ENDOSCOPY AND BEYOND[†]

I begin with a description of some of the basic general problems in the theory of automorphic forms. They are already familiar to the participants in this seminar.

Functoriality and Related Matters

The following statement is a rough formulation of the problem of functoriality.

> (I) *If H and G are two reductive groups over the global field F and the group G is quasi-split then to each homomorphism*
>
> $$\phi: \quad {}^L H \longrightarrow {}^L G$$
>
> *there is associated a* transfer *of automorphic representations of H to automorphic representations of G.*

The word *transfer* anticipates the complete solution of the problem; it has not been generally accepted, and the word *lift* is frequently used instead. Whether those who are fond of this terminology would want to use it in all cases is not clear. The transfer associated to the determinant, which maps the L-group of $GL(n)$ to the L-group of $GL(1)$, takes a representation π to its central character, an unlikely *lift*. At all events, transfer would have the property that

$$L(s, \pi, r \circ \phi) = L(s, \Pi, r),$$

if Π is the transfer of π. The formulation (I) does not take difficulties associated with L-packets and endoscopy into account.

A second problem that arose some time after functoriality is that of associating to an automorphic representation π, now on the group G, a subgroup H_π of ${}^L G$ that would only be defined up to conjugacy but would have the following property.

[†]Notes for a lecture delivered by R. Langlands, IAS, March 30, 2000. I have left these notes as they were when distributed at the lecture. They are not free of errors and false starts.

(II) *If r is a representation of ${}^LG$ then the multiplicity of the trivial representation of H_π in the restriction of r to H_π is the order of the pole of $L(s,\pi,r)$ at $s=1$.*

Once again, this is not intended as an absolutely precise statement.

Some Touchstones

There are three. I take for expository purposes the ground field F to be an arbitrary number field (of finite degree).

(T1) *Take H to be $GL(2)$, G to be $GL(m+1)$ and ϕ to be the mth-symmetric power representation.*

(T2) *Take H to be the group consisting of a single element and G to be $GL(2)$. Then ${}^LH$ is a Galois group and problem* (I) *is that of associating an automorphic form to a two-dimensional Galois representation.*

(T3) *Take G to be $GL(2)$ and π to be an automorphic representation such that at every infinite place v of the π_v is associated to a two-dimensional representation not merely of the Weil group but of the Galois group over F_v. Show that H_π is finite.*

A positive solution of the first problem has as consequence the Ramanujan-Petersson conjecture and the Selberg conjecture in their strongest forms; the Artin conjecture follows from the second. As is well-known, all these problems have been partially solved; some striking results for the first problem are very recent. For various reasons, the partial solutions all leave from a methodological point of view a great deal to be desired.

As far as (T1) is concerned, a fundamental strategic error was, I now believe, introduced at the very beginning of functoriality. I am perhaps as responsible as anyone. Recognizing that functoriality would entail the solution of these standard problems of analytic number theory, I inferred, perhaps unwisely, that all attention should be directed to functoriality. It would perhaps have been better to conclude that if functoriality had consequences for algebraic number theory and, especially, analytic number theory, then analytic number theory should be a part of all efforts to establish functoriality. As an excuse, I observe that at the time no-one had any clear idea of the possibilities or the limits of the trace formula, which was then the most promising tool.

The problem (T2) has, for example, been solved when the image of the Galois group is solvable, thus not icosahedral. This is a consequence of base change, a problem that is almost trivial for $GL(1)$, requiring only Hilbert's Theorem 90. Even for $GL(2)$ it requires no more in the way of number theory, although it does require some analysis not needed for $GL(1)$. Since we are attempting to construct a generalization of abelian class field theory, it is fair to expect that we need, when dealing with more difficult extensions, to combine the analysis with some number theory at least at the level of that required for $GL(1)$. Thus some analogue of Kummer extensions need to be studied. This has never been attempted; we have refused to invest enough resources in the problem (T2). I would like to think that the problems (T2) and (T3) were to be treated simultaneously, but this is by no means clear. Indeed it is at first glance doubtful, but I shall come back to this point at the very end of the lecture. The problem (T3) was not originally a part of functoriality, and without the evidence made available by, for example, the Deligne-Serre theorem, I would find it far-fetched to suggest that from a knowledge of π_∞ alone, it might be possible to deduce information about H_π.

In conclusion, and this is the point of view I want to represent today, any serious attempt to solve the problems of functoriality, in particular these three standard problems, should enlist, at a very early stage, the resources of analytic number theory – in my view this has *not yet* been done – and should recognize that its algebraic techniques have to encompass those of abelian class field theory. There has been an understandable tendency, of which I certainly am guilty, to exploit the formal consequences of class field theory while ignoring the concrete ingredients.

On the other hand, we can begin to contemplate developments beyond the trace formula only because the trace formula has been sufficiently developed that we understand, even if not so clearly as we might wish, its consequences and its limitations. Serious problems remain before it yields even a part of that of which it is capable, but serious progress is being made by serious people, and some confidence in their ultimate success is justified. Moreover, in my view, further progress will not be made by turning away from the trace formula, but by going beyond it. In particular, we are going to have to deal with comparisons of spectra and Arthur's techniques

I balanced all, brought all to mind,

will be indispensable. One only knows what to do with a trace formula when both sides of both formulas are there and the search for *small* comparable objects in the two formulas has begun. So I shall spend some time on a brisk and superficial review of Arthur's efforts to date, before turning to some very tentative suggestions, and then only in some very special cases, for going beyond them.

I am returning to Arthur's papers after a long absence, and do not yet really understand them. The review is not satisfactory. A satisfactory review would entail making two aspects of the papers clear: first the general structure of the argument which is cumulative, for the papers are almost without exception directed toward one specific end; and then the details of the arguments, isolating not only the techniques, often combinatoric or inductive, that reappear repeatedly but also various analytic difficulties specific to certain difficult stages. I shall not discuss the second aspect. Indeed I have not yet had the time to recall for myself the arguments I once knew or to understand the later ones. Since the second aspect is not discussed, there will be no occasion to be explicit about an otherwise important point. There are assumptions in Arthur's treatment, not only the fundamental lemma, but various other more analytic assumptions. These are all established in considerable generality, but not always in the generality envisaged, so that anyone contemplating applying Arthur's results for twisted trace formulas should note these assumptions carefully. Of course, there is, so far as I know, no reason to think they are out of reach. Anyone who wants to make himself useful might well consider trying to prove them.

You may take the suggestions as seriously as you like. The lecture is not intended to have any mathematical content; it is purely exhortatory. I want to encourage analytic number theorists to think more deeply about automorphic forms and specialists in the theory of automorphic forms to begin to exploit the deeper techniques of analytic number theory.

Outside Functoriality

There are of course problems in the theory of automorphic forms that lie outside functoriality, although they are related. None the less for the first of the two I list, many of the same methods are pertinent; for the second they are unlikely to be.

(P1) *Express the L-functions of a Shimura variety in terms of automorphic L-functions.*

(P2) *Show that the L-function of an elliptic curve over the number field F is equal to the standard automorphic L-function associated to an automorphic representation π of $GL(2)$*

Once again, both these problems are, so far as I know, only partially solved, the second with great *éclat*. There is a certain seepage between (T2) and (P2) that has led to some new results for (T2). Interesting as these results are, and important as it is to keep in mind any possibilities for counting in connection with (T2) suggested by (P2), I am afraid that in the long run it will turn out that to treat (T2) outside of functoriality is misguided, and what appears, unfortunately, to be a disingenuous attempt to separate (T2) from its origins in functoriality, labeling it without further explanation as a now, for unexplained reasons, generally accepted *strong* Artin conjecture, is to be deprecated as it can easily mislead the inexperienced reader about the true nature of the problem.

It may be useful to recall again, for the younger participants in this seminar, the pedigree of functoriality. More than three decades ago, it was observed that the theory of Eisenstein series gave rise to large classes of Euler products with analytic continuation, which could all be fit into a much larger class, defined by means of parameters that might not without reason be called the Hecke class or parameter, and which included the Artin L-functions. The *only* way to deal with the *abelian* Artin L-functions – then or now – is to establish that they are equal to *standard* L-functions – a term that I wish were restricted to those associated to $GL(n)$. Perhaps *basic* would be a more widely acceptable term. It was clear then, in essence from the ideas of Tamagawa as the complete theory of Godement-Jacquet remained to be worked out, that the basic analytic properties of these functions could be established. So the conjecture suggested itself that each and every one of these L-functions was equal to a standard L-function, and this led pretty much immediately to the general notion of functoriality, descended directly both from abelian class field theory and from Eisenstein series. It seems to me, for various reasons, that when assessing the utility of developing various proposed techniques, it is well to keep this dual pedigree in mind.

