Representation Theory and Arithmetic *

Although some of the books of Hermann Weyl, especially those dealing with algebraic matters, are notoriously difficult, the papers on geometry and analysis were often models of ease and transparency, as much in the incidental papers as in the major ones, like those on the spectral theory of ordinary differential equations or the representation theory of compact Lie groups.

This lecture is a brief introduction to some problems in the contemporary theory of automorphic forms, a part of the spectral theory of group actions, a topic that perhaps began with the theorem of Peter-Weyl on the representation theory of general compact groups; but the clue to the present investigations, and indirectly the major link to Hermann Weyl, is provided by the spectral theory of Harish-Chandra for non-compact semisimple groups. The influence of Weyl's techniques for studying characters and of the spectral theory of ordinary differential equations is manifest throughout the work of Harish-Chandra. Specifically, however, the clue is given by the geometrical and cohomological properties of the discrete series.

None the less our major concerns will be arithmetical and owe more to Weyl's fellow student Hecke than to Weyl himself, for two subjects that began with Hecke play the principal roles, the extension of the theory of complex multiplication to higher-dimensional varieties, a subject that has become the theory of Shimura varieties, and the theory of Hecke operators and the associated L-series. Even so, Weyl was fascinated by arithmetic from the beginning of his career, Hilbert's *Klassenkörperbericht* being one of the first papers he read as a student, and, as his monograph on ideal theory and other papers testify, it continued to attract his interest until the end.

I begin by recalling some familiar, but fundamental and ultimately very difficult concepts. We begin with a smooth projective variety V over a finite field k. If k_n is the extension of k of degree n, let N_n be the number of points on V with coefficients in k_n, and form

$$Z(t,V) = \exp\left(\sum_{\mathrm{n}=1}^{\infty} \frac{\mathrm{N_n}}{\mathrm{n}} \mathrm{t^n}\right).$$

It is the zeta-function of V introduced by Weil.

If, for example, k has q elements and V is the projective line, then $N_n = q^n + 1$ and

* Appeared in *Proceedings of Symposia in Pure Mathematics 48*, Amer. Math. Soc., Providence, RI (1988)

$$Z(t,V) = \exp\left(\sum_{n=1}^{\infty}\frac{(qt)^n}{n} + \sum_{n-1}^{\infty}\frac{t^n}{n}\right) = \frac{1}{(1-qt)(1-t)}.$$

It is by now very well known that for any variety V the function $Z(t,V)$ is a rational function of t of the form

$$Z(t,V) = \prod_{0\leq i\leq 2\dim\, V} L_i(t,V)^{(-1)^{i-1}},$$

where $L_i(t,V)$ is a polynomial

$$L_i(t,V) = \prod_{j=1}^{d_i}(1-a_{ij}t).$$

In addition, $|a_{ij}| = q^{i/2}$ and d_i has cohomological significance.

If we take a variety of V over a global field F, in particular over **Q**, then V will be defined by a finite number of equations with coefficients that are integral outside a finite set S of primes, and thus can be reduced modulo any prime not in S, and if S is taken to be sufficiently large will even give upon reduction a smooth variety over the residue field and thus a zeta-function

$$Z(t,V;\mathfrak{p}) = \prod_i L_i(T,V;\mathfrak{p})^{(-1)^{i-1}}.$$

It has been suggested, somewhat casually and in specific cases by Hasse and then systematically, and independently, by Weil, that the Euler products

$$L_i(t,V;S) = \prod_{\mathfrak{p}\notin S}\frac{1}{L_i(N\mathfrak{p}^{-s},V)}$$

would be of interest. For example, if V is just a point, the global field is **Q**, and if S is empty then $L_0(t,V;S)$ is simply the Riemann zeta-function.

In general these functions are of interest for at least two reasons.

(i) They pose an obvious problem of analytic continuation.

(ii) Although the functions are defined in terms of local data, they yield information about the global arithmetic of the variety. For example, for varieties of dimension zero this is expressed by the classical class-number formulas and for elliptic curves by the conjectures of Birch and Swinnerton-Dyer.

