HARISH-CHANDRA HOMOMORPHISMS FOR p-ADIC GROUPS

Roger Howe

Conference Board of the Mathematical Sciences
REGIONAL CONFERENCE SERIES IN MATHEMATICS

supported by the
National Science Foundation

Number 59

HARISH-CHANDRA HOMOMORPHISMS FOR $\mathfrak{p}$-ADIC GROUPS

Roger Howe
with the collaboration of Allen Moy

Published for the
Conference Board of the Mathematical Sciences
by the
American Mathematical Society
Providence, Rhode Island

Expository Lectures
from the CBMS Regional Conference
held at the University of Chicago
August 8–12, 1983

Research supported in part by National Science Foundation Grant DMS 83-03885.

1980 *Mathematics Subject Classifications* (1985 *Revision*). Primary 22E50; Secondary 16A27, 16A45.

Library of Congress Cataloging-in-Publication Data

Howe, Roger.

Harish-Chandra homomorphisms for *p*-adic groups.

(Regional conference series in mathematics, ISSN 0160-7642; no. 59)

On t.p. the German (Fraktur) lower case "*p*" is used "*p*-adic" in the title.

"Supported by the National Science Foundation."

"Expository lectures from the CBMS regional conference held at the University of Chicago, August 8–12, 1983"–T.p. verso.

Bibliography: p.

1. Lie groups. 2. Representations of groups. 3. *p*-adic groups. 4. Homomorphisms (Mathematics). I. Conference Board of the Mathematical Sciences. II. Title. III. Series.

QA1.R33 no. 59 510s [512′.55] 85-18649
[QA387]

ISBN 0-8218-0709-9 (alk. paper)

Preface

The Harish-Chandra homomorphism, between the centers of the universal enveloping algebras of a reductive Lie algebra and the Levi component of a parabolic subalgebra, plays a key role in representation theory of semisimple Lie groups. It is used in the study of asymptotics of matrix coefficients, the study of Lie algebra cohomology, the study of characters and the classification schemes of Langlands and of Vogan. In the context of real Lie groups, a particularly useful feature of the Harish-Chandra homomorphism is that, since it is purely algebraic in nature, it can be defined for θ-stable parabolics as well as for real ones, and thus can be used to control cohomological induction as well as ordinary induction from parabolics.

There are many parallels between representation theory on real reductive groups and p-adic groups, and each theory has illuminated the other. In particular, Harish-Chandra's philosophy of cusp forms has been basic for both theories. However, so far there has been no analog of the Harish-Chandra homomorphism in the p-adic case. To some extent the Jacquet functor has substituted for it. But the Jacquet functor makes sense only for rational parabolic subgroups, so it passes over supercuspidal representations in silence. And it is a construction on modules, not directly on convolution algebras, so it is not so organically a part of harmonic analysis as is the Harish-Chandra homomorphism.

The purpose of these notes is to propose a partial analogue for p-adic groups of the Harish-Chandra homomorphism. The overall form of the construction presented here is as follows. Let G be a reductive p-adic algebraic group, and let $G' \subseteq G$ be the centralizer of a semisimple element in G. (If the program begun here gets carried far enough, G' may become an "endoscopic" group and not actually contained in G; but the present formulation suffices for our purposes here.) In favorable cases, one has an intuition that certain representations of G are related to certain corresponding representations of G'. If G' is the Levi component of a parabolic subgroup P, then the desired correspondence is realized by taking a representation of G', extending it to P by letting the unipotent radical of P act trivially, then inducing the extended representation from P to G. (For real groups, this construction can be augmented by cohomological induction to cover a more or less general G', but for p-adic groups such a construction is currently unavailable.)

What is shown here (for the case $G = \mathrm{GL}_n$ and suitable G') is that one can find certain convolution algebras $\mathscr{H} \subseteq C_c^\infty(G)$ and $\mathscr{H}' \subseteq C_c^\infty(G')$ which are isomorphic in a straightforward and natural way. This allows one to transfer whole the subset of the representations of G "seen" by $\mathscr{H}$ (i.e., in which $\mathscr{H}$ is represented nontrivially) to the subset of the representations of G' seen by $\mathscr{H}'$. When G' is the Levi component of a parabolic subgroup, the correspondence between representations of G' and of G is just that given by extending to a parabolic containing G' and inducing (using normalized induction to preserve unitarity). However, the same phenomenon holds for G' which are not Levi components of parabolic subgroups (more properly, they are Levi components of parabolic subgroups defined only over some extension of the ground field). From such G' there is the possibility of constructing supercuspidal representations of G. In fact most of the constructions of supercuspidal representations in the literature rely on essentially the phenomenon under study here, though not in an explicit fashion.

Some convenient features of the construction given here are listed.

(i) Since the correspondences of representations it yields arise from identifications of Hecke algebras, there is an obvious sense in which they are canonical.

(ii) The isomorphisms constructed are very simple-minded and obviously allow transfer of important properties of representations expressible in terms of matrix coefficients: rates of decay, supercuspidality, square-integrability, temperedness, etc. (However, it is not obvious whether unitarity is always preserved, though there is some positive evidence.)