The Artin L-functions fly none the less two colors. They are automorphic L-functions – associated to the trivial group $\{1\}$ – and they are the simplest motivic L-functions – associated to motives of dimension zero. I suggest in this lecture that it may now be possible to approach three key problems related to functoriality and, indeed, functoriality in general, and therefore the analytic continuation of automorphic L-functions as well. The suggestion may or may not hold water. The problem of dealing with motivic L functions and thus with Hasse-Weil zeta-functions, of which (P2) is a particular manifestation, is another matter and may – my present view – or may not demand that Artin L-functions be dealt with first and may or may not require the use of Shimura varieties. I would love to hear the views of someone who has thought seriously about these questions.

Endoscopy

Until now, there have been three kinds of trace formula envisaged: what Jacquet has called the absolute trace formula, which is just the original trace formula for an approximation to the trace of

$$\int_{G(\mathbb{A})} f(g)R(g)dg,$$

R being the representation on the space of automorphic forms; the twisted trace formulas, whose importance in the case of base change was first recognized by Saito and then by Shintani and in the case of transfer from $SL(2)$ to $GL(3)$ by Jacquet; and the relative trace formulas, for which I refer you to a paper of Jacquet.

A) H. Jacquet, *Automorphic spectrum of symmetric spaces*, Proc. Symp. Pure. Math., vol. 61 (1996)

In all these trace formulas, endoscopy at first appears as a distracting impediment. The early applications, apart perhaps from its use to calculate various dimensions or traces explicitly, entailed a comparison of two geometric sides, thus often a comparison of, for example, conjugacy classes in two different groups. It turned out that conjugacy classes are not in general the right objects to compare, because one begins the comparison with two groups isomorphic or otherwise related only over an algebraic closure, so that there is no possibility of comparing objects attached to them but well defined only as objects over the ground field. Only objects defined over the algebraic closure have meaning for the comparison.

Now the basic principle in applications of the trace formula is the comparison of *small* terms on the geometric side, small meaning in essence simple enough that equality can be easily arranged and easily recognized, but these small terms have also to correspond bijectively so that they can then be cancelled one by one. What appears in the absolute trace formula are sums of orbital integrals over conjugacy classes, so that if the orbital integrals are chosen as the small objects, then the classes have to be in bijective correspondence. If they are not, then orbital integrals are not appropriate.

The solution is to stabilize the trace formula, so that the expressions of the absolute trace formula become a sum over stabilized trace formulas. Arthur's program is to use, in a manner to be explained, the twisted trace formula to establish the existence of a transfer from the classical groups to $GL(n)$. There is much to be done. The twisted trace formula has to be stabilized, but this being rather difficult and demanding even more from the local theory, he begins with the stabilization of the absolute trace formula. The basic local results that he has to assume deal with transfer and the fundamental lemma, both of which are difficult, although transfer has been proved by Waldspurger to follow from the fundamental lemma. (Some caution is called for as there is a fundamental lemma for the group, the original form, and one for the Lie algebra. I am not yet clear about their relation. Waldspurger seems to use that for the algebra.) The fundamental lemma, apart from some special cases, turns out to be a problem of quite a different nature than many other problems arising in the trace formula. It appears to be a problem in algebraic combinatorics, but a problem at a very high level that is beyond the purview of this lecture.

When the trace formula for a group G is expressed as a sum over stabilized trace formulas, these are stabilized trace formulas for lower dimensional groups, the endoscopic groups. For the absolute trace formula, these groups are in essence – for precision, see the papers referred to by Kottwitz and Shelstad – quasi-split groups H whose L-groups ${}^LH$ are the centralizers of semisimple elements s in ${}^LG$. For example, if $s = 1$, then the corresponding endoscopic group is the quasi-split group of which G is an inner twisting.

The twisted trace formula arises when one treats a group G with an outer automorphism θ. Take, for example as this is the case to which Arthur devotes his attention, G to be $GL(n)$ and θ to be

$$A \to J^t A^{-1} J^{-1},$$

with

$$J = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & -1 & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ (-1)^{n-1} & 0 & \dots & 0 & 0 \end{pmatrix}$$

Twisted endoscopy is treated in the monograph of Kottwitz and Shelstad.

B) R. Kottwitz and D. Shelstad, *Foundations of twisted endoscopy*, Astérisque (1999)

For the same reasons as in the absolute trace formula, the twisted trace formula has to be stabilized. The result is that the trace is expressed as a sum of stabilized absolute trace formulas. Such an expression can also be regarded as a comparison – of the twisted trace formula with absolute trace formulas – from which the spectral results are to be deduced.

For the absolute trace formula, there is, as we saw, a distinguished endoscopic group and comparison with it leads or, more often, is expected to lead to results about the transfer of automorphic representations on G to automorphic representations on its quasi-split inner form, of which a typical example is to be found in the final chapter of Jacquet-Langlands. There are also smaller endoscopic groups, and sometimes the stabilized trace formula can be used to establish transfer from these groups to the group G. For the twisted trace formula there is no one distinguished endoscopic group. There are several, perhaps an infinite number. To define endoscopic groups, we have first to attach to θ an automorphism $\hat\theta$ of ${}^LG$. For $GL(n)$, the automorphism $\hat\theta$ looks just like θ. Then an endoscopic group is basically a group H whose L-group ${}^LH$ in ${}^LG$ is the group of fixed points of an automorphism $Int(s)\circ\hat\theta$. Taking for example s such that $X = sJ$ is symmetric, this leads to the relation

$$sJ^tA^{-1}(sJ)^{-1} = A$$

or

$$X = AX^tA,$$

so that ${}^LH$ is symplectic or orthogonal and H orthogonal or symplectic. When more care is given to the definitions and the action of the Galois group taken into account, it appears that there are an infinite number of possibilities for H, all of the same maximal dimension. These are the principal endoscopic groups.

Thus, for example, the use of the twisted trace formula for $GL(2n)$ will lead to a comparison with a sum over all possible quasi-split forms of the orthogonal group $SO(2n)$, and there are infinitely many. There are also endoscopic groups of lower dimension, but I understand from Arthur that in this case, the lower-dimensional endoscopic groups do not lead to interesting transfers. What he proposes, by means of an elaborate inductive procedure, is to establish the existence of the transfer associated to the standard imbeddings of the L-groups of orthogonal and symplectic groups in that of $GL(n)$ with the aid of the twisted trace formula and the principal endoscopic groups, thus those of largest dimension. He has already done a great deal, but the final goal has not yet been reached, although, the fundamental lemma aside, he seems to have a clear idea of the procedure to be carried out.

There is at least one competing method for establishing the existence of such transfers, the converse theorem.

For both the absolute and the twisted trace formulas, the stabilization can only be carried out after the formulas have been rendered invariant. For the absolute, both steps have, so far as I can see, pretty much been completed. For the twisted trace formula, only the first is available. Since this review is to be brisk, I consider only the absolute formula.

Absolute Trace Formula

The first form of the absolute trace formula is established in the following two papers.

7) *A trace formula for reductive groups I: terms associated to classes in* $G(\mathbb{Q})$, Duke Jour. (1978)

9) *A trace formula for reductive groups I: applications of a truncation operator*, Comp. Math. (1980)

As a stark equality it is:

$$\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(f) = \sum_{\chi\in\mathfrak{X}} J_{\chi}^T(f)$$

In this formula f is a function with compact support in a certain subgroup $G(\mathbb{A})^1$ of $G(\mathbb{A})$, the kernel of a certain collection of characters (7, p. 917). The sum on the left is over classes in $G(\mathbb{Q})$, two elements in the group lying in the same class if their semisimple parts are conjugate. The sum on the right is over equivalence classes of pairs (M,ρ), where M is a Levi factor of a parabolic and ρ a cuspical representation of $M(\mathbb{A})^1$. According to the theory of Eisenstein series, to each such pair corresponds an invariant subspace of $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$. The trace formula presupposes of course the theory of Eisenstein series. The left side has been known since the special year on automorphic forms at the IAS in 1983/84 as the geometric side and the right side as the spectral side. In the seminars held during that year, the techniques for truncation introduced by Arthur for the absolute trace formula were extended to the twisted trace formula, along lines suggested, I believe, by ideas of Kottwitz. The speakers in the seminar, myself in particular, were, for reasons no longer of any importance, not aware of the source of these ideas.

The trace formula thus depends on a truncation, a truncation that requires a choice both of the parameter T, which is an element of the Lie algebra of an appropriate abelian subgroup of G, and of a minimal rational (over the given global field which can be taken to be $\mathbb{Q}$) parabolic P_0 that all parabolic subgroups that appear in the following are to contain. We fix a Levi factor M_0 of P_0 that is to lie in all Levi factors that appear.

It is proved that each term on both sides is a polynomial in T in the following paper.