The problem (i) is of course patent and in comparison with those posed by the ideas implicit in (ii) puerile. None the less it leads not only to serious analytic questions but also to serious arithmetic questions. Even for varieties of dimension 0 it requires class-field theory to solve it even in part.

Depth aside, it is certain that the problem is solved in very few cases:

(i) varieties of dimension zero associated to abelian extensions;

(ii) abelian varieties with complex multiplication, in particular, for elliptic curves with complex multiplication but not, except for a few isolated examples, for other elliptic curves.

Thus even for curves there is a great deal left to do. There is one class of curves for which much is known, the modular curves, and more generally Shimura curves. A fairly general family of modular curves is obtained by dividing the upper half-plane by the discrete groups

$$\Gamma_N = \left\{\gamma \in SL(2,\mathbf{Z}) | \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod N\right\}.$$

The associated complex algebraic curve Sh_N can be made projective by adding a finite number of points and then given a structure over **Q**.

It is possible to show that $L_1(s, Sh_N; S)$ can be analytically continued by showing that it is a product of the L-functions attached by Hecke to automorphic forms on the upper half-plane, which are of the form

$$L_S(s,\pi) = \prod_{\mathfrak{p}\notin S} \frac{1}{(1-\alpha_p/p^s)(1-\beta_p/p^s)},$$

and thus of degree two. Here π denotes the form or, what amounts to the same thing, the associated representation. Thus

$$L_1(s, Sh_N; S) = \prod_{\pi} L_S(s - 1/2, \pi), \tag{1}$$

only a finite number of π, and these not necessarily distinct, intervening in the product.

Such a result poses further problems, for if one of the curves Sh_N appears as a ramified covering of some curve C, not itself an Sh_N, then one may hope and expect to deduce from (1) a similar representation for $L_1(s, C; S)$, and thus verify that it too can be analytically continued. This is the method that has been proposed—for very good reasons—for dealing with elliptic curves. In order to deal with other base fields, one needs a theory of base change for automorphic forms, but that is only partially developed [L2] and not pertinent to this lecture. It is more important to stress that the methods that lead to (1) and its refinements are also important for apparently quite different arithmetic problems, like the structure of the ideal-class group of cyclotomic fields [M-W].

There is another class of varieties for which an analogue of (1) is valid, those attached to the names of Hilbert and Blumenthal. They can be of any dimension, but the surfaces of this type—associated to real quadratic fields—are perhaps of most interest at the moment because for them a number of important conjectures can be tested with the help of (1), the conjectures of Tate relating algebraic cycles to the Galois action of étale cohomology and to the order of the poles of the Hasse-Weil zeta-function [HLR] and the conjectures of Beilinson [Ra].

All this is by way of preface to stress the importance of the problem of analytic continuation and to observe that its solution even for what appear to be very special varieties can lead to unpredictable and valuable arithmetic consequences.

The one class of varieties that offers hope for substantial advances is that of Shimura varieties. There are several problems involved and on all but one progress was being made, especially by R. Kottwitz, but there is one central obstacle that it was not clear would be removed in the near future, so that I feared that like Jean Débardeur we would remain “toujours à terre, jamais au large”, but the obstacle has now been removed by Kottwitz himself [K6], and by H. Reimann and T. Zink as well [RZ]. These are important developments, and the purpose of this lecture is to draw attention to them.

There are three types of Shimura varieties to be distinguished:

(a) the general type;

(b) those associated to a moduli problem for abelian varieties with endomorphism algebra and polarization;

(c) those associated to the Siegel upper half-spaces.

The problems can be posed for all of them, but it is often a major step to pass from the solution for those of type (b) to the general solution for those of type (a). At the moment one is attempting only to deal with those of type (b). The methods that work for those of type (c) usually work for those of type (b) with little change. Thus I confine myself to type (c).

