative polynomial algebra on ℓ generators, where ℓ is the rank of a maximal split torus.

When $G_p = GL(n, \mathbb{Q}_p)$, this means there are n algebraically independent Hecke operators. For $n = 2$, one gets the usual Hecke operator of the theory of modular forms; the second is just the operator $K_p \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} K_p$ which determines the central character of π_p. The Hecke algebra is calculated in terms of invariant theory, but for the Langlands dual group ${}^LG$, a complex algebraic group, rather than for G itself. One has

$$\mathcal{H}_p \xrightarrow{\sim} \mathbb{C}[{}^LT_d]^W$$

where T_d is the maximal split torus in G, ${}^LT_d$ is the corresponding torus in ${}^LG$, and W is the relative Weyl group. If $G_p = GL(n, \mathbb{Q}_p)$, then ${}^LG = GL(n, \mathbb{C})$, ${}^LT_d$ is the diagonal maximal torus (up to conjugation), and W is the permutation group. The Hecke algebra is then isomorphic to the algebra of permutation-invariant functions on the diagonal torus. In other words, characters of the Hecke algebra are in one-to-one correspondence with diagonal matrices up to conjugation. This was one of Langlands’ first insights:

Rule 3. *The (equivalence classes of) irreducible spherical representations of* G_p *are in natural bijection with semi-simple conjugacy classes in* ${}^LG$.

What else is in bijection with semi-simple conjugacy classes in ${}^LG$? A tautological answer is the set of equivalence classes of homomorphisms $\mathbb{Z} \to {}^LG$ with semi-simple image. When $G = GL(n, \mathbb{Q}_p)$, this is the set of n-dimensional semi-simple representations of $\mathbb{Z}$. Langlands' next insight is that the relevant $\mathbb{Z}$ is not the abstract group $\mathbb{Z}$ but rather the natural generator $Frob_p$ of the Galois group $Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p)$.

Here is a quick review of the Galois theory of finite fields. Every finite field of characteristic p has order $q = p^f$ for some positive integer f, and then $Gal(\mathbb{F}_q/\mathbb{F}_p)$ has a natural generator, the arithmetic Frobenius operator $\phi_p : a \mapsto a^p$. Note that ϕ_p is a homomorphism of fields. Every $\mathbb{F}_q$ has the property that its multiplicative group is cyclic of order $q-1$, hence $\mathbb{F}_q \xrightarrow{\sim} \mathbb{F}_p(\mu_{q-1})$, and we can let $\overline{\mathbb{F}}_p = \cup_f \mathbb{F}_p(\mu_{p^f-1})$. With respect to this union, the operator ϕ_p acts consistently, hence defines an element $\phi_p \in Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p) = \varprojlim_f Gal(\mathbb{F}_{p^f}/\mathbb{F}_p)$ that is a topological generator of this profinite group. By convention we set $Frob_p = \phi_p^{-1}$. Thus

Rule 3'. *The (equivalence classes of) irreducible spherical representations of* G_p *are in natural bijection with conjugacy classes of homomorphisms* $< Frob_p >^{\mathbb{Z}} \to {}^LG$ *with image in the semisimple elements.*

The basis of algebraic number theory is the existence, for all primes p, of a short exact sequence

$$1 \to I_p \to Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p) \to 1$$

where I_p is the inertia group. The group I_p is complicated, but the right-hand term is very simple. We make it simpler still by defining the Weil group W_p to be the subgroup of $Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ that fits in this short exact sequence:

$$1 \to I_p \to W_p \to < Frob_p >^{\mathbb{Z}} \to 1.$$

A homomorphism from W_p to a group H is *unramified* if it is trivial on I_p; hence

Rule 3". *The (equivalence classes of) irreducible spherical representations of* G_p *are in natural bijection with conjugacy classes of* **unramified** *homomorphisms* $W_p \to {}^LG$ *with image in the semisimple elements.*

Local Langlands conjecture. *The (equivalence classes of) irreducible admissible representations of* G_p *are partitioned into finite* L**-packets** *parametrized by conjugacy classes of all homomorphisms* $WD_p \to {}^LG$ *such that any lift of* $Frob_p$ *has image in the semisimple elements.*

Here WD_p is the Weil-Deligne group, a slightly more complicated version of W_p; representations of W_p all define representations of WD_p.

