2 The Trace Formula and Hecke Operators

JAMES ARTHUR

This lecture is intended as a general introduction to the trace formula. We shall describe a formula that is a natural generalization of the Selberg trace formula for compact quotient. Selberg also established trace formulas for noncompact quotients of rank 1, and our formula can be regarded as an analogue for general rank of these. As an application, we shall look at the "finite case" of the trace formula. We shall describe a finite closed formula for the traces of Hecke operators on certain eingenspaces.

A short introduction of this nature will by necessity be rather superficial. The details of the trace formula are in [1(e)] (and the references there), while the formula for the traces of Hecke operators is proved in [1(f)]. There are also other survey articles [1(c)], [1(d)], [5], and [1(g)], where some of the topics in this paper are discussed in more detail and others are treated from a different point of view.

1

Suppose that G is a locally compact group that is unimodular and that Γ is a discrete subgroup of G. There is a right G-invariant measure on the coset space $\Gamma\backslash G$ that is uniquely determined up to a constant

NUMBER THEORY,
TRACE FORMULAS and
DISCRETE GROUPS

ISBN 0-12-067570-6

multiple. We can therefore take the Hilbert space $L^2(\Gamma\backslash G)$ of square integrable functions on $\Gamma\backslash G$. Define

$$(R(y)\phi)(x) = \phi(xy), \qquad \phi \in L^2(\Gamma\backslash G),\ x,\ y \in G.$$

Then R is a unitary representation of G on $L^2(\Gamma\backslash G)$.

One would like to obtain information on the decomposition of R into irreducible representations. Selberg approached the problem by studying the operators

$$R(f) = \int_G f(y)R(y)\,dy, \qquad f \in C_c(G),$$

on $L^2(\Gamma\backslash G)$. If ϕ belongs to $L^2(\Gamma\backslash G)$, one can write

$$\begin{aligned}(R(f)\phi)(x) &= \int_G f(y)\phi(xy)\,dy \\ &= \int_G f(x^{-1}y)\phi(y)\,dy \\ &= \int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma} f(x^{-1}\gamma y)\phi(\gamma y)\,dy \\ &= \int_{\Gamma\backslash G}\Big(\sum_{\gamma\in\Gamma} f(x^{-1}\gamma y)\Big)\phi(y)\,dy.\end{aligned}$$

Therefore, $R(f)$ is an integral operator with kernel

$$K(x, y) = \sum_{\gamma\in\Gamma} f(x^{-1}\gamma y). \tag{1.1}$$

Suppose first that $\Gamma\backslash G$ is compact. Then under some mild restriction on f, the operator $R(f)$ is of trace class, and its trace is equal to the integral of the kernel on the diagonal. This is so, for example, if G is a Lie group and f is smooth. One can then write

$$\begin{aligned}\operatorname{tr} R(f) &= \int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma} f(x^{-1}\gamma x)\,dx \\ &= \int_{\Gamma\backslash G}\sum_{\gamma\in\{\Gamma\}}\ \sum_{\delta\in\Gamma_\gamma\backslash\Gamma} f(x^{-1}\delta^{-1}\gamma\delta x)\,dx \\ &= \sum_{\gamma\in\{\Gamma\}}\int_{\Gamma_\gamma\backslash G} f(x^{-1}\gamma x)\,dx \\ &= \sum_{\gamma\in\{\Gamma\}}\int_{G_\gamma\backslash G}\int_{\Gamma_\gamma\backslash G_\gamma} f(x^{-1}u^{-1}\gamma u x)\,du\,dx \\ &= \sum_{\gamma\in\{\Gamma\}} a_\Gamma^G(\gamma)\int_{G_\gamma\backslash G} f(x^{-1}\gamma x)\,dx,\end{aligned}$$

where

$$a_\Gamma^G(\gamma) = \text{volume}(\Gamma_\gamma \backslash G_\gamma).$$

Here, $\{\Gamma\}$ is a set of representatives of conjugacy classes in Γ, and Γ_γ and G_γ denote the centralizers of γ in Γ and G. Implicit in the discussion is the absolute convergence of the various sums and integrals. Now it can also be seen that R decomposes into a direct sum of irreducible unitary representations with finite multiplicities. It follows that

$$\text{tr } R(f) = \sum_{\pi \in \Pi(G)} a_\Gamma^G(\pi) \text{ tr } \pi(f),$$

where $\Pi(G)$ is a set of equivalence classes of irreducible representations of G, and $a_\Gamma^G(\pi)$ is a positive integer. We can therefore write

$$\sum_{\gamma \in \{\Gamma\}} a_\Gamma^G(\gamma) I_G(\gamma, f) = \sum_{\pi \in \Pi(G)} a_\Gamma^G(\pi) I_G(\pi, f), \tag{1.2}$$

where

$$I_G(\gamma, f) = \int_{G_\gamma \backslash G} f(x^{-1}\gamma x)\, dx$$

and

$$I_G(\pi, f) = \text{tr } \pi(f).$$

This is the Selberg trace formula for compact quotient, introduced in [6(a)]. (Selberg's original formula actually took a slightly different form. The present form is due to Tamagawa [9].)