The Shimura varieties associated to the Siegel upper half-spaces are, properly speaking, attached to the group of symplectic similitudes, the group G of $2n \times 2n$ matrices U for which

$$ {}^tUJU = \lambda J, $$

λ a scalar, and

$$ J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. $$

To describe, even approximately, the form that (1) is expected to take we have to introduce at the same time the L-group ${}^LG$ of G. The group G is a group over **Q**; the L-group is in contrast a group over **C**. It is the Clifford group attached to the orthogonal form in $2n+1$ complex variables. The spin representation of the corresponding orthogonal group is of dimension 2^n and ${}^LG$ consists of all matrices that can be written as the product of a scalar matrix and an element of the spin group, so that ${}^LG$ has a natural representation r of degree 2^n.

According to the general definition of L-functions associated to automorphic forms there is attached to every finite-dimensional representation ρ of ${}^LG$ and every automorphic representation π of G an Euler product

$$ L_S(s, \pi, \rho). \tag{2} $$

Here S is some large finite set of primes of **Q**.

The Euler products attached to $\rho = r$ are of particular importance for the zeta-functions of the Shimura varieties attached to G. Questions of completeness and connectedness aside, these are as complex manifolds essentially quotients $\Gamma\backslash H$, where Γ is a congruence subgroup of $G(\mathbf{Z})$ and H is the set of all complex symmetric matrices $Z = X + iY$ with $Y > 0$ and

$$ \gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} : Z \to (AZ+B)(CZ+D)^{-1}. $$

The structure of these varieties over **Q**, or over a number field if that is appropriate, is given by the theory of Shimura, completed by Deligne [D].

To obtain a Shimura variety in the proper sense, one must in fact take the disjoint union of several of these varieties, obtaining for this particular group varieties over **Q**. The question of completeness is more vexing, and forces us to enlarge the notion of a zeta-function with the help of intersection cohomology to deal with singular varieties. The conjecture of Zucker, proved by Looijenga and Saper-Stern, allows one for many purposes to argue as though the quotients $\Gamma\backslash H$ were compact, and in order to arrive without too much delay at the problems that have actually been settled we do so here.

The bulk of the cohomology of the Shimura variety Sh is contained in the middle dimension $q = n(n+1)/2$ and if calculated by means of the theory of continuous cohomology [BW] is given by the discrete-series representations of $G(\mathbf{R})$ that annihilate the Casimir operator. The set Π_∞ of such representations V has 2^{n-1} elements. If π_∞ is one of them, and if K is the open compact subgroup of the adelic group $G(\mathbf{A}_f)$ that must be introduced when Sh is defined completely, then each time that an automorphic representation $\pi = \pi_\infty \otimes \pi_f$, π_f being an irreducible representation of $G(\mathbf{A}_f)$, occurs in $L^2(G(\mathbf{Q})Z(\mathbf{R})\backslash G(\mathbf{A}))$ there is a contribution to the cohomology in degree q of dimension $2d(\pi_f^K)$. We denote by $d(\pi_f^K)$ the dimension of the space of vectors fixed by K under π_f. The critical observation is that $2 \cdot 2^{n-1} = 2^n$, the dimension of r and thus the degree of the Euler product $L_S(s, \pi, r)$.

If $\pi' = \pi'_\infty \otimes \pi_f$, where $\pi'_\infty \in \Pi_\infty$ then, by definition,

$$L_S(s, \pi', r) = L_S(s, \pi, r).$$

(This would be valid even if Γ-factors had been incorporated into the L-functions.) Thus if, as is often but not always the case, whenever $\pi_\infty \otimes \pi_f$ occurs in $L^2(G(\mathbf{Q})Z(\mathbf{R})\backslash G(\mathbf{A}))$ then $\pi'_\infty \otimes \pi_f$ also occurs for any $\pi'_\infty \in \Pi_\infty$ then the representations $\{\pi_\infty \otimes \pi_f | \pi_\infty \in \Pi_\infty\}$ contribute a space of dimension $2^n d(\pi_f^K)$ to the cohomology each time that they occur, and thus should contribute a factor of degree $2^n d(\pi_f^K)$ to the L-function $L_q(s, Sh; S)$. If the Eichler-Shimura theory for the upper half-plane which leads to (1) is kept in mind, then a natural guess is that this factor is

$$L_S(s - q/2, \pi_f, r)^{d(\pi_f^K)}. \tag{3}$$

The shift by $q/2$ is to account for the absolute value of the roots of the local L-functions.