10) *The trace formula in invariant form*, Ann. Math. (1981)

Then $J_{\mathfrak{o}}(f)$ and $J_{\chi}(f)$ are introduced by

$$J_{\mathfrak{o}}(f) = J_{\mathfrak{o}}^{T_0}(f), \quad J_{\chi}(f) = J_{\chi}^{T_0}(f),$$

for an appropriate value of $T = T_0$ (10, p. 19). The exact choice of T_0 is not of much importance but Arthur has a favorite choice. The trace formula then becomes

$$\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}(f) = \sum_{\chi\in\mathfrak{X}} J_\chi(f)$$

The Invariant Trace Formula

Arthur puts the trace formula in invariant form twice, first in (10) and then in a recent preprint,

A stable trace formula I. General expansions.

There is very little difference between the two versions. The difficulty is that so long as the local correspondence has not been proved, or rather so long as the local L-factors $L(s,\pi,r)$ are not defined in general, there is no completely unambiguous way to normalize the intertwining operators. In principle the normalization provides the appropriate definition of what are called the *weighted characters*, so that some substitute has to be found. The second substitute has the advantage for purposes of comparison of various trace formulas for different groups that it is defined in terms of the Plancherel measure, and therefore unambiguous, and not in terms of only approximately correct normalizing factors.

Analytic questions aside, to render the trace formula stable the proper formal viewpoint is necessary. First of all, it is necessary to use all Levi factors, or at least a complete set of representatives L, including of course G itself. We indicate that an object is attached to L by adding a supserscript, thus $J_{\mathfrak{o}}^L$ or J_χ^L. Both these objects satisfy the same formal relations.

If Q is a parabolic subgroup whose Levi factor M_Q lies between M_0 and L, then for $f \in \mathcal{C}_c^\infty(L(\mathbb{A})^1)$ and for $y \in L(\mathbb{A})^1$ we define

$$f_{Q,y} = \delta_Q(m)^{1/2} \int_{K\cap L(\mathbb{A})} \int_{N_Q} (\mathbb{A}) f(k^{-1}mnk) \cdot u'_Q(k,y)\, dn dk,$$

where u'_Q is a combinatorially defined function. Neither u'_Q nor the form $f_{Q,y}$ are of much importance to us here. The point is that both $J_{\mathfrak{o}}$ and J_χ satisfy the same formal relations and that these can be expressed in terms of $f_{Q,y}$.

These relations are the following:

$$J_{\mathfrak{o}}^L(f^y) = \sum_Q |\Omega^{M_Q}||\Omega^L|^{-1} J_{\mathfrak{o}}^{M_Q}(f_{Q,y}) \tag{1}$$

and

$$J_\chi^L(f^y) = \sum_Q |\Omega^{M_Q}||\Omega^L|^{-1} J_\chi^{M_Q}(f_{Q,y}) \tag{2}$$

In both formulas the sum over Q runs over the rational parabolic subgroups between M_0 and L. The formulas are formally identical. In later papers, Arthur replaces the symbol Ω with W or W_0. In order to make comparison easier, I shall make the same change when I come to the later material.

It is verified in (10) that there is a uniform way to create invariant distributions from distributions satisfying (1) or (2). What it takes are functionals

$$\phi_M^L : U(L) \to V(M)$$

such that

$$\phi_M^L(f^y) = \sum_Q \phi_M^{M_Q}(f_{Q,y}).$$

What $U(M)$ and $V(M)$ are is a technical matter. Basically, $U(M)$ is a space of smooth functions on a product over a finite number of local places,

$$\prod_{v\in S} M(F_v), \tag{3}$$

while $V(M)$ is a space of functions on the irreducible tempered representations of the group in (3). There are technical constraints to be imposed, basically that any invariant functional I on $U(M)$ is obtained by pulling back a uniquely determined invariant functional $\hat{I}$ on $V(M)$. This is not the place to go into detail.

These functions are introduced by Arthur as multi-dimensional combinatorial generalizations of logarithmic derivatives, by means of what he calls (G, M)-families. This notion is basic to the combinatorial aspects of Arthur's treatment and to understand his papers, it is essential to have mastered it. It is not difficult or deep, but to acquire some ease with it takes time and a study of examples, such as those found in (10). A (G, M)-family is a collection of functions c_P of a variable λ that runs over a vector space $i\mathfrak{a}_M$, the parabolic group running over those whose Levi factor is M. This family is called a (G, M)-family if the functions satisfy the constraint

$$c_P(\lambda) = c_{P'}(\lambda) \tag{4}$$

on the hyperplane defined by an adjacent pair (P, P'). From such a family, we can build a well-defined smooth function c_M on $i\mathfrak{a}_M$ as

$$c_M(\lambda) = \sum_P c_P(\lambda)\theta_P(\lambda)^{-1}, \tag{5}$$

where θ_P is a product of linear functions that I do not describe further.

The pertinent (G, M)-family for our present purposes is of the form

$$\phi_P(\lambda, f, \pi, P_0) = \mathrm{tr}(I_{P_0}(\pi, f)R_P(\lambda, \pi, P_0)). \tag{6}$$

The group P_0, of which this expression turns out to be independent is a parabolic group with Levi factor M. The function f can be thought of as lying in $U(G)$ so that this expression defines a function of π, thus an element in $V(M)$. Thus the (G, M) family will depend on two additional parameters f and π. Moreover $I_{P_0}(\pi, \cdot)$ is the representation induced from π, which we may integrate against f and $R_P(\lambda, \pi, P_0)$ is, at least in the paper (10), a normalized intertwining operator. The function of the normalization is basically to remove a noninvariant part that depends, in essence, only on finitely many places from the trace formula and to leave the automorphic L-functions, which are, on the one hand, infinite products and, on the other, invariant.

Although the expression (6) was introduced for G, it could be defined for any Levi factor L between M and G, and is then denoted

$$\Phi_P^L(\lambda, f, \pi, P_0),$$

the function f now lying in $U(L)$. Arthur then sets, using (5), $\phi_M^L(f)$ at π equal to the value of

$$\sum_P \Phi_P^L(\lambda, f, \pi, P_0)\theta_P(\lambda)^{-1}$$

at $\lambda = 0$

Arthur then proves that for any functional J satisfying (1) or (2), there is an invariant functional I such that

$$J^L(f) = \sum_M c(M)c(L)^{-1}\hat{I}^M(\phi_M^L(f),$$

with $c(M) = |\Omega^M|$. The definition proceeds inductively.

$$I^L(f) = J^L(f) - \sum_{M \neq L} c(M)c(L)^{-1}\hat{I}^M(\phi_M^L(f)).$$

The invariant trace formula then takes the form,

$$\sum_{\mathfrak{o}\in\mathcal{O}} I^L_{\mathfrak{o}}(f) = \sum_{\chi\in\mathfrak{X}} I^L_{\chi}(f).$$

When $L = G$, it is suppressed from the notation and this expression becomes simply

$$\sum_{\mathfrak{o}\in\mathcal{O}} I_{\mathfrak{o}}(f) = \sum_{\chi\in\mathfrak{X}} I_{\chi}(f).$$

So we have moved back from J to I. Before leaping ahead to S (for stabilization), we consider a basic application of the general trace formula.

An Application

In most applications that have been made, or even contemplated, until now of the trace formula, there is a comparison to be made, and this comparison is impossible without stabilization. The only – rather almost the only – groups for which stabilization is not necessary are the inner forms of $GL(n)$, so that no stabilization is necessary either in the comparison of automorphic forms on a quaternion algebra with automorphic forms on $GL(2)$ or for base change, nor is any needed for the simplest Shimura varieties, the familiar ones associated to $GL(2)$. Otherwise stabilization is necessary. The consequences for a group as simple as $U(3)$ are discussed in the following two books.

B) J. Rogawski, *Automorphic representations of unitary groups in three variables*, PUP

C) R. Langlands and D. Ramakrishnan, eds. *The zeta functions of Picard modular surfaces*, Les Publ. CRM

The attention shown to these volumes does not suggest any widespread interest in stabilization or its applications to the study of Shimura varieties. Indeed, the mere process of transforming the trace formula to an invariant form is arduous enough that no-one is likely to undertake it without some clearly defined purpose. I give an application in which the invariant trace formula does not appear explicitly because a large number of terms vanish, but not the term most important for the purpose, the one contributed by the identity, and in which it is also possible, but not easy, to circumvent the problems caused by instability.