Example 1. Suppose that $G = \mathbb{R}$ and $\Gamma = \mathbb{Z}$. Then (1.2) becomes

$$\sum_{n \in \mathbb{Z}} f(n) = \sum_{n \in \mathbb{Z}} \hat{f}(2\pi n), \qquad f \in C_c^\infty(\mathbb{R}),$$

the Poisson summation formula.

Example 2. Suppose that G is a finite group and that $\Gamma \subset G$ is an arbitrary subgroup. Let π be an irreducible unitary representation of G and set

$$f(x) = \text{tr } \pi(x^{-1}).$$

Writing the left-hand side of the trace formula as

$$\sum_{x \in \Gamma \backslash G} \sum_{\gamma \in \Gamma} \text{tr } \pi((x^{-1}\gamma x)^{-1}) = |G|\,|\Gamma|^{-1} \sum_{\gamma \in \Gamma} \text{tr } \pi(\gamma^{-1}),$$

and applying the theory of characters, we find that $a_\Gamma^G(\pi)$ equals the multiplicity of the trivial representation in the restriction of π to Γ. But $a_\Gamma^G(\pi)$ is, by definition, the multiplicity of π in the representation of G induced from the trivial representation of Γ. The equality of these two multiplicities is just Frobenius reciprocity for finite groups. Frobenius reciprocity applies more generally to an arbitrary irreducible unitary representation of Γ, but so in fact does the Selberg trace formula. The arguments above apply equally well to spaces of square integrable, Γ-equivariant sections on G. The Selberg trace formula is therefore a generalization of Frobenius reciprocity.

We have chosen the notation $I_G(\gamma, f)$ and $I_G(\pi, f)$ in (1.2) to emphasize that as distributions in f, these functions are invariant. They remain unchanged if f is replaced by a conjugate

$$f^y(x) = f(yxy^{-1}).$$

The importance of such distributions is that they are completely determined from only partial information on f. One could expect to be able to evaluate any invariant distribution only knowing the orbital integrals of f on the conjugacy classes of G, or alternatively, the values of the characters at f of the irreducible unitary representation of G.

Consider, for example, the special case that $G = \mathrm{SL}(2, \mathbb{R})$. Assume that f is smooth and bi-invariant under the maximal compact subgroup $\mathrm{SO}(2, \mathbb{R})$. This was the case Selberg treated in greatest detail. The value at f of any invariant distribution depends only on the symmetric function

$$g(u) = |e^{u/2} - e^{-u/2}| \int_{A\backslash G} f\left(x^{-1}\begin{bmatrix} e^{u/2} & 0 \\ 0 & e^{-u/2} \end{bmatrix} x\right) dx, \qquad u \in \mathbb{R},\ u \neq 0,$$

where A denotes the subgroup of diagonal matrices in $\mathrm{SL}(2, \mathbb{R})$. It could equally well be expressed in terms of the function

$$h(r) = \int_{-\infty}^{\infty} e^{iru} g(u)\, du = \mathrm{tr}\ \pi_r(f), \qquad r \in \mathbb{R},$$

in which $\{\pi_r\}$ is the principal series of induced representations. Written in terms of g and h, (1.2) becomes the more concrete formula given on page 74 of [8(a)]. Selberg noticed a remarkable similarity between this formula and the "explicit formulas" of algebraic number theory. The analogue of the numbers

$$\{\log (p^n) : p \text{ prime},\ n \geq 1\}$$

is the set of lengths of closed geodesics on the Riemann surface

$$X_\Gamma = \Gamma\backslash \mathrm{SL}(2, \mathbb{R})/\mathrm{SO}(2, \mathbb{R}),$$

while the role of the zeros of the Riemann zeta-function is played by the points

$$\left\{\frac{1}{2} + \sqrt{-1}\,r : \pi_r \in \Pi(G)\right\}.$$

By choosing g suitably, Selberg obtained a sharp asymptotic estimate for the number of closed geodesics of length less than a given number. By varying h instead, he established an asymptotic formula for the distribution of the eigenvalues of the Laplacian of X_Γ.

Now, suppose only that $\Gamma\backslash G$ has finite invariant volume. For example, Γ could be a congruence subgroup

$$\{\gamma \in \mathrm{SL}(2, \mathbb{Z}) : \gamma \equiv 1 (\mathrm{mod}\ N)\}$$

of $\mathrm{SL}(2, \mathbb{R})$. Then everything becomes much more difficult. For $G = \mathrm{SL}(2, \mathbb{R})$, Selberg derived a trace formula in detail. Among other things, he gave a finite closed formula for the trace of the Hecke operators on the space of modular forms of weight $2k$, for $k > 1$. (See [8(a), p. 85].) In the later paper [8(b)], Selberg outlined an argument for establishing a trace formula for noncompact quotient when G has real rank 1. He also emphasized the importance of establishing such a result in general.

In this lecture we shall describe a general trace formula. It will be valid if G is an arbitrary reductive Lie group, and Γ is any arithmetic subgroup that is defined by congruence conditions.