There are two distinct questions implicit here: (a) can the Euler products (2) be analytically continued; (b) can the zeta-function of the variety Sh really be expressed in terms of these functions? These are two very different aspects of the problems posed by the introduction of the general Euler products into the theory of automorphic forms. The problem of analytic continuation can be approached in various ways [GS] and is in particular tied to functoriality, so that although a great deal remains to be done, it is clear that we are dealing with promising material methods [AC].

The question (b) emphasizes a distinct consideration. Even if the functions (2) have interesting analytic properties and lead to an internally rich theory of automorphic forms, is it a theory that bears on other domains of mathematics, in particular, on arithmetic? At first, after the Eichler-Shimura theory, an almost but not quite decisive response to this question is to show that the zeta-functions of Shimura varieties can be expressed in terms of these functions, for then we may hope that even those varieties not defined by groups have zeta-functions that can be so expressed.

We are here concerned with question (b), which requires that we give a precise expression for the zeta-function as a product of the functions (2) (and their inverses) and that we prove it. Since the precise expression is not so important, simply whatever the proof yields, it is the strategy of the proof that counts, and that is elaborate. It has to be recognized immediately that the occurrence of $\pi_\infty \otimes \pi_f$, $\pi_\infty \in \Pi_\infty$, in $L^2(G(\mathbf{Q})Z(\mathbf{R})\backslash G(\mathbf{A}))$ does not always entail the occurrence of $\pi'_\infty \otimes \pi_f$ with the same multiplicity. This is the subject of endoscopy and the stable trace formula, which have only begun to be developed [K1, K2, L3, LS, Ro]. Our experience so far [L1] suggests that there are subgroups ${}^L H \hookrightarrow {}^L G$, attached to groups H over $\mathbf{Q}$, and that $r' = r|{}^L H$ decomposes into a direct sum $\oplus r_i$ of irreducible representations, so that for a representation π obtained by functoriality from a representation π' of $H(\mathbf{A})$ there is a factorization

$$L(s - q/2, \pi, r) = L(s - q/2, \pi', r') = \prod_i L(s - q/2, \pi', r_i)$$

and that it is not $L(s - q/2, \pi, r)$ that occurs in the zeta-function but only some of the factors $L(s - q/2, \pi', r_i)$.

To compare two L-functions, and that is what one is attempting, it is simpler to compare their logarithms, or rather for each p and n the coefficients of $1/p^{ns}$ in the expansion of their logarithms.

On one side, for the product of automorphic L-functions, this will turn out to be a sum

$$\sum_H c_H ST(f_H), \tag{4}$$

where f_H is a function in $H(\mathbf{A})$ that depends on p and n and ST denotes the stable trace.

For the zeta-function this is, apart from difficulties with the cusps,

$$N_{p,n} \tag{5}$$

the number of points on the variety with coefficients from F_{p^n}.

To compute (4) we use the stable trace formula, which in principle expresses (4) as a sum over stable conjugacy classes in the various H and thus as a sum over conjugacy classes in G. Thus to make the comparison we need a method of calculating $N_{p,n}$ as a similar sum.

Now, to reach this stage, we have had to proceed as though some developments that were only beginning had been carried successfully to completion, but at least they have been inching forward. Until the recent work of Kottwitz and Zink, however, $N_{p,n}$ offered quite different difficulties, and there were some who felt that we were dealing with a problem that would remain for the forseeable future intractable.