The Tamagawa Number

The theory of quadratic forms was developed in the nineteenth century by Eisenstein, H. J. S. Smith and Minkowski, who found in particular famous formulas for the number of weighted classes in a genus and for the number of weighted representations of a number by a genus. This work was then extended by Siegel, in whose hands it led to the modern theory of automorphic forms on general reductive groups. The original formula was perceived by Tamagawa to be an assertion about the volume of $G(F)\backslash G(\mathbb{A})$ for an orthogonal group, and Weil conjectured in general that this volume was 1 if G is semisimple and simply connected, thus that the Tamagawa number of a semisimple, simply connected group is 1. Elegant as the general statement is, the original theorems of Eisenstein, Smith and Minkowski remain its most appealing manifestations. It is none the less important to appreciate that the statement that the Tamagawa number is 1 belongs to the very elements of the analytic theory of automorphic forms, thus to the theory that arose by combining the generality created by Siegel with the more analytic ideas of Hecke, Maaß and Selberg. Siegel was captivated by complex analysis in both one and several variables, but, in his published work at least, largely indifferent to real analysis. The general statement is, moreover, still inaccessible by any other means than those provided by the basic real-analytic ideas.

It is proved in two steps. For quasi-split groups, it is an easy consequence of the elements of the theory of Eisenstein series. Every other group G is an inner twist of a quasi-split group G' and for them, it is proved by comparing the trace formula for G and G'. A baby trace formula will not do the trick. There is some harmonic analysis required and harmonic analysis requires functions that do not vanish at the identity. There is also some arithmetic, for example, the Hasse principle, but that is not our concern here.

The second step was carried out by Kottwitz, who exploited Arthur's trace formula.

D) R. Kottwitz, *Tamagawa numbers*, Ann. Math. (1988)

Kottwitz compares the trace formula on G for a function f with that on G' for a function f'. These functions are chosen so that they are products $f = \prod f_v$ and $f' = \prod f'_v$. At all but a finite number of places,, say those in a set S, G_v and G'_v will be the same group. We choose at one of these places $f_v = f'_v$ to be the matrix coefficient of a supercuspidal representation. This reduces the spectral side of the trace formula for both G and G' to a sum over cuspidal automorphic forms. At all the places

$w \in S$ he arranges, and this requires some effort, that the invariant orbital integrals of f_w and f'_w be comparable, in effect that

$$\int_{G_\gamma(F_w)\backslash G(F_w)} f(g^{-1}\gamma g)dg = 1 \tag{7}$$

and

$$\int_{G'_\gamma(F_w)\backslash G'(F_w)} f'(g^{-1}\gamma g)dg = 1. \tag{8}$$

These are approximate statements as attention has to be made to normalization.

Stabilization is necessary because there is no way to compare conjugacy classes in $G(F)$ and $G(F')$. Only stable conjugacy classes can be compared, thus, in essence, only conjugacy over the algebraic closure of F, because it is only over this closure that the two groups are isomorphic. The advantage of (7) and (8) is that because the orbital integrals are so simple this does not matter.

More on Invariance

If we had looked more closely at the arguments of Kottwitz, we would have seen that we had not carried the discussion of either the geometric terms $J_{\mathfrak{o}}$ or $I_{\mathfrak{o}}$ or of the spectral terms J_χ or I_χ far enough. Indeed, we did not describe them at all. The principal step is perhaps to obtain an adequate representation of the terms $J_{\mathfrak{o}}$ and J_χ appearing in the trace formula, thus in the terminology of the Morning Seminar of 1983/84 to pass from the coarse expansions to the fine. Once the fine expansions of $J_{\mathfrak{o}}$ and of J_χ are available and it is certain that they satisfy relations similar to (1) and (2), the formal principles already described – which cannot be established by formal arguments alone – can be exploited to introduce the fine expansions of $I_{\mathfrak{o}}$ and I_χ.

The coarse geometric terms $J_{\mathfrak{o}}(f)$ have an expansion

$$J_{\mathfrak{o}}(f) = \sum_M \sum_\gamma |W_0^M||W_0^G|^{-1} a^M(S,\gamma) J_M(\gamma, f),$$

in which M runs over all the Levi factors of G containing the given minimal Levi factor M_0. There are two extreme cases. In the first, $\mathfrak{o}$ consists entirely of semisimple elements and there is in $\mathfrak{o}$ an element γ contained in a Levi factor M_γ which is such that the centralizer of γ in G also lies in M_γ. Then

$$J_{\mathfrak{o}} = |D(\gamma)|^{1/2} \int_{G_\gamma(F_S)\backslash G(F_S)} f(x^{-1}\gamma x) v_M(x) dx,$$

is a weighted orbital integral, the weight $v_M(x)$ being defined combinatorially in terms of the roots of the group M. In any application, for example to the study of the Tamagawa number, the combinatorial properties of the weights, especially those related to (G, M)-pairs, are important. In the second extreme case $\mathfrak{o}$ consists of unipotent elements. This case is treated in a separate paper.

20) *A measure on the unipotent variety.*,Can. J. Math (1985)

The general case is a combination of the two. The set of places S has to be chosen sufficiently large, a notion that depends on γ and f and F_S is the sum

$$\oplus_{v\in S} F_v$$

The first fine expansion of the spectral side appears in the paper

14) *On a family of distributions obtained from Eisenstein series, II*, Amer. Jour. Math. (1982)

It is a consequence of a multi-dimensional version of familiar properties of the kernel $\sin(Tx)/T$ that deserves to be better known. The formula for $J_\chi(f)$ that results is rather long. It is a sum over M containing M_0, over L containing M, over cuspical representations of $M(\mathbb{A})$, taken in the usual way modulo twisting with families of characters, and over subsets $W^L(\mathfrak{a}_M)_{\mathrm{reg}}$ of the Weyl group of the product of

$$|W_0^M||W_0|^{-1}|\det(s-1)_{\mathfrak{a}_M^L}|^{-1},$$

a factor to which we need not pay too much attention, and of

$$\int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} |\mathcal{P}(\mathcal{M})|^{-\infty} \sum_{\mathcal{P}} \mathrm{tr}(\mathfrak{M}_{\mathcal{L}}(\mathcal{P},\lambda)\mathcal{M}(\mathcal{P},\int)\rho_{\chi,\pi}(\mathcal{P},\lambda,\{\})\lceil\lambda. \tag{9}$$

This expression is not agreeable, one one could wish for a more compact notation. The theory of Eisenstein series decomposes $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}))$ into a direct integral of induced representations, the subspaces being parametrized first by χ and then, for each M, a class of π, basically all $\pi\otimes\omega_\lambda$. The induced representation, with the necessary multiplicity, attached to χ and $\pi\otimes\omega_\lambda$ is $\rho_{\chi,\pi}(P,\lambda)$, the subgroup P being as usual any parabolic with Levi factor M. All these representations are equivalent and there are intertwining operators between them. The operators

$$\mathfrak{M}_L(P,\lambda), \tag{10}$$

defined generally by a certain family of (G, M)-pairs, thus by generalized logarithmic derivatives of the intertwining operators are – because of the differentiation – not invariant, reflecting the lack of invariance of J_χ. The difficulty with (9) is that the operators (10) contain two quite different difficulties, those arising from the lack of invariance of the logarithmic deriviatives and those arising from the automorphic L-functions that appear in the global intertwining operators. The next step, undertaken first in two papers,

28) *Intertwining operators and residues I. Weighted characters*, J. Funct. Anal. (1989)

27) *The invariant trace formula. II. Global theory*, J. Amer. Math. Soc. (1988)

and then later, in a form that for technical reasons is slightly modified, in the papers on stabilization, is to separate these two difficulties.

The convergence of the spectral side is, as one knows, someone delicate. There are still problems to be resolved along the lines of those treated in

W. Müller, *The trace-class conjecture in the theory of automorphic forms*, Ann. of Math. (1989)

So Arthur introduces a decomposition of the collection of parameters according to the size t of what I would think of as the real part of their infinitesimal character but which appear in his papers as the complex part and sets

$$J_t(f) = \sum_{\Im(\nu_\chi)=t} J_\chi(f).$$

Then

$$J_t(f) = \sum_M |W_0^M||W_0^G|^{-1} \sum_{M_1} \int_{\Pi_{M_1}(M,t)} a^M(\pi) J_M(\pi, f) d\pi.$$

Here M_1 runs over Levi factors containing M. The term $a^M(\pi)$ contains the logarithmic derivatives of automorphic L-functions and $J_M(\pi, f)$ contains the logarithmic derivatives of *normalized* local intertwining operators. Thus it contains only local information. On the other hand $a^M(\pi)$ is scalar!

What one sees from this sketch is that the trace formula is technically elaborate. It is highly unlikely that many mathematicians will undertake to master it until more results for transfer or for the zeta-functions of Shimura varieties are available, thus results that they can use in a context more familiar to them. In my view, there is something to be said for careful, complete treatments of low-dimensional examples beyond $U(3)$.