2

In dealing with congruence subgroups, it is most efficient to work over the adèles. Therefore, we change notation slightly and take G to be a reductive algebraic group over $\mathbb{Q}$. The adèle ring

$$\mathbb{A} = \mathbb{R} \times \mathbb{A}_0 = \mathbb{R} \times \mathbb{Q}_2 \times \mathbb{Q}_3 \times \mathbb{Q}_5 \times \cdots$$

is locally compact ring that contains $\mathbb{Q}$ diagonally as a discrete subring. Moreover, $G(\mathbb{A})$ is a locally compact group that contains $G(\mathbb{Q})$ as a discrete subgroup. At first glance, $G(\mathbb{Q})\backslash G(\mathbb{A})$ might seem an ungainly substitute for the quotient of $G(\mathbb{R})$ by a congruence subgroup. However, the study of the two are equivalent. We shall assume for simplicity that G is semisimple and simply connected, and that $G(\mathbb{R})$

has no compact factors. Let K_0 be an open compact subgroup of the group $G(\mathbb{A}_0)$ of finite adèles. The strong approximation theorem asserts that

$$G(\mathbb{A}) = G(\mathbb{Q})K_0G(\mathbb{R}). \tag{2.1}$$

If follows that there is a $G(\mathbb{R})$-equivariant homeomorphism

$$\Gamma\backslash G(\mathbb{R}) \xrightarrow{\sim} G(\mathbb{Q})\backslash G(\mathbb{A})/K_0,$$

where

$$\Gamma = G(\mathbb{Q})K_0 \cap G(\mathbb{R})$$

is a congruence subgroup of $G(\mathbb{R})$. Conversely, any congruence subgroup can be obtained in this way. Thus, instead of working with $L^2(\Gamma\backslash G(\mathbb{R}))$, one can work with the $G(\mathbb{R})$-invariant subspace of functions in $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})$ that are right invariant under K_0. The advantage of the adèlic picture is that the conjugacy classes $G(\mathbb{Q})$ are much easier to deal with than those of Γ.

Suppose that f is a function in $C_c^\infty(G(\mathbb{A}))$. This means that f is a finite linear combination of functions $f_{\mathbb{R}}f_0$, where $f_{\mathbb{R}}$ belongs to $C_c^\infty(G(\mathbb{R}))$ and f_0 is a locally constant function of compact support on $G(\mathbb{A}_0)$. For example, if one is interested in the action of a function $f_{\mathbb{R}}$ on $L^2(\Gamma\backslash G(\mathbb{R}))$, one could take f to be the product of $f_{\mathbb{R}}$ with the unit function 1_{K_0}. (By definition, 1_{K_0} is the characteristic function of K_0 divided by the volume of K_0). Consider the values

$$K(x, x) = \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1}\gamma x), \qquad x \in G(\mathbb{Q})\backslash G(\mathbb{A}),$$

of the kernel on the diagonal. Since $G(\mathbb{Q})\backslash G(\mathbb{A})$ is not in general compact, this function is not generally integrable. What causes the integral to diverge? Experiments with examples suggest that the contribution to the integral of a conjugacy class in $G(\mathbb{Q})$ diverges when the conjugacy class intersects a proper parabolic subgroup defined over $\mathbb{Q}$. The more parabolics it meets, the worse will generally be the divergence of the integral. It turns out that by adding a correction term for each standard parabolic subgroup of G, one can truncate $K(x, x)$ in a uniform way so that its integral converges. Let us briefly describe this process in the special case that $G = \mathrm{SL}(n)$. (For a fuller illustration of the case of $\mathrm{SL}(n)$, see the survey [1(c)].)

The standard parabolic subgroups P of $\mathrm{SL}(n)$ are parametrized by partitions $(n_1, \ldots, n_k)$ of n. To each such partition corresponds subgroups

$$P = \left\{ \begin{pmatrix} \boxed{*}_{n_1} & * & \cdots & * \\ & \boxed{*}_{n_2} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & \boxed{*}_{n_k} \end{pmatrix} \right\},$$

$$M = \left\{ \begin{pmatrix} \boxed{*}_{n_1} & & & 0 \\ & \boxed{*}_{n_2} & & \\ & & \ddots & \\ 0 & & & \boxed{*} \end{pmatrix} \right\},$$

$$N = \left\{ \begin{pmatrix} I & \cdots & \cdots & * \\ & \boxed{I} & & \vdots \\ & & \ddots & \vdots \\ 0 & & & I \end{pmatrix} \right\},$$

and

$$A_M = A_P = \left\{ \begin{pmatrix} t_1 I & & & 0 \\ & \boxed{t_2 I} & & \\ & & \ddots & \\ 0 & & & t_k I \end{pmatrix} \right\}$$

of $\mathrm{SL}(n)$. One has the decomposition

$$\mathrm{SL}(n, \mathbb{A}) = P(\mathbb{A})K = N(\mathbb{A})M(\mathbb{A})K,$$

where

$$K = \mathrm{SO}_n(\mathbb{R}) \times \left(\prod_p \mathrm{SL}_n(\mathbb{Z}_p) \right)$$

is the maximal compact subgroup. For any point

$$x = n \begin{pmatrix} m_1 & & \\ & \ddots & \\ & & m_k \end{pmatrix} k, \qquad n \in N(\mathbb{A}),\ m_i \in GL(n_i, \mathbb{A}),\ k \in K,$$

in $\mathrm{SL}(n, \mathbb{A})$, set

$$H_P(x) = (\log |\det m_1|, \ldots, \log |\det m_k|).$$

Then $H_P(x)$ takes values in the real vector space

$$\mathbf{a}_M = \mathbf{a}_P = \{(u_1, \ldots, u_k) : \textstyle\sum u_i = 0\}.$$