There are two things to be done: (i) to find a group-theoretical description of the points on the variety with coefficients in F_{p^n} that allows one to calculate $N_{p,n}$ in terms of G; (ii) to put the resulting expression in a form that can be compared term-by-term with the expansion of (4). Kottwitz had already shown that step (ii) could be effected by the fundamental lemma for the endoscopic groups for base change [K4], and thus reduced to a problem in harmonic analysis for which at least some serious progress could be made [K5, AC]. In addition he had isolated the algebro-geometrical problem that has to be regarded as the irreducible form of (i), namely to show that an invariant introduced by him, and referred to in [LR] as the Kottwitz invariant, was 1 for abelian varieties over finite fields. Only recently have Kottwitz himself [K6] and Reimann-Zink [RZ] succeeded in showing that this is so, thus overcoming what seemed to me the major obstacle to a successful treatment of the zeta-function of Shimura varieties, so that, in spite of the many difficulties that remain and that I hope have not been slighted here, we can at last be sanguine about the prospect of obtaining utilizable results in the not-too-distant future.

The Kottwitz invariant for the group of symplectic similitudes G is attached to a triple $(\gamma, \delta, \varepsilon)$. Here ε lies in $G(\mathbf{Q})$, is elliptic in $G(\mathbf{R})$, and

$$\langle \varepsilon x, \varepsilon y\rangle = c(\varepsilon)\langle x, y\rangle, \quad |c(\varepsilon)|_p = |q|_p,$$

$$q = p^r, r > 0, \quad \langle x, y\rangle = {}^t xJy.$$

Moreover $\gamma = \{\gamma_l | l \neq p\}$, $\gamma_l \in G(\mathbf{Q}_\lambda)$ and γ_l is conjugate to ε in $G(\overline{\mathbf{Q}_l})$ for all l and in $G(\mathbf{Q}_l)$ for almost all l. If F is the unramified extension of $\mathbf{Q}_p$ of degree r and σ the Frobenius element in $\mathrm{Gal}(F/\mathbf{Q}_p)$ then $\delta \in G(F)$ and

$$\delta\sigma(\delta)\cdots\sigma^{r-1}(\delta)$$

is conjugate to ε in $G(\overline{\mathbf{Q}}_p)$.

The associated invariant $k(\gamma, \delta; \varepsilon)$ is of cohomological nature, and is most easily defined when the centralizer of ε in G is a torus I. Suppose $\Gamma = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The invariant takes values in the dual of $\pi_0(\hat{I}^\Gamma)$, the connected component of the group of Γ-invariant elements in $\hat{I}$. The group $\hat{I}$ is that complex torus on which Γ acts in such a way that

$$\mathrm{Hom}(\hat{I}, G_m) \simeq \mathrm{Hom}(G_m, I)$$

is a homomorphism of Γ-modules.

If v is a place of $\mathbf{Q}$ let $\Gamma_v \subseteq \Gamma$ be $\mathrm{Gal}(\overline{\mathbf{Q}_v}/\mathbf{Q}_v)$. The invariant is a product $\Pi_v\beta(v)$, where $\beta(v)$ is a homomorphism from $\hat{I}^{\Gamma_v}$ to $\mathbf{C}^\times$, or properly speaking the restriction of such a homomorphism to $\hat{I}^\Gamma$. In the definition of $\beta(v)$ three types of places are distinguished.

(i) If $v = l \neq p$, then

$$\gamma_l = c\varepsilon c^{-1}, \quad c \in G(\overline{\mathbf{Q}_l}).$$

Since both γ_l and ε lie in $G(\mathbf{Q}_l)$, the cochain

$$\{c^{-1}\sigma(c)\}$$

defines an element of $H^1(\mathbf{Q}_l, I)$ and thus by Tate-Nakayama theory a homomorphism from $I^{\hat{\Gamma}_v}$ to $\mathbf{C}^\times$.

(ii) For $v = p$ we write

$$\delta\sigma(\delta)\cdots\sigma^{r-1}(\delta) = c\varepsilon c^{-1}, \quad c \in G(\mathbf{Q}_p^{un}),$$

and then

$$b = c^{-1}\delta\sigma(c) \in I(\mathbf{Q}_p^{un}).$$

In [K3] Kottwitz associates to this b a coweight of $\hat{I}^{\Gamma_v}$. It is taken as $\beta(v)$.