Stabilization

Since the bulk of the audience will be totally unfamiliar with stabilization, I shall not say much about it. Suppose, just to consider a simple example, we wish to compare automorphic forms not on $GL(2)$ and on the multiplicative group of a quaternionic division algebra but on $SL(2)$ and the group of elements of norm one in such an algebra.

At a local level, say over $\mathbb{R}$, this could mean comparing representations of $SL(2)$ and $SU(2)$, thus, as it turns out, even comparing their characters; but to compare characters, which are functions, we have to compare the sets on which they are defined, thus we have to pair conjugacy classes in the two groups. A conjugacy class $\Gamma'(\theta)$ in $SU(2)$ is determined by the eigenvalues $e^{i\theta}$ and $e^{-i\theta}$, given up to order, of its members. On the other hand a conjugacy class $\Gamma(\theta)$ in $SL(2)$ is determined by the rotation it contains, which will be through an angle θ, $0 \leq \theta < 2\pi$. Although the rotations through an angle θ and through $2\pi - \theta$ have the same eigenvalues, $e^{i\theta}$ and $e^{-i\theta}$, and are conjugate inside $GL(2)$, they are not conjugate inside $SL(2)$. The class $\Gamma'(\theta)$ has to be paired with the union of $\Gamma(\theta)$ and $\Gamma(-\theta)$. This is the instability associated to $SL(2)$. There is a similar instability, much more complex, inherent in comparisons of higher dimensional groups.

The inability to compare conjugacy classes as such means that we are also not able to compare characters, at least not without a small modification. As is well known, the value of a character of a representation of $SU(2)$ at an element of $\Gamma'(\theta)$ is equal to -1 times the sum of a holomorphic discrete-series character and of the corresponding anti-holomorphic discrete-series character at $\Gamma(\theta)$. Of course, *holomorphic* and *anti-holomorphic* are not absolute notions. They depend on the choice of an orientation of the plane.

We could also introduce a *stable* character, the sum of the holomorphic and of the anti-holomorphic character. This leads to the notion of stable harmonic analysis, in which one has stable characters, not true characters but sums of associated true characters. Stable harmonic analysis is best for purposes of comparison but incomplete and must be accompanied by *endoscopy*, in the case at hand the study of the difference of a holomorphic and anti-holomorphic character. These can be expressed in terms of ordinary, thus abelian, characters of the group of rotations in $SL(2)$. This group, a reductive group in its own right, is therefore referred to as an endoscopic group. Thus, for many purposes local harmonic analysis can be reduced to the study of stable harmonic analysis and of endoscopy. An ordinary character is the sum of a stable character and an endoscopic transfer from the character an endoscopic group of smaller dimension, in the present case from a Cartan subgroup.

This reduction, which may seem artificial locally, is absolutely necessary globally. The local problems, which appear not only at infinity but at the finite places as well, are exacerbated globally. On the one hand, the trace formula appears as a sum over conjugacy classes, not over stable conjugacy classes, and only stable conjugacy classes can be compared. On the other hand, harmonic analysis functions at the level of conjugacy classes in $G(\mathbb{A})$, not at the level of conjugacy classes in $G(\mathbb{Q})$, and it is by no means the case that every class in $G(\mathbb{A})$ stably conjugate to a class in $G(\mathbb{Q})$ is realizable in $G(\mathbb{Q})$. The upshot is that the absolute trace formula has to be represented as the sum of a stable trace formula and an infinite number of endoscopic transfers of trace formulas on lower-dimensional groups, the endoscopic groups, in the case at hand again all abelian Cartan subgroups but now those defined over the global field $\mathbb{Q}$.

A similar program has to be undertaken for arbitrary groups, but now the lower dimensional endoscopic groups are themselves nonabelian, so that their trace formulas, or in the local context their characters, have to be stabilized before being transferred. This is easier said than done. Ths situation at infinity, at least for tempered characters, has been well understood for some time, thanks to Shelstad. The stabilization of the terms in the trace formula associated to elliptic semisimple elements is elaborate but fairly well understood, except that the fundamental lemma which is the keystone of the undertaking remains unproved in general.

One of Arthur's tasks if he is to arrive not at a formula simplified for special purposes but one that can be applied to establish the transfer from the classical groups to $GL(n)$ already described is to stabilize the remaining terms on the geometric side, those associated to orbits that are not semisimple, and, I suppose, to sort out the consequences for the spectral side. He has begun this for the absolute trace formula although, as I recall, he will ultimately need it for the twisted formula as well.

In addition to a preprint already mentioned, there are at least three more.

A stable trace formula II. Global descent.

Stabilization of a family of differential equations.

Endoscopic L-functions and a combinatorial identity.

I do not try to present them here, as my goal is rather to persuade the younger, more vigorous members of the audience, that the path to a successful treatment of the number-theoretical problems to whose solution functoriality promises to contribute may very well lie through the trace formula and that they had best be prepared to master it; it is not to discuss in any serious way the arguments that enter into the complete treatment of the formula. That would be a task not for a single lecture but for a year's seminar.

Comparisons

The general principle governing the use of the trace formula is the term-by-term comparison of two geometric sides, whether these are the geometric sides of two trace formulas or, as with Shimura varieties, the geometric side of a trace formula and of some Lefschetz formula. Certainly, experience suggests that to arrive at some notion of which comparisons will be possible and useful, the full trace formulas is not needed, only some inspiration, for which the geometric terms associated to regular elliptic elements provide the clue; the rest is hard work – usually on the part of a large number of people over several decades – accompanied by further inspiration. Sustained reflection on the equalities, beyond those for the simpler geometric and spectral terms, required for the successful functioning of the comparison demands an understanding of the trace formula in all its complexity and leads to a different kind of insight. Several of the most striking suggestions to arise from such reflections are the conjectures of Arthur on nontempered unitary representations of groups over local fields and on the nontempered spectrum in the space of automorphic forms. The local conjectures are, as is well-known, already difficult enough; the global conjectures are, no doubt, the correct general form of the conjecture of Ramanujan-Petersson.

A Pipe Dream

Rather than consider, as suggested at the beginning of the lecture, the order of the pole of $L(s,\pi,r)$ at $s=1$, it is more convenient to consider the residue of

$$-\frac{L'}{L}(s,\pi,r). \tag{11}$$

It is an integer $m(\pi,r)$ that we can expect to be equal to the multiplicity of the trivial representation in the restriction of r to H_π. In general if J is a subgroup of ${}^LG$, let $m(J,r)$ denote the multiplicity of the trivial representation in the restriction of r to J. Although it may be too much to expect that the linear form

$$r \longrightarrow m(J,r)$$

uniquely determines the conjugacy class of J, it should come close to it, and that is enough for our purposes. Then the linear form

$$r \longrightarrow m(\pi, r)$$

will determine all π with H_π in some small, perhaps even finite, collection. So our goal is to try to isolate those π associated to a given linear form. That should mean that we can find a formal finite linear combination

$$\oplus a_i r_i$$

such that

$$-\sum_i a_i \frac{L'}{L}(s, \pi, r_i)$$

is nonzero or at least equal to a given integral value if and only if H_π lies in some small collection of conjugacy classes containing a prescribed one.

So we concentrate on the analysis of (11). In fact the parameter s is factitious. It will be eliminated, but first we observe that rather than using the full L-function we could use a partial L function, the product over the primes outside of some finite set S that includes of course all the infinite places. This should not change the residue for the representations of interest. At the places in S, we introduce a function ϕ_v, smooth with compact support, so that the traces $\operatorname{tr}\pi_v(\phi_v)$ are defined. Then the residue of

$$-\prod_S \operatorname{tr}\pi_v(\phi_v)\frac{L'}{L}(s, \pi, r) \tag{12}$$

is

$$m(\pi, r)\prod_S \operatorname{tr}\pi_v(\phi_v).$$

The expression (12) is

$$\prod_S \operatorname{tr}\pi_v(\phi_v)\sum_{\mathfrak{p}}\sum_{j=1}^{\infty}\frac{1}{j}\operatorname{tr}(r(A_\pi(\mathfrak{p})^j))\mathrm{N}\,\mathfrak{p}^{js},$$

if $A_\pi(\mathfrak{p})$ is the conjugacy class in ${}^LG$ associated to π and $\mathfrak{p}$ in the usual way (sometimes known as the Hecke or Satake parameter). Ignoring for the moment exceptional π, thus those that do not satisfy the general Ramanujan-Petersson conjecture, we can expect that

$$m(\pi, r) = \lim_{N\to\infty}\frac{\sum_{\mathrm{N}\,\mathfrak{p}<N}\operatorname{tr}(r(A_\pi(\mathfrak{p})))}{\sum_{\mathrm{N}\,\mathfrak{p}<N} 1}$$

Thus what appears to be pertinent are the individual terms

$$\prod_S \operatorname{tr} \pi_v(\phi_v) \cdot \operatorname{tr}(r(A_\pi(\mathfrak{p}))) \tag{13}$$

and their averages over $\mathfrak{p}$. It is best, however, to acknowledge that it appears at first that at some point we are going to have to deal with the Ramanujan-Petersson conjecture, and then it would be necessary to examine for individual primes $\mathfrak{p}$ the averages

$$\frac{1}{N} \sum_{j=1}^{N} \operatorname{tr}(r(A_\pi(\mathfrak{p})^j)).$$

They look more difficult as they have no predictable value. It may be, and this is just as well, that they are irrelevant. The reason will appear toward the end of the lecture.