The truncation of $K(x, x)$ will depend on a point

$$T = (t_1, \ldots, t_n), \qquad t_i \in \mathbb{R}, \textstyle\sum t_i = 0,$$

which is suitably regular, in the sense that t_i is much larger than t_{i+1} for every i. For any P, write

$$T_P = (t_1 + \cdots + t_{n_1}, t_{n_1+1} + \cdots + t_{n_2}, \ldots)$$

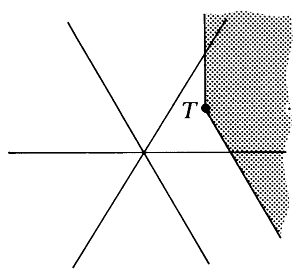
for the corresponding point in $\mathbf{a}_P$, and let $\hat{\tau}_P$ be the characteristic function of

$$\{(u_1, \ldots, u_k) \in \mathbf{a}_P : u_1 + \cdots + u_i > u_{i+1} + \cdots + u_k, \qquad 1 \leq i \leq k\}.$$

Example 3. Suppose that $G = \mathrm{SL}(3)$ and that P corresponds to the minimal partition $(1, 1, 1)$. Then $\mathbf{a}_P$ is a two-dimensional space with six chambers. The function

$$H \to \hat{\tau}_P(H - T_P), \qquad H \in \mathbf{a}_P,$$

is the characteristic function of the convex shaded dual chamber.



The notion of a standard parabolic subgroup exists, of course, for arbitrary G, as do the other objects described for $\mathrm{SL}(n)$. For any P, we have a kernel

$$K_P(x, y) = \int_{N(\mathbb{A})} \sum_{\gamma \in M(\mathbb{Q})} f(x^{-1}\gamma n y)\, dn,$$

for the right convolution operator of f on $L^2(N(\mathbb{A})M(\mathbb{Q})\backslash G(\mathbb{A}))$. We can now write down the truncated kernel. It is an expression

$$k^T(x, f) = \sum_P (-1)^{\dim A_P} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_P(\delta x, \delta x)\hat{\tau}_P(H_P(\delta x) - T_P), \quad (2.2)$$

defined by a finite alternating sum over standard parabolic subgroups P of G. The inner sum over δ may also be taken over a finite set, but this depends upon x. The purpose of this inner sum is to make the corresponding term left $G(\mathbb{Q})$-invariant. Observe that the term corresponding to $P = G$ is just $K(x, x)$. A term corresponding to $P \neq G$ can be regarded as a function that is supported on a neighborhood of infinity in $G(\mathbb{Q})\backslash G(\mathbb{A})$.

Theorem 2.1.

(a) *The function $k^T(x, f)$ is integrable over $G(\mathbb{Q})\backslash G(\mathbb{A})$.*

(b) *The function*

$$J^T(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})} k^T(x, f)\, dx$$

is a polynomial, for values of T that are suitably regular.

(See [1(a), Theorem 7.1] and [1(b), Proposition 2.3].)

As a polynomial, $J^T(f)$ can be extended to all values of T, even though it is defined as a convergent integral only for T in some chamber. Set

$$J(f) = J^0(f).$$

Then J is a distribution on $C_c^\infty(G(\mathbb{A}))$. An obvious question is how to evaluate it more explicitly.

3

Theorem 2.1 is just the first of a number of steps. We have described it in order to give some flavor of what is involved. The remaining steps are more elaborate, and we shall discuss them in only the most cursory manner.

Theorem 2.1 provides only a definition of a distribution $J(f)$. There is not yet any trace formula. For this, one needs to look at the representation theoretic expansion of $K(x, x)$. Since $G(\mathbb{Q})\backslash G(\mathbb{A})$ is generally noncompact, R is no longer a direct sum of irreducible representations. Rather, we have

$$R = R_{\mathrm{disc}} \oplus R_{\mathrm{cont}},$$

where R_{disc} is a direct sum of irreducible representations, and R_{cont} decomposes continuously. The decomposition of R_{cont} can be described in terms of the constituents of the analogues of R_{disc} for Levi components of standard parabolic subgroups. This description is part of the theory of Eisenstein series initiated by Selberg [8(a)] and established for general groups by Langlands [6(b)]. We shall say here only that the theory of Eisenstein series provides a second formula for the kernel. One obtains

$$K(x, y) = \sum_{P_1} \sum_{\phi} \sum_{i\mathfrak{a}_{P_1}^*} E(x, I_{P_1}(\lambda, f)\phi, \lambda)\overline{E(y, \phi, \lambda)}\, d\lambda, \tag{3.1}$$

where P_1 is summed over standard parabolic subgroups, ϕ is summed over an orthonormal basis of $L^2_{\mathrm{disc}}(P_1(\mathbb{Q})A_{P_1}(\mathbb{R})^0 N_{P_1}(\mathbb{A})\backslash G(\mathbb{A}))$ (the subspace that decomposes discretely under the action of $G(\mathbb{A})$), $I_{P_1}(\lambda)$ is an induced representation, and $E(\cdot, \cdot, \cdot)$ is the Eisenstein series (or, rather, its analytic continuation to imaginary λ). Notice that the term corresponding to $P_1 = G$ is just the kernel of the operator $R_{\mathrm{disc}}(f)$. More generally, for any P, Eisenstein series give a second formula for the kernel $K_P(x, y)$. One has only to restrict the sum in (3.1) to those P_1 that are contained in P and to take partial Eisenstein series from P_1 to P. Substituted into (2.2), these formulas provide a second expression for $k^T(x, f)$.