(iii) If $v=\infty$ then $I(\mathbf{R})\cap G_{sc}(\mathbf{R})$ is compact Cartan subgroup of $G_{sc}(\mathbf{R})$, the symplectic group. All of these are conjugate and possess a standard coweight that is used to define $\beta(\infty)$.

Precise general definitions can be found in [K6] and [LR]. To pass from the Kottwitz invariant for triples to the Kottwitz invariant for abelian varieties with polarization, observe that if the variety and the polarization are defined over a field with q elements then the l-adic cohomology together with the Frobenius endomorphism yields γ_l, $l\neq p$, so that γ is defined. The element δ is provided by the Dieudonné module attached to the variety. All the γ_l have the same eigenvalues. They are algebraic numbers and there is at least one element of $G(\mathbf{Q})$ with these eigenvalues. Any such element serves as ε, and the geometric theorem essential to the calculation of the $N_{p,n}$ is that for triples arising in this way the invariant is 1.

The argument of Kottwitz has a strong functorial flavor and uses Fontaine's theory for Galois modules attached to p-divisible groups, while Reimann and Zink use more explicit methods based on classifications of group schemes over finite fields due to Raynaud.

References

[AC] J. Arthur and L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Ann. of Math. Studies (to appear).

[BW] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Ann. of Math. Studies (1980).

[D] P. Deligne, *Variétés de Shimura: Interprétation modulaire, et techniques de contruction de modèles canoniques*, Proc. Sympos. Pure Math., vol. 33, part 2, Amer. Math. Soc., Providence, RI, 1979, pp. 247–289.

[GS] S. Gelbart and F. Shahidi, *Analytic properties of automorphic L-functions* (to appear).

[HLR] G. Harder, R.P. Langlands, and M. Rapoport, *Algebraische Zyklen auf Hilbert-Blumenthal-Flächen*, J. Reine Agnew. Math. **366** (1986).

[K1] R. Kottwitz, *Stable trace formula: cuspidal tempered terms*, Duke Math. J. **57** (1984).

[K2] ________ , *Stable trace formula: elliptic singular terms*, Math. Ann. **275** (1986).

K3] ________ , *Isocrystals with additional structure*, Compositio Math. **56** (1985).

[K4] ________ , *Shimura varieties and twisted orbital integrals*, Math. Ann. **269** (1984).

[K5] ________ , *Base change for unit elements of Hecke algebras*, Compositio Math. **60** (1986).

[K6] ________ , in preparation.

[KS] R. Kottwitz and D. Shelstad, in preparation.

[L1] R.P. Langlands, *On the zeta-functions of some simple Shimura varieties*, Canad. J. Math. **31** (1979).

[L2] ________ , *Base change for $GL(2)$*, Ann. of Math. Studies (1980).

[L3] ________ , *Les débuts d'une formule des traces stable*, Publ. de l'Univ. Paris VII, No. 13 (1983).

[LR] R.P. Langlands and M. Rapoport, *Shimuravarietäten und Gerben*, J. Reine Agnew. Math. **378** (1987).

[LS] R.P. Langlands and D. Shelstad, *On the definition of transfer functors*, Math. Ann. **278** (1987).

[MW] B. Mazur and A. Wiles, *Class fields and abelian extensions of* **Q**, Invent. Math. **76** (1984).

[Ra] D. Ramakrishnan, *Valeurs de fonctions L des surfaces d'Hilbert-Blumenthal en $s = 1$*, C. R. Acad. Sci. Paris, Sér. I Math. **301** (1985).

[RZ] H. Reimann and T. Zink, *Der Dieudonnémodul einer polarisierten abelschen Varietät vom CM-Typ*, Ann. of Math. (2) (to appear).

[Ro] J. Rogawski, *Automorphic representations of unitary groups in three variables*, in preparation.