There is a spherical function $\phi_\mathfrak{p}^r$ such that

$$\operatorname{tr}(r(A_\pi(\mathfrak{p}))) = \operatorname{tr} \pi_\mathfrak{p}(\phi_\mathfrak{p}^r).$$

Let $\phi_\mathfrak{q}$ for $\mathfrak{q} \neq \mathfrak{p}$, q not in S, be the unit element in the Hecke algebra at $\mathfrak{q}$ and set

$$\Phi_\mathfrak{p}^r(g) = \phi_\mathfrak{p}^r(g_\mathfrak{p}) \prod_{v \in S} \phi_v(g_v) \prod \phi_\mathfrak{q}(g_\mathfrak{q}),$$

the final product being taken over $\mathfrak{q}$ not in S and different from $\mathfrak{p}$. Then $\operatorname{tr} \pi(f)$ is equal to (13) if π is unramified outside of S and to 0 otherwise.

The conclusion is, in essence, that averaging

$$\sum_i a_i \operatorname{tr} \pi(\Phi_\mathfrak{p}^{r_i})$$

will pick out exactly the π of interest. (If the form $J \to \sum a_i m(J, r_i)$ takes several integral values, supplementary induction processes will certainly be necessary.) Thus we are reduced to considering the trace of $\Phi_\mathfrak{p}^r$ on the space of automorphic forms. So we are at first reduced to the trace formula, but now to be accompanied by a supplementary and clearly very difficult averaging. The result will still have to be compared with another expression of which we have as yet no very clear notion. To continue, it is best to look at a particular case.

Icosahedral Representations

Take G to be $PGL(2)$ so that ${}^LG$ is $SL(2)$ and suppose we are looking for π such that H_π is icosahedral. The pertinent J are abelian, dihedral, tetrahedral, octrahedral and icosahedral. We can assume that we understand well those π such that H_π is abelian or dihedral and that we are in a position to eliminate their contributions from any formula that appear. This leaves the other three. If we examine $m(\pi, r)$ for r the k-th symmetric power, $k = 1, 2, 3 \ldots, 30$, we find the following values.

$$\begin{array}{ll} \text{tetrahedral}: & 0,0,0,0,0,1,0,1,0,0,0,2,0,1,0,1,0,2,0,2,0,1,0,3,0,2,0,2,0,3 \\ \text{octahedral}: & 0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,1,0,1,0,1,0,0,0,2,0,1,0,1,0,1 \\ \text{icosahedral}: & 0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,1 \end{array}$$

Thus we first examine the sixth symmetric power to get at the tetrahedral group, then the eighth to get at the icosahedral group, and finally the twelfth to get at the icosahedral group.

Nothing will be easy, but we first ask ourselves what the result obtained by averaging is to be compared with. We need some construction of representations with an icosahedral Galois group, a construction that is presumably similar to the construction of Kummer extensions. As with these extensions, we are permitted to pass to a solvable extension of the ground field because base change is available – in principle for any group! At least it is not unreasonable to suppose that once we have dealt with matters in the larger field a descent will be possible. Consider an arbitrary finite group $\mathfrak{G}$ as well as an irreducible d-dimensional representation ρ of it. Kummer extensions arise when $\mathfrak{G}$ is abelian. Since we are permitting abelian base change, we can assume that the representation ρ is defined over F. This only requires a cyclotomic extension.

Suppose K is a Galois representation of F with group $\mathfrak{G}$. It contains a normal base. Since the representation r occurs d times in the regular representation, there are elements $x_{i,j}$, $1 \le i, j \le d$ in K such that

$$X_j^\sigma = r(\sigma) X_j$$

if X_j is the column vector $(x_{i,j})$. We can suppose that the matrix

$$\mathfrak{X} = (X_1, X_2, \ldots, X_d)$$

is nonsingular. Another choice of the X_j leads to a matrix $\mathfrak{X}A$, where A has coefficients in F.

Suppose

$$a = (a_1, a_2, \ldots, a_d), \quad b = (b_1, b_2, \ldots, b_d)^{\mathrm{tr}}$$

are vectors with coefficients in F, the first a row vector, the second a column vector. Then any symmetric function $E(\{a\mathfrak{X}^\sigma b\})$ of $\{a\mathfrak{X}^\sigma b\}$ is the same function of $\{ar(\sigma)\mathfrak{X}b\}$,

$$E(\{ar(\sigma)\mathfrak{X}b\}). \tag{14}$$

This function can be expanded as

$$\sum a^I b^J F_{I,J},$$

where I and J are vectors with positive integral values and the coefficients $F_{I,J}$ are in F and satisfy certain algebraic relations,

$$P(F_{I,J}) = 0. \tag{15}$$

They are also functions $E_{I,J}(x_{i,j})$ of the $x_{i,j}$.

It seems to me likely that, except for some degenerate cases, given a solution of (15) in the field F, we can solve

$$E_{I,J}(x_{i,j}) = F_{I,J}, \tag{16}$$

and that then

$$\mathfrak{X}^\sigma = r(\sigma)\mathfrak{X},$$

where on the left σ is an element of the Galois group of the extension $F(\{x_{i,j}\})$ of F and on the right an element of the group $\mathfrak{G}$.

If B is a matrix with coefficients in F and if in (14) the matrix $\mathfrak{X}$ is replaced by $\mathfrak{X}B^{-1}$ and b by Bb the result is not changed. Thus the group of matrices with coefficients in F acts on the solutions of (15) and (16). The action on solutions of (15) is clear; the action on solutions of (16) less so. If $d = 1$, the action is familiar. In the equation $\mu^n = \nu$, the left side is replaced by $\beta\mu$ and the right side by $\beta^n\nu$.

What has to be done, just as when studying Kummer extensions, is to examine the orbits under $GL(d, F)$ of the solutions of (15) and (16) that are unramified outside of S. It is by no means clear to me at the moment how this can be done in a useful way! I would be much easier about this lecture if I had at least an inkling of an idea.

Modifying the Trace Formula

We are going to consider only the problem (T2), so that r will be the m-th symmetric power. Consider first, not the result of applying the trace formula, but the trace itself. It is

$$\sum_\pi \mathrm{tr}(\pi(\Phi_{\mathfrak{p}}^r)). \tag{17}$$

This trace is the sum of two parts, those terms corresponding to one-dimensional representation and those terms corresponding to cuspidal representations. They are quite different. The first dwarfs the second and has to be removed, not approximately but exactly. This is easy to do in (17), but not immediately obvious on the geometric side of the trace formula.

If we are to retain the freedom to make an abelian or solvable base change, we have pretty much to work with an arbitrary number field, but as I have only made calculations for $\mathbb{Q}$, I prefer to confine myself in this part of the discussion to that field. Rather than $\phi^r_{\mathfrak{p}}$ or $\Phi^r_{\mathfrak{p}}$, I write ϕ^r_p and Φ^r_p. Then, barring any errors on my part, the function $\phi^r_{\mathfrak{p}}$ is the projection modulo the center of

$$\frac{T^m_p}{p^{m/2}}, \tag{18}$$

if T^m_p is the characteristic function of the integral matrices whose determinant is a unit times p^m. The sum over the one-dimensional representations in (17) leads to

$$\sum_{\chi}\prod_{v}\chi(\phi_v)\chi^m(\varpi_{\mathfrak{p}})p^{m/2}\frac{1-1/p^m}{1-1/p}, \tag{19}$$

the sum being over all one-dimensional representations that are unramified outside of S and of order two. (We are dealing with $PGL(2)$!) The remainder looks just like (17) itself, except that the sum is confined to cuspidal representations unramified outside of S, thus

$$\sum_{\pi}\prod_{v\in S}\operatorname{tr}(\pi_v(\phi_v))\operatorname{tr}(r(A_p(\pi))). \tag{20}$$

For a given collection of ϕ_v, $v\in S$, this expression is thought to be $O(1)$ as p varies. On the other hand, (19) is of the order of $p^{m/2}$, which is very large if, for example, $m=12$.