Thus, we have two distinct expressions for the integrable function $k^T(x, f)$. One is a geometric expansion related to conjugacy classes, which originates with the formula (1.1), and the other is a spectral expansion related to representation theory, which originates with the formula (3.1). We therefore obtain two expressions for the integral

$$J^T(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})} k^T(x, f)\, dx.$$

At this stage, the two expressions are too abstract to be of much value. Nevertheless, it turns out that each can be rewritten in a rather explicit form. This is by far the most difficult part of the process. In the end, however, one obtains two different formulas for the polynomial $J^T(f)$. In particular, by specializing T one obtains two different formulas for the distribution $J(f)$. One is a geometric expansion

$$J(f) = \sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \sum_{\gamma \in \{M(\mathbb{Q})\}} a^M(\gamma) J_M(\gamma, f), \tag{3.2}$$

and the other is a spectral expansion

$$J(f) = \sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \int_{\Pi(M)} a^M(\pi) J_M(\pi, f)\, d\pi. \tag{3.3}$$

The trace formula can be regarded as the equality of the two.

We shall have to be content to say only a few words about the terms in (3.2) and (3.3). In each case, M is summed over Levi components of standard parabolic subgroups P of G, and $|\mathrm{ch}(\mathbf{a}_M)|$ denotes the number of chambers in the vector space $\mathbf{a}_M$. In (3.2), $\{M(\mathbb{Q})\}$ stands for the conjugacy classes in $M(\mathbb{Q})$. In (3.3), $\Pi(M)$ is a set of irreducible unitary representations of the group

$$M(\mathbb{A})^1 = \{x \in M(\mathbb{A}) : H_P(x) = 0\},$$

equipped with a certain measure $d\pi$. The functions $a^M(\gamma)$ and $a^M(\pi)$ are global in nature and depend only on the subgroup M. If $M(\mathbb{Q})\backslash M(\mathbb{A})^1$ is compact, they are equal to the coefficients $a^{M(\mathbb{A})^1}_{M(\mathbb{Q})}(\gamma)$ and $a^{M(\mathbb{A})^1}_{M(\mathbb{Q})}(\pi)$ that occur in (1.2). The distributions $J_M(\gamma, f)$ and $J_M(\pi, f)$ are local in nature. If $M = G$, they equal the distributions $I_{G(\mathbb{A})}(\gamma, f)$ and $I_{G(\mathbb{A})}(\pi, f)$ in (1.2). For $M \neq G$, however, they are more complicated. For example, $J_M(\gamma, f)$ is the orbital integral of f over the conjugacy class of γ, but not with respect to the invariant measure. The invariant measure has instead to be weighted by the volume of a certain convex hull, which depends on x.

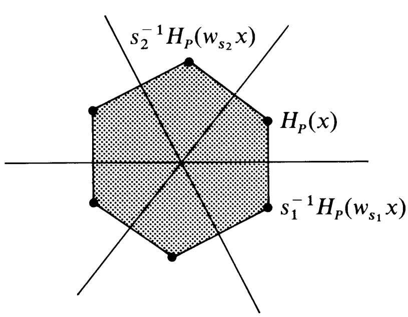
Example 3 (continued). Suppose that $G = \mathrm{SL}(3)$, that M corresponds to the minimal partition $(1, 1, 1)$, and that γ is a diagonal element with distinct eigenvalues. Then

$$J_M(\gamma, f) = \int_{A_M(\mathbb{A})\backslash G(\mathbb{A})} f(x^{-1}\gamma x) v_M(x)\, dx,$$

where $v_M(x)$ is the volume of the convex hull of the set

$$\{s^{-1}H_P(w_s x) : s \in W(\mathbf{a}_M)\}.$$

The Weyl group $W(\mathbf{a}_M)$ here is isomorphic to S_3, and for every element s, w_s is the associated permutation matrix. The convex hull is represented by the shaded region



The example illustrates a weakness of the identity obtained from (3.2) and (3.3). Unlike the formula (1.2) for compact quotient, the individual terms are not invariant distributions. They depend on more than just the irreducible characters of f. Fortunately, there is a way to rectify this. For technical reasons, one must insist that f be K-finite, but this is of no great consequence. Then there is natural process that associates an invariant linear functional $I(f)$ to $J(f)$. The same process attaches invariant linear functions $I_M(\gamma, f)$ and $I_M(\pi, f)$ to $J_M(\gamma, f)$ and $J_M(\pi, f)$ such that

$$I(f) = \sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \sum_{\gamma \in \{M(\mathbb{Q})\}} a^M(\gamma) I_M(\gamma, f)$$

and

$$I(f) = \sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \int_{\Pi(M)} a^M(\pi) I_M(\pi, f)\, d\pi.$$