Rule 4 (MH-Taylor, Henniart). *The local Langlands conjecture is true for* $G = GL(n, F)$ *(any* p*-adic field* F*).*

The corresponding local conjecture with W_∞ an extension of $\mathbb{C}^\times$ by $Gal(\mathbb{C}/\mathbb{R})$, is known and due to Langlands for any G_∞. We henceforward restrict attention to representations of W_p that extend to representations of $Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$; this makes the statements much simpler and is not a very serious restriction.

Thus, for $G = GL(n)$, each factor π_p of an automorphic representation π can be parametrized by something very close to a representation $\rho_p : Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to GL(n, \mathbb{C})$, with all but finitely many ρ_p unramified.

Langlands' version of Artin's conjecture. *Let $\pi = \pi_\infty \otimes \bigotimes_p \pi_p$ be an admissible irreducible representation of $GL(n, \mathbf{A})$. Then $m(\pi) > 0$ if and only if for all p, ρ_p has finite image, and there is an irreducible representation $\rho : Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL(n\mathbb{C})$ such that $\rho|_{Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)} \xrightarrow{\sim} \rho_p$.*

Here we need to know that there is a restriction map from representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ to $Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ for each prime p, or, more concretely, an inclusion $Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \hookrightarrow Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, well-defined up to conjugation. This is the local-global principle for algebraic number theory, just as if X is a punctured surface, then the loop around a puncture gives rise to an element of $\pi_1(X)$, well-defined up to conjugation.

Rule 5 (strong multiplicity one). *Let $G = GL(n)$. For any π, $0 \leq m(\pi) \leq 1$. Moreover, if π' and π are cuspidal automorphic representations of $GL(n, \mathbf{A})$ with the property that $\pi_p \xrightarrow{\sim} \pi'_p$ for all but finitely many p, then $\pi' = \pi$.*

For Galois representations, the corresponding property is the Chebotarev density theorem that says that an irreducible representation ρ as above is determined up to equivalence by the ρ_p for a set of primes p of Dirichlet density one.

There is a similar parametrization for general automorphic representations, based on Langlands' theory of Eisenstein series. For $G = GL(n)$, the parabolic subgroups are indexed by partitions $n = n_1 + \cdots + n_r$. The Eisenstein series for the corresponding P, with Levi factor $\prod_i GL(n_i)$, are parametrized by sums of irreducible Galois representations of dimensions n_i, $i = 1, \ldots, r$. Indeed, the analogous parametrization works for irreducible admissible representations of $GL(n, \mathbb{Q}_p)$. More generally, Eisenstein series on G are parametrized by (conjugacy classes of) representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ with values in a proper parabolic subgroup of ${}^LG$.

In any case, an automorphic representation $\pi = \pi_\infty \otimes \bigotimes_p \pi_p$ can always be labelled by the local Langlands parameters of the individual π_p (though multiplicity one fails for general groups). This labeling is equivalent to the determination of the eigenvalues of Hecke operators at almost all primes p, plus some additional information at the remaining primes and at ∞. The labeling is systematized in terms of the L-functions. Each π_p has a local L-function $L(s, \pi_p)$. If π_p is unramified and associated to the representation ρ_p of the Weil group, then

$$L(s, \pi_p) = det(1 - \rho_p(Frob_p)p^{-s})^{-1}.$$

In general there is a modified version. One knows that

$$L(s, \pi) = \prod_p L(s, \pi_p)$$

satisfies a functional equation:

$$\Lambda(s, \pi) = \varepsilon(s, \pi)\Lambda(1 - s, \pi^\vee).$$

To define $L(s, \pi)$ one multiplies $L(s, \pi)$ by the Γ factors, which can be defined in terms of the local Langlands parameters for π_∞, and the local ε factors are also determined by the local Langlands parametrization. Note that the L-function can also be defined for the Galois representation ρ If $G = GL(n)$ with $n > 1$, or if $n = 1$ but π is not trivial, then $L(s, \pi)$ is an *entire* function. Thus Langlands' conjectures include the classical

Artin conjecture. *the L-function of an irreducible non-trivial complex representation of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ is entire.*

Indeed the only known way to prove that the L-function of a Galois representation is entire is by proving that the Galois representation is automorphic. The Artin conjecture is one of the outstanding open questions in algebraic number theory.

However, not all automorphic representations of $G = GL(n)$ are indexed by global representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, and even among those that are, the representations with values in $GL(n, \mathbb{C})$ – the *Artin* representations – form a minority. In the remainder of this talk I will state a conjecture about automorphic representations of $GL(n)$ and Galois representations, and say a few words about other groups.