It is for this reason that we have to see (19) reproduced exactly on the geometric side. It is the average over p of the expression (20), perhaps modified to take into account unwanted abelian, dihedral, tetrahedral and octahedral representations, that is to be compared with whatever results from the concrete study of icosahedral representations, and we want to be able to do this numerically if nothing else. If we cannot do it numerically, we are unlikely to be able to do it theoretically, and the presence of a term of order p^6, corresponding to $m=12$ would, as Erez Lapid pointed out to me, destroy all possibility of this.

I have only performed the calculation that removes (19) approximately. There are many details to work out, but the result so far is persuasive enough and I explain it, but in broad outline only. What occurs in (19) are the integrals

$$\int_{G(\mathbb{Q}_v)}\phi_v(g)dg,$$

and we want to see how they arise from the geometric terms of the trace formula. I have contented myself with seeing this for $v = \infty$, the calculations at the other places being, I suppose, somewhat different. (I am not very far along and want to understand the critical issues before spending too much time on technical matters.)

First of all, I use from the geometric side of the trace formula only the sum over the regular elliptic elements, assuming that the others are few in number and will take care of themselves. We obtain then a double sum, the outer sum being over quadratic extensions of $\mathbb{Q}$, or better over fundamental discriminants, and the inner sum over integral γ in the corresponding field whose norm is $\pm Ap^m$, A being one of a finite set of integers that depends upon the ϕ_v, v a finite place in S. Since the all the functions ϕ_v including ϕ_∞ have compact support, the total sum will be finite. It is multiplied by a factor $1/p^{m/2}$ arising from (18) that is clearly important for determining the size of any part of the expression. The inner sum is provided with a factor $\epsilon(d)$ that is given as twice the discrimant times the regulator divided by the number of units. As Lapid also pointed out, the orbital integrals occurring permit a rewriting of the sum as a sum over all discriminants, with a similar factor $\epsilon(d_f)$ multiplying the inner factor, and with γ now constrained to lie in the corresponding order. It is in fact somewhat more complicated than this because the choice of the ϕ_v at the finite places in S amount, for our purposes, to congruence conditions at the places in S, conditions that are also the source of the factor A. So the fundamental discriminants can not simply be multiplied by all squares. Some additional attention has to be paid to these places.

I usually write the fundamental discriminant as d, the square as f^2, and the discriminant f^2d as d_f. The factor $\epsilon(d_f)$ can be replaced by

$$\sqrt{|d_f|}\sum_{n\geq 1}\left(\frac{d_f}{n}\right)\frac{1}{n},$$

This sum is of course only conditionally convergent, but the original sum being finite, we can interchange it with the sum over n. For a given n, we are to divide by $np^{m/2}$.

Write $Ap^m = N/4$ and set $\gamma = a+b\beta$, where β is defined in terms of the fundamental discriminant as $\sqrt{d}/2$ when $d \equiv 0 \pmod 4$ or $(1+\sqrt{d})/2$ if $d \equiv 1 \pmod 4$. If $\gamma = a + b\sqrt{\beta}$, then

$$N/4 = \mathrm{N}\,\gamma = \begin{cases} a^2 - b^2d/4, & d \equiv 0 \pmod 4 \\ (a+b/2)^2 - b^2d/4, & d \equiv 1 \pmod 4 \end{cases}$$

In the first of these two cases, I take $r = 2a$ and in the second I take $r = 2a+b$. Thus, in the first case r runs over the even integers, and in the second over all integers. In both cases

$$\frac{\mathrm{tr}(\gamma)}{2|\,\mathrm{N}(\gamma)|^{1/2}} = \frac{r}{N^{1/2}}.$$

Moreover

$$N - r^2 = -b^2 d = -b_f^2 d_f, \quad d_f = f^2 d.$$

Thus we are going to sum over the indicated r and then over the integers b_f such that b_f^2 divides $r^2 - N$ and yields a result congruent to 0 or 1 modulo 4. (There are further constraints for the prime divisors of b_f in S, but they are not important at the present level of discourse.) The map $\mathrm{Ch}: \gamma \to (\mathrm{N}\,\gamma, \mathrm{tr}\,\gamma)$ is a map from regular semisimple conjugacy classes to the plane with the line $x = 0$ and the parabola $4x - y^2$ removed. Moreover, as our functions are homogeneous, the orbital integrals are functions of $y/\sqrt{|x|}$ and of the sign of x. The orbital integral at the real place can be represented as

$$\int \phi_\infty(g^{-1}\gamma g)dg = \psi'_\infty(\mathrm{Ch}(\gamma))\Gamma_\infty(\gamma) + \psi''_\infty(\mathrm{Ch}(\gamma))\frac{|\mathrm{N}\,\gamma|}{|\gamma_1 - \gamma_2|}\Delta_\infty(\gamma), \tag{21}$$

where Δ_∞, introduced for formal purposes is identically 1, but Γ_∞ is the characteristic function of the set where $4x - y^2 \geq 0$. The eigenvalues of γ are γ_1 and γ_2. The two functions are both smooth and of compact support in the plane with the y-axis removed, but the values of the first only count inside the parabola. There are similar formulas at the finite places that are useful, but that need not be introduced here, as I have not yet treated the finite places.

Both ψ'_∞ and ψ''_∞ are homogeneous functions and may be written as

$$\psi'(\mathrm{tr}\,\gamma/2\,\mathrm{N}\,\gamma), \quad \psi''(\mathrm{tr}\,\gamma/2\,\mathrm{N}\,\gamma).$$

The function ψ'' is smooth with compact support and ψ' is smooth inside and up to the boundary of $-1 \leq t \leq 1$ but 0 outside this set.

The sum that appears for a given n will therefore be the sum of two expressions:

$$\frac{1}{2np^{m/2}} \sum_{|r|\leq\sqrt{N}} \sum_{s^2|r^2-N} \psi'\left(\frac{r}{\sqrt{N}}\right)\frac{\sqrt{N-r^2}}{s}\left(\frac{(N-r^2)/s^2t^2}{n}\right) \tag{22}$$

and

$$\frac{1}{2np^{m/2}} \sum_r \sqrt{\frac{N}{4}} \sum_r \sum_{s^2|r^2-N} \psi''\left(\frac{r}{\sqrt{N}}\right)\frac{1}{s}\left(\frac{(N-r^2)/s^2t^2}{n}\right). \tag{23}$$

The t^2 has something to do with numbers t that are products of primes in S and over which there is a sum to be taken, a sum that is ignored here. The additional 2 appears in the denominator because we take s positive.

For each n and s (t being ignored), both (22) and (23) are sums over periodic sequences, that we expect to express approximately through the value of the Fourier transforms of ψ' and ψ'' at 0. This turns out to be so. Moreover when we perform the sum on n and s, we obtain (I only include that part of the result which is pertinent.)

$$\frac{p^{m/2}}{1-1/p}\{\Phi(\Gamma(r)\sqrt{1-r^2}\psi'(r)+\frac{1}{2}\Phi(\psi''(r))\}=c\frac{p^{m/2}}{1-1/p}\int\phi_\infty(g)dg. \tag{24}$$

I have only retained the part that depends on p and ϕ_∞ and have not concerned myself with the two components of $G(\mathbb{R})$. Here $\Phi(f(r))$ is the value of the Fourier transform of an arbitrary function $f(r)$ at 0 and Γ is the characteristic function of $[-1,1]$. The number c is some trivial universal constant. What is important is that the constants in (24) are exactly those that will make it equal to (19). The extra term $p^{m/2}/p^m=p^{-m/2}$ in (19) is of no importance.

What we can now expect is that what remains when this dominant part is removed is $O(1)$. We cannot, however, expect to prove it. Indeed, when we set about to do so, we find that we are led to appeal as usual to estimates for Kloosterman sums, so that we find ourselves in a dead end. But on reflection, we see that we do not need so much; we only need to control the average over primes of what is left. This may be possible, and one of the reasons for delivering this lecture is to solicit the advice of experts on exactly this point.

The sums in (22) and (23) are periodic in r of period $K=Bns^2t^2$, where B is some constant that depends on the ϕ_v, v finite in S. When Poisson summation is applied and the term at 0 that contributes to the dominant term removed, what is left is a weighted sum over values of the Fourier transforms

$$\mu'(x)=\int\exp(2\pi ixr)\sqrt{1-r^2}\psi'(r)dr$$

and

$$\mu''(x)=\int\exp(2\pi ixr)\psi''(r)dr$$

of $\sqrt{1-r^2}\,\psi'(r)$ and of $\psi''(r)$ at points $k\sqrt{N}/K$, $k\neq 0$. The factor $4A$ in $N=4Ap^m$ aside, we are basically to let N run over the m-th powers of the first M primes, M growing larger and larger, add the contributions, divide by M, and see what results. At the moment, I have no idea how such averages are to be estimated. The result will presumably be found by taking all values of k, K and N for which $k\sqrt{N}/K$ falls in some short interval $[x,x+dx]$ not containing 0, adding the contributions, and showing that the result tends to a limit $\nu(x)dx$.