(iii) The correspondence preserves Plancherel measure. Thus, in particular, it provides an inductive computation of Plancherel measure and, especially, a method for finding the formal degrees of discrete series representations.

The main drawbacks of the construction are that it involves quite a lot of work, and that it is not clear how general it can be. Some remarks can be offered in partial amelioration of these points. First, the reader will be pleased to know that the presentation of the main results is considerably streamlined over the original treatment, given in the lectures to which these are the notes, and the results are improved. Second, the results given here can be extended with relatively little extra effort to cover centralizers of tamely ramified tori. They have also been extended to centralizers of unramified tori in other classical groups than GL_n. Third, though the details of proofs given here are rather painstaking, the formal scheme of the results has a robust simplicity to it that gives one hope that, properly formulated, they are quite general. It seems reasonable to hope that these methods could lead to an explicit Plancherel formula for groups with only tamely ramified tori. There is even some slight evidence that they remain valid for wildly ramified tori. But that is work for the future.

The notes follow more or less the outline of the lectures (given in Chicago, August 1983). In Chapter 1, the basic phenomena are illustrated in the case of $GL_n(\mathbf{F}_q)$, where $\mathbf{F}_q$ is the finite field with q elements, and they are applied to give an efficient sketch of the representation theory of $GL_n(\mathbf{F}_q)$. In Chapter 2, the transition is made to local fields, and the basic result is formulated and proved in the case of the Levi component of a (rational) parabolic. Chapter 3 then completes the proof of the main theorem by treating the case $G' = GL_m(F') \subseteq GL_n(F) = G$, where F' is an unramified extension of degree n/m of the ground field F. The results of Chapter 1 have been known to me for a long time; Allan Silberger and I discussed these questions at the Institute for Advanced Study in 1972. A treatment of such results for general Chevalley groups over finite fields, not quite from the same point of view, was given by Howlett and Lehrer [**HL**]. A more or less definitive discussion has been given by Lusztig in his recent book [**L2**]. By contrast, the results of Chapters 2 and 3 are recent and are the fruits of joint work with Allen Moy.

I would like to thank Paul Sally for his interest in these results and for organizing the CBMS conference where I reported on them. I would also like to thank him for preparing a preliminary draft of Chapter 1 and David Manderscheid for a preliminary draft of Chapter 2. I am pleased to recognize Ms. Donna Belli and Mrs. Mel DelVecchio for their rapid, accurate production of the typescript. I am very appreciative to the Guggenheim Foundation whose support gave me time to finish this and other writing projects. But, especially, I gladly yield a debt of thanks to Allen Moy, without whose aid, both mathematical and logistical, these notes would still be in limbo. Working with Allen has been a pleasure.

New Haven, CT
February, 1985

Contents

1. A Hecke Algebra Approach to the Representations of $\mathrm{GL}_n(\mathbf{F}_q)$

1. Introduction. Let $\mathbf{F}_q$ be the finite field with q elements, and let $\mathrm{GL}_n(\mathbf{F}_q)$ be the general linear group of n by n invertible matrices with coefficients in $\mathbf{F}_q$. The irreducible characters of $\mathrm{GL}_n(\mathbf{F}_q)$ were described by Green in [**Gr**]. Much more recently, Lusztig [**L1**] gave a description of some of the cuspidal representations of $\mathrm{GL}_n(\mathbf{F}_q)$. Using l-adic cohomology, Deligne-Lusztig [**DL**] then gave a construction of a large family of cuspidal representations of general reductive algebra groups over $\mathbf{F}_q$. Lusztig followed this, in a tour de force, by classifying the irreducible representations of more or less general reductive groups [**L2**].

In this chapter, we consider $\mathrm{GL}_n(\mathbf{F}_q)$ and give a rough description of the set $(\mathrm{GL}_n(\mathbf{F}_q))^\wedge$ of irreducible representations. We will not describe the characters or explicit modules for the representations. What we gain in exchange for setting such a modest goal is a fairly self-contained account, which uses only the most standard tools of representation thery, along with Harish-Chandra's "philosophy of cusp forms". The main observation underlying this account is that there is a natural isomorphism between certain spherical function algebras associated to different groups. The account of $\mathrm{GL}_n(\mathbf{F}_q)$ given in this chapter is introductory to the discussion of GL_n over p-adic fields given in Chapters 2 and 3. Certain results needed to complete the story for $\mathrm{GL}_n(\mathbf{F}_q)$ but not needed later are treated in appendices. This account may be regarded as introductory to the elegant treatment of $\mathrm{GL}_n(\mathbf{F}_q)$ by A. Zelevinski [**Z1**].