The invariant trace formula is just the identity

$$\sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \sum_{\gamma \in \{M(\mathbb{Q})\}} a^M(\gamma) I_M(\gamma, f) = \sum_M |\mathrm{ch}(\mathbf{a}_M)|^{-1} \int_{\Pi(M)} a^M(\pi) I_M(\pi, f)\, d\pi. \quad (3.4)$$

(The details of the construction are contained in [1(e), Sections I.2–I.3, Sections II.2–II,4]. For a general idea of how it works, see the introduction to [1(b)].) If $M = G$, the distributions $J_G(\gamma, f)$ and $J_G(\pi, f)$ are already invariant, and the process does not alter them. Consider the special case that G is anisotropic over $\mathbb{Q}$. Then there are no proper (rational) parabolic subgroups, and the only summands come from $M = G$. The formula (3.4) reduces to (1.2), which is to be expected since $G(\mathbb{Q})\backslash G(\mathbb{A})$ is compact. Thus, (3.4) is a generalization of (1.2) in which the additional terms are contributed by the proper Levi components M.

In the interests of simplicity, we have passed over two technical complications. On the left-hand side of (3.4) (as well as in (3.2)), the notation should actually include a large finite set S of valuations of $\mathbb{Q}$ (that depends in a simple way on f). For if $\gamma \in M(\mathbb{Q})$ is unipotent, the orbital integral on $G(\mathbb{A})$ at γ diverges. It must instead be taken over $G(\mathbb{Q}_S)$. The functions $a^M(\gamma)$ and $I_M(\gamma, f)$, and also the conjugacy relation $\{M(\mathbb{Q})\}$, really depend implicitly on S. The other point is that the integrals over $\Pi(M)$ in (3.3) and (3.4) are not known to converge. This is tied up with the fact that the operator $R_{\mathrm{disc}}(f)$ is not known to be of trace class. However, there is a way to group the terms in the sum-integral over M and $\Pi(M)$ to make them converge. This is supposed to be implicit in (3.4). These complications are serious if one wants to derive the kind of asymptotic formulas available for compact quotient. However, they do not seem to be of any consequence in applications such as base change, that entail a comparison of two trace formulas.

The expression (3.4) is certainly a formula, but the reader is perhaps wondering where the trace is. It is buried in the term corresponding to $M = G$, on the spectral side of (3.4). We have not described $a^G(\pi)$ in general, but this function is actually defined explicitly as a finite sum of terms, one of which is the multiplicity with which π occurs discretely in $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})^1)$. If we were able to transfer everything but these multiplicities to the left-hand side of (3.4), we would be left with a formula for the trace of $R_{\mathrm{disc}}(f)$. This, however, is not allowed, since we don't know at the moment that $R_{\mathrm{disc}}(f)$ is of trace class. What is known to be of trace class is $R_{\mathrm{cusp}}(f)$, the restriction of $R_{\mathrm{disc}}(f)$ to the space of cusp forms. One can always rewrite (3.4) as a formula for the trace of $R_{\mathrm{cusp}}(f)$. It simply entails a convergent grouping of the terms that would otherwise be the formula for the trace of $R_{\mathrm{disc}}(f)$.

4

Suppose that K_0 is an open compact subgroup of $G(\mathbb{A}_0)$. The Hecke algebra

$$H_{K_0} = C_c(K_0\backslash G(\mathbb{A}_0)/K_0)$$

acts on $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})/K_0)$ on the right by convolution. If

$$\Gamma = G(\mathbb{Q})K_0 \cap G(\mathbb{R})$$

as before, we have a $G(\mathbb{R})$-isomorphism

$$L^2(G(\mathbb{Q})\backslash G(\mathbb{A})/K_0) \xrightarrow{\sim} L^2(\Gamma\backslash G(\mathbb{R})).$$

It is easy to describe the action of H_{K_0} in terms of the space on the right. Writing

$$g = g_{\mathbb{R}}g_0, \qquad g_{\mathbb{R}} \in G(\mathbb{R}),\ g_0 \in G(\mathbb{A}_0),$$

for any element $g \in G(\mathbb{Q})$, we first observe from (2.1) that

$$\{g_0 : g \in G(\mathbb{Q})\}$$

is dense in $G(\mathbb{A}_0)$. We can therefore assume that $h \in \boldsymbol{H}_{K_0}$ is the product of $\mathrm{vol}(K_0)^{-1}$ with the characteristic function of

$$K_0 g K_0, \qquad g \in G(\mathbb{Q}),$$

since any element in $\boldsymbol{H}_{K_0}$ is a linear combination of such functions. Let ϕ be a function in the Hilbert space, and take $x \in G(\mathbb{R})$. Then