The weight is, in essence and barring as usual the many possible errors and many possible missing factors,

$$\frac{1}{nsp^{m/2}K}\sum_{0\le r<K}\exp\left(\frac{2\pi irk}{K}\right)\left(\frac{r^2-N}{ns^2}\right) \tag{25}$$

Since K is essentially ns^2, the relation between k, N, n and s is

$$\frac{k\sqrt{N}}{ns^2} \sim x, \quad k \sim \frac{ns^2 x}{\sqrt{N}}$$

The expression (25) is to be summed over n and s and averaged over N. Since N runs only over m-th powers of primes, the averaging will not be easy, but one might hope that the result is the same as if it ran over all m-th powers, for that is much easier. Presumably to pass from the first to the second requires more than an application of Dirichlet's theorem on primes in arithmetic progressions, but how much more is not clear to me at present.

Just to see, how this might lead to great simplification over the usual treatments, and that, in particular, the Kloosterman sums might be replaced by explicitly calculable Gauss and Jacobi sums, I take $s = t = 1$ in (22) and (23), drop the sum over s, and average over $N = k^m$, the exponent m being fixed. For this I make use of results of Weil on Jacobi sums and on the number of points on the curves

$$y^2 = x^m + \delta, \tag{26}$$

where δ is to be taken to be $-r^2$.

E) A. Weil, *Numbers of solutions of equations in finite fields*, BAMS (1949)

F) —, *Jacobi sums as "Größencharaktere"*, TAMS (1952)

The expression

$$\frac{1}{M}\sum_{k=1}^{M}\left(\frac{k^m - r^2}{n}\right)$$

is going to be equal to the product over the prime divisors p of n of

$$\frac{1}{p}\sum_{k=1}^{p}\left(\frac{k^m - r^2}{n}\right).$$

So we consider this expression, denoting it Δ/p. It is pretty clear that

$$p + \Delta + \epsilon,$$

where ϵ is 1 if m is odd and 2 if it is even, is the number of points over $\mathbb{F}_p$ on the curve (26). Thus $\Delta + \epsilon - 1$ is the coefficient of the linear term in the numerator of the ζ-function of the curve. Weil provides a formula for this as a sum of terms

$$\chi((-1)^{b_0}\delta^{a_0+b_0})j = \chi((-1)^{a_0}r^{2(a_0+b_0)})j$$

where χ is a Größencharakter and j a Jacobi sum. Substitution in (25) is going to lead to a Gauss sum.

Whether the measure $\nu(x)dx$ is indeed absolutely continuous with respect to Lebesgue measure is perhaps a question to which one should not anticipate the response. We can certainly predict what it must be. We are going to end up with a distribution

$$\int \{\mu'(x) + \frac{1}{2}\mu''(x)\}\nu(x)dx,$$

which is an invariant distribution on $G(\mathbb{R})$. I am not sure about the factor $1/2$ and I am ignoring the finite places in S. This distribution will have two parts, one contributed by the dihedral representations unramified outside of S and not too ramified at the primes of S, thus, in particular, by dihedral representations associated to a finite number of quadratic extensions, so that there should be no particular problem analyzing the sum of their character, the other by representations π_∞ associated to two-dimensional representations of the Galois group $\mathrm{Gal}(\mathbb{C}\backslash\mathbb{R})$, whose characters are of course known. So we can anticipate the form of ν. That should be a help.

There will remain the problem of getting a handle on the two-dimensional analogues of Kummer extensions and of making the comparison with whatever that gives with whatever result the averaging process just described yields! This will presumably be a matter of finding *small* comparable objects on both sides.

It should perhaps be observed that if the limit of the average of the geometric side – with the contribution corresponding to the one-dimensional representations removed – is shown to exist, then the limit

$$\sum_\pi \prod_{v\in S} \mathrm{tr}\,\pi(\pi_v) \frac{\sum_{p<N} \mathrm{tr}(r(A_\pi(p)))}{\sum_{p<N} 1},$$

the sum being taken over cuspidal π, will also exist and be of the form

$$\sum_\pi c_\pi \prod_{v\in S} \mathrm{tr}\,\pi(\pi_v), \tag{27}$$

where c_π is a constant. It will have to be shown that these constants are positive integers or zero. The formula (27) may be regarded as the spectral side of a *trace formula* in which the geometric side is whatever is yielded by the averaging of the original geometric side as described – if it yields something. The discussion in the section *Icosahedral representations* will have to be developed so that it too yields a *geometric* expression, and the two will then have to be compared.

Final Comments

There is hardly a guarantee that these ideas will lead anywhere, but in some respects the theory of automorphic forms is at present in a parlous state, and needs a push in the right direction. Formulas are relatively easy to come by – indeed this ease is the curse of the theory – and represent a tendency that was tolerable in Siegel's day, but has now perhaps to be regarded as a vice. They are, in my view, the doubtful side of his legacy. Many of the deeper structural questions on the other hand – not all covered by functoriality – are, like those in nearby branches of algebraic geometry and number theory, difficult to broach and in danger of being neglected.

Whether the ideas sketched here are indeed in the right direction, I do not know. I would like to think about them as concretely as possible, but it is clear that a single mathematician, especially an aging one who is inclined to scatter his forces and has no knowledge of the techniques of analytic number theory, is not likely to take them very far.

There is a question that we have left aside and to which we promised to return. There appears to be no danger of proving much about Ramanujan if one manages to treat icosahedral representations in the manner proposed. The means proposed require estimates weaker than usual, as we average over a fixed power of primes, and we would end up with at best the Ramanujan-Petersson for representations of icosahedral type, thus, as far as that conjecture goes, with nothing at all. Observe, for what it is worth, that the method would use an average on primes outside of S but the conclusion would apply to primes inside S. If we then try to apply the same method to problem (T1), which does have consequences for the Ramanujan-Petersson conjecture, as we propose applying to (T2), thus to take G to be $GL(m+1)$ and J to be the image in ${}^LG$ under the m-th symmetric power of the L-group of GL(2), then we will be averaging again over primes, but in the context of $GL(m+1)$ and therefore be attempting, I believe, to deal with (T1) with methods quite different than ones usually applied to establish weak forms of the Ramanujan-Petersson conjecture, but I am not sure! We can hope – if we dare – that the varieties (26) will be replaced by other explicitly treatable varieties!

We will then be applying the trace formula first for $GL(m+1)$ and to a function whose component at p is a spherical function ϕ_p such that

$$\operatorname{tr}\pi(\phi_p) = \operatorname{tr} r(A_p(\pi)),$$

where $A_p(\pi)$ is the Hecke parameter of π and r is an irreducible representation of the L-group of $GL(m+1)$, chosen such that its restriction to the image of $GL(2)$ contains the trivial representation a large number of times. Since we shall have to examine the spectral side and peel off all contributions that violate Ramanujan, and then turn to the geometric side, and again isolate and remove the matching contributions, we are going to need a solid grounding in the trace formula in order to see such an enterprise through to the end. The results of the averaging will have to be so arranged that they can be compared with the trace formula for $GL(2)$. But I am rushing things!

I return briefly to (T3). We now want to project not on those representations for which $m(\pi, r)$ has a given value but on those for which π_∞ is one of two representations, those corresponding to the two representations of $PGL(2,\mathbb{R})$ corresponding to the two different representations of $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ in $SL(2,\mathbb{C})$, thus to

$$\iota \to \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon \end{pmatrix},$$

where ι is complex conjugation and $\epsilon = \pm 1 = (-1)^e$. These are both representations induced from characters of the diagonal matrices, namely

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \to \left(\frac{\operatorname{sgn}\alpha}{\operatorname{sgn}\beta}\right)^e.$$

Therefore we want to apply the trace formula to a function whose component ψ_∞ has a Fourier transform which is the δ-function concentrated at these two tempered representations. If I am not mistaken this will be a function for which the term ψ'_∞ in (21) is zero and the term ψ''_∞ is identically one. Of course, such a function does not have compact support, but has to be approximated. This approximation can perhaps be regarded as an analogue of the averaging used for (T2). For (T3), the full function Φ would not contain the factor ϕ_p^r; there would be no distinguished place p; and ϕ_q would be the unit element everywhere outside of S. Otherwise we would proceed in the same way, first stripping off the terms on the geometric side corresponding to the one-dimensional representations. If I am right about the orbital integrals that we need to approximate, this process might be fairly elegant. We can expect a result similar to that for (T2), although perhaps simpler, for the dihedral representations will no longer occur, but once again, it is perhaps best not to anticipate too much. There is considerable room for disappointment.