Say $\pi \subset A^2_{G,\mathcal{I}}$ is *cohomological* if $\mathcal{I}$ is the annihilator of a finite-dimensional representation of G (for any G). This condition is equivalent to the condition that $\pi_f = \bigotimes'_p \pi_p$ is realized as a subspace of the cuspidal cohomology of the (adelic) locally symmetric space $G(\mathbb{Q})\backslash G(\mathbf{A})/K$ attached to G. If $G = GL(n)$ and π is cohomological, then one always expects there to be an associated continuous representation $\rho : Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL(n, V)$ but now V is an ℓ-adic vector space where ℓ is any prime number. In fact, there should be a (ρ_ℓ, V_ℓ) for each prime ℓ, but with the property that the characteristic polynomial of $\rho_\ell(Frob_p)$ is independent of ℓ, for every prime p such that π_p is unramified, and of course one expects $\rho_\ell(Frob_p)$ to be the local Langlands parameter of π_p. (We have to identify ℓ-adic and complex matrices, but I ignore this problem, which has a rigorous foundation.) The collection (ρ_ℓ, V_ℓ) is then a *compatible system of ℓ-adic representations*, in the sense of Serre, and should indeed be associated to the representations on ℓ-adic cohomology of some algebraic variety.

Theorem. *If π is cohomological and $\pi \xrightarrow{\sim} \pi^\vee \otimes \xi_n$ where $\pi^\vee$ is the contragredient and ξ_n is an appropriate character factoring through the determinant, then the associated (ρ_ℓ, V_ℓ) exist and satisfy all the expected properties.*

This theorem has been proved in many stages: for $n = 2$ it was begun by Eichler and Shimura, continued by Deligne, and completed by Langlands and Carayol. For $n > 2$ it was begun by Clozel, continued in my book with Taylor, and nearly completed in the Paris automorphic forms book project; the final steps ("all the expected properties") were proved by Taylor's students Shin and Caraiani, the latter just last fall.

Conversely, a compatible family (ρ_ℓ, V_ℓ) satisfying "all the expected properties" is expected to come from a cohomological π. Weak versions of this have been proved in recent years by a number of people; for $n = 2$ this is the method of Taylor and Wiles, extended by Kisin and Emerton, among others; for $n > 2$ this was begun in my joint work with Clozel, Shepherd-Barron, and Taylor, and continued in subsequent years, culminating in an important paper by Taylor and three collaborators. Similar results are known for $GL(n, F)$ whenever F is a totally real field or a totally imaginary quadratic extension of a totally real field.

The need for ℓ-adic coefficients, rather than complex coefficients, is concealed in the factor π_∞. The codimension 1-ideals $\mathcal{I} \subset Z(\mathfrak{g})$ are parametrized, just like the Hecke operators, by characters of $\mathbb{C}[\mathfrak{t}]^W$ where now $\mathfrak{t}$ is the (diagonal) Cartan subalgebra of $Lie(G_\infty)$ and W is again the permutation group. Thus the ideals are parametrized by points in $\mathfrak{t}^*/W$, or again by points in the closed positive Weyl chamber $\bar{C}^+$. The integral points $\lambda \in \bar{C}^+$ give rise to the annihilators of finite-dimensional representations by sending $\lambda \mapsto \lambda + \rho$ where

$$\rho = (\frac{n-1}{2}, \frac{n-3}{2}, \ldots, \frac{1-n}{2})$$

is half the sum of the positive roots. An ideal $\mathcal{I}$ is called *algebraic* if its parameter lies in $(\mathbb{Z}^n + \rho) \cap \bar{C}^+$; this includes the cohomological ideals but also some whose parameters lie on the boundary of C^+.

Conjecture (Clozel). *If π_∞ is algebraic then there is an associated family of ℓ-adic Galois representations.*

There is exactly one $\mathcal{I}$ for which the corresponding π are expected to give rise to Artin representations; a continuous representation with values in $\mathbb{C}$ in fact factors through a finite quotient $Gal(L/\mathbb{Q})$ of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, for some L, and therefore defines continuous representations with values in any field. Almost nothing is known about this more general conjecture.

Finally, for general G, a similar theorem is expected, with n-dimensional representations replaced by homomorphisms to a (modified version of) the L-group; it is just the L-group in the case of Artin representations.