$$\begin{aligned}(\phi * h)(x) &= \int_{G(\mathbb{A}_0)} \phi(xy)h(y)\,dy \\ &= \sum_{k \in K_0/K_0 \cap g_0 K_0 g_0^{-1}} \phi(xkg_0) \\ &= \sum_{\gamma \in \Gamma_{\mathrm{diag}}/\Gamma_{\mathrm{diag}} \cap g \Gamma_{\mathrm{diag}} g^{-1}} \phi(\gamma_0 g_0 x) \\ &= \sum_{\gamma_{\mathbb{R}} \in \Gamma/\Gamma \cap g_{\mathbb{R}} \Gamma g_{\mathbb{R}}^{-1}} \phi(g_{\mathbb{R}}^{-1} \gamma_{\mathbb{R}}^{-1} x),\end{aligned}$$

where

$$\Gamma_{\mathrm{diag}} = G(\mathbb{R})K_0 \cap G(\mathbb{Q}).$$

This is closer to the classical definition. Actually, in the special case that $G = \mathrm{SL}(2)$ and $\Gamma = \mathrm{SL}(2, \mathbb{Z})$, the prescription above gives only the classical Hecke operators

$$T(n), \qquad n \in \mathbb{N},$$

in which n is a square. To get them all, one would need to take GL(2), a nonsemisimple group that we excluded with our original simplifying assumption.

Suppose that $\pi_{\mathbb{R}}$ is an irreducible unitary representation of $G(\mathbb{R})$. Let $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$ be the multiplicity with which $\pi_{\mathbb{R}}$ occurs discretely in $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})/K_0)$. If h belongs to $\boldsymbol{H}_{K_0}$, let $R_{\mathrm{disc}}(\pi_{\mathbb{R}}, h)$ be the operator obtained by restricting h to the $\pi_{\mathbb{R}}$-isotypical subspace of $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})/K_0)$. It can be identified with an $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0) \times m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$-matrix. One would like an explicit formula for $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$ or, more generally, a formula for the trace of $R_{\mathrm{disc}}(\pi_{\mathbb{R}}, h)$. This, of course, is too much to ask in general. However, it is reasonable to ask the question when $\pi_{\mathbb{R}}$ belongs to $\Pi_{\mathrm{disc}}(G(\mathbb{R}))$, the discrete series of $G(\mathbb{R})$. This is essentially what Selberg's formula [8(a), p. 85] gives in the special case that $G = \mathrm{SL}(2)$, $\Gamma = \mathrm{SL}(2, \mathbb{Z})$, and $\pi_{\mathbb{R}}$ is any but the lowest discrete series.

Recall that $G(\mathbb{R})$ has a discrete series if and only if it has a compact Cartan subgroup. Assume that this is the case. Then $\Pi_{\mathrm{disc}}(G(\mathbb{R}))$ is disjoint union of finite sets $\Pi_{\text{disc}}(\mu)$, parametrized by the irreducible finite dimensional representations μ of G. The number $w(G)$ of elements in a set $\Pi_{\text{disc}}(\mu)$ equals the order of a quotient of Weyl groups and is independent of μ; the set itself contains all the representations in the discrete series with the same infinitesimal character as μ. These facts are, of course, part of the classification [4(b)] of Harish-Chandra.

We shall describe a formula for the sum over $\pi_{\mathbb{R}} \in \Pi_{\text{disc}}(\mu)$ of the traces of the Hecke operators. We must first define the terms that appear. Suppose that M is a Levi component of a standard parabolic subgroup of G over $\mathbb{Q}$. Let $\Phi'_M(\gamma, \mu)$ be the function on $M(\mathbb{R})$ which equals

$$|\det(1 - \mathrm{Ad}(\gamma))_{\mathbf{g}/\mathbf{m}}|^{1/2} \sum_{\pi_{\mathbb{R}} \in \Pi_{\text{disc}}(\mu)} \Theta_{\pi_{\mathbb{R}}}(\gamma)$$

if $\gamma \in M(\mathbb{R})$ is $\mathbb{R}$-elliptic (i.e., belongs to a Cartan subgroup that is compact modulo $A_M(\mathbb{R})$), and which equals 0 otherwise. Here $\Theta_{\pi_{\mathbb{R}}}$ stands for the character $\pi_{\mathbb{R}}$, and $\mathbf{g}$ and $\mathbf{m}$ are the Lie algebras of G and M. One can express $\Phi'_M(\gamma, \mu)$ in terms of formulas of Harish-Chandra [4(a)], which are reminiscent of the Weyl character formula. Observe that $\Phi'_M(\gamma, \mu)$ vanishes unless γ is semisimple. Now, suppose that γ belongs to $M(\mathbb{Q})$ and is semisimple. Write

$$h_M(\gamma) = \delta_P(\gamma_0)^{1/2} \int_{K_{0,\max}} \int_{N_P(\mathbb{A}_0)} \int_{M_\gamma(\mathbb{A}_0)\backslash M(\mathbb{A}_0)} h(k^{-1}m^{-1}\gamma mnk)\, dm\, dn\, dk,$$

where $\delta_P(\gamma_0)$ is the modular function of P, evaluated at the finite adèlic component of γ, and $K_{0,\max}$ is a suitable maximal compact subgroup of $G(\mathbb{A}_0)$. This is essentially an invariant p-adic orbital integral and is no more complicated than the distributions in the trace formula for compact quotient. Finally, there is a constant $\chi(M_\gamma)$ which is defined if $\gamma \in M(\mathbb{Q})$ is $\mathbb{R}$-elliptic. If G has no factors of type E_8,

$$\chi(M_\gamma) = (-1)^{q(M_\gamma)} \operatorname{vol}(\overline{M}_\gamma(\mathbb{Q})\backslash \overline{M}_\gamma(\mathbb{A}_0)) w(M_\gamma)^{-1},$$

where $q(M_\gamma)$ is the dimension of the symmetric space attached to M_γ and $\overline{M}_\gamma$ is any inner twist of M_γ such that $\overline{M}_\gamma(\mathbb{R})/A_M(\mathbb{R})^0$ is compact. This relies on a theorem of Kottwitz [5] that requires the Hasse principle. Otherwise, $\chi(M_\gamma)$ must be given by a slightly more complicated formula.

Theorem 4.1. *Suppose the μ is an irreducible finite dimensional representation of G whose highest weight is nonsingular. Then*

$$\sum_{\pi_{\mathbb{R}} \in \Pi_{\text{disc}}(\mu)} \operatorname{tr}(R_{\text{disc}}(\pi_{\mathbb{R}}, h))$$

equals

$$\sum_{M} (-1)^{\dim A_M} |\mathrm{ch}\,(\mathbf{a}_M)|^{-1} \sum_{\gamma \in \{M(\mathbb{Q})\}} \chi(M_\gamma) \Phi'_M(\gamma, \mu) h_M(\gamma).$$

This theorem is proved in [1(f), Corollary 6.2]. It expresses the trace of Hecke operators as a finite closed formula. The reader might want to compare the formula with those in [8(a), p. 85] (for $G = \mathrm{SL}(2)$), [2, p. 283] ($G = P\mathrm{GL}(2)$), and [7, p. 307] (G of real rank 1). The main step would be to convert the p-adic orbital integrals into suitable finite sums. However, for future applications to Shimura varieties, the formula is best left in adèlic form. This is more natural for comparison with the Lefshetz fixed point formula in characteristic p, as one can see from the formulas for GL(2) in [6(a), Sections 5–6].

References

[1.] Arthur, J.

(a) "A trace formula for reductive groups I: Terms associated to classes in G(Q)," *Duke Math. J.* **45** (1978), 911–952.

(b) "The trace formula in invariant form," *Ann. of Math.* **114** (1981), 1–74.

(c) "The trace formula for reductive groups," In: *Journées Automorphes*, Publications Mathématiques de l'Université de Paris VII, 1–41.

(d) "The trace formula for noncompact quotient," *Proceedings of the International Congress of Mathematicians, Warsaw, 1983,* **2** (1983), 849–859.

(e) "The invariant trace formula I. Local theory," *J. Amer. Math. Soc.* 1 (1988), 323–383; "II. Global theory," *J. Amer. Math. Soc.* **1** (1988), 501–554. Update.

(f) "The L^2-Lefschetz numbers of Hecke operators," preprint.

(g) "Characters, harmonic analysis, and an L^2-Lefschetz formula," update to appear in *Proc. Sympos. Pure. Math.* (Weyl Symposium), A.M.S.

[2.] Duflo, M., and Labesse, J.-P., "Sur la formule des traces de Selberg," *Ann. Scient. Éc. Norm. Sup.*, **4** (1971), 193–284.

[3.] Harish-Chandra

(a) "Discrete series for semisimple groups I," Acta Math. **113** (1965), 241–318; also *Collected Papers*, Vol. **III**, 403–481.

(b) "Discrete series for semisimple groups II," *Acta. Math.* **116** (1966), 1–111; also *Collected Papers*, Vol. **III**, 537–648.

[4.] Kottwitz, R., "Tamagawa numbers," *Ann. of Math.*, **127** (1988), 629–646. Update.

[5.] Labesse, J.-P., "La formule des traces d'Arthur-Selberg," (Séminaire Bourbaki, 636), *Astérisque* **133** (1986), 73–88.

[6.] Langlands, R. P.

(a) "Modular forms and l-adic representations," *Lecture Notes in Mathematics*. **349** (1973), 361–500.

(b) "On the Functional Equations Satisfied by Eisenstein Series," Lecture Notes in Mathematics. **544** (1976).

[7.] Osborne, M. S., and Warner, G. "Multiplicities of the integrable discrete series," *J. Funct. Anal.* **30** (1978), 287–310.

[8.] Selberg, A.

(a) "Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series," *J. Indian Math. Sol.* **20** 1956), 47–87.

(b) "Discontinuous groups and harmonic analysis," *Proceedings of the International Congress of Mathematicians, Stockholm.* (1962), 177–189.

[9.] Tamagawa, T. "On Selberg's trace formula," J. Fac. Sci. Univ. Tokyo, Sect. I. **8** (1960), 363–386.

Note Added in Proof.

The trace class problem has been recently solved by W. Müller ("The trace class conjecture in the theory of automorphic forms," preprint). In particular, (3.4) can now be written as a formula for the trace of $R_{\mathrm{disc}}(f)$. It would be interesting to investigate the convergence properties of the other terms on the right-hand side of (3.4).