2. Notation and G/B. We will abbreviate $\mathrm{GL}_n(\mathbf{F}_q) = G$ when there is no need for more specific notation. The group G acts naturally on $\mathbf{F}_q^n = V_n$, the space of column vectors of length n over $\mathbf{F}_q$, by matrix multiplication. For $m \leqslant n$, we identify $\mathbf{F}_q^m = V_m$ with the subspace of V_n whose elements have their last $n-m$ coordinates equal to zero. Let V'_{n-m} be the subspace of V_n whose elements have their first m coordinates equal to zero. Then V'_{n-m} is a complement to V_m in V_n, and we have the direct sum decomposition

(2.1) $$V_n = V_m \oplus V'_{n-m}.$$

Let $A = \{0 = a_0 < a_1 < a_2 < \cdots < a_l = n\}$ be an increasing subsequence of the nonnegative integers up to n. We define the set of subspaces

$$\mathcal{F}_A = \{V_{a_i}: a_i \in A\} \tag{2.2}$$

to be the *standard flag* associated to A. We define groups

$$\begin{aligned} &\text{(i)} \quad P_A = \{g \in G: g(V_{a_i}) = V_{a_i}, a_i \in A\};\\ &\text{(ii)} \quad U_A = \{g \in P_A: (1-g)V_{a_i} \subseteq V_{a_{i-1}}, a_i \in A\};\\ &\text{(iii)} \quad M_A = \{g \in P_A: g(V'_{n-a_i}) = V'_{n-a_i}, a_i \in A\}. \end{aligned} \tag{2.3}$$

Then P_A is a *standard parabolic* subgroup of G with *unipotent radical* U_A and *Levi component* M_A. We have

$$\begin{aligned} P_A &= M_A U_A,\\ M_A &\simeq \prod_i \mathrm{GL}(V_{a_i} \cap V'_{n-a_{i-1}}) \simeq \prod_i \mathrm{GL}_{a_i - a_{i-1}}(\mathbf{F}_q). \end{aligned} \tag{2.4}$$

In particular, if we take $A = N = \{0, 1, 2, \ldots, n\}$, then $P_N = B$ is the standard Borel subgroup of upper triangular matrices, U_N is a maximal unipotent subgroup of G, and the group $M_N = D \simeq (\mathrm{GL}_1(F_q))^n \simeq (\mathbf{F}_q^\times)^n$ is the group of diagonal matrices in G.

If we take $A = \{0 < m < n\}$, then $P_A = P_m$ is a maximal parabolic subgroup of G. The unipotent radical U_m of P_m is abelian and isomorphic to $\mathrm{Hom}(V'_{n-m}, V_m)$, and the Levi component of P_m is isomorphic to $\mathrm{GL}(V_m) \times \mathrm{GL}(V'_{n-m})$.

Recall that two standard parabolics P_A and $P_{A'}$ are called *associate* if M_A and $M_{A'}$ are conjugate in G. From (2.3)(iii), we see that this is so if and only if the sets of numbers $\{a_i - a_{i-1}\}$ and $\{a'_j - a'_{j-1}\}$, counted with multiplicities, are the same. Observe that the collection of numbers $\{a_i - a_{i-1}\}$ forms a partition of n which we call the *partition associated to A*. Thus, P_A and $P_{A'}$ are associate if and only if their associated partitions are equal.

Let W denote the group of n by n permutation matrices. Any element of W simply permutes the coordiates of a vector in V_n. We may identify W in a canonical way with the symmetric group S_n of permutations of $\{1, 2, \ldots, n\}$. We may also identify W with the Weyl group $N_G(D)/D$ of D, the diagonal matrices in G.

For $A = \{0 = a_0 < a_1 < \cdots < a_l = n\}$ as above, a nested sequence $\{X_i\}$ of subspaces of V_n, such that $X_i \subseteq X_{i+1}$ and $\dim X_i = a_i$, is called a *flag of type A*. Every flag of type A may be moved by some $g \in G$ into the standard flag $\mathcal{F}_A$ of type A, so the space of flags of type A may be identified with the homogeneous space G/P_A, which is called the *flag variety of type A*.

For another set $A' = \{0 = a'_0 < a_1 < \cdots < a'_k = n\}$, the orbits of $P_{A'}$ in G/P_A may be identified with the double coset space $P'_A \backslash G/P_A$. According to

Bruhat's lemma [**Bb**, p. 25], we have

$$(2.5)\qquad \text{(a)}\quad B\backslash G/B \simeq W, \quad \text{or} \qquad \text{(b)}\quad G = \bigcup_{w\in W} BwB,$$

where the union is disjoint.

More generally, given a parabolic subgroup P_A, set

$$(2.6)\qquad W_A = W \cap P_A.$$

Then

$$(2.7)\qquad P_{A'}\backslash G/P_A \simeq W_{A'}\backslash W/W_A.$$

3. The representation $\mathrm{Ind}_B^G 1$ and the Hecke algebra $\mathscr{H}(G//B)$. One can begin to approach the study of the irreducible representations of G through consideration of the induced representations

$$(3.1)\qquad \rho_A = \mathrm{Ind}_{P_A}^G 1,$$

where 1 denotes the trivial representation of P_A. It is well known from general theory [**CR**] that the space $\mathrm{Hom}_G(\rho_A, \rho_{A'})$ of intertwining maps from ρ_A to $\rho_{A'}$ may be identified with the space of complex valued functions on G such that

$$(3.2)\qquad f(p_1 g p_2) = f(g), \qquad p_1 \in P_A,\ g \in G,\ p_2 \in P_{A'}.$$

This space has an obvious basis—the characteristic functions of the $(P_A, P_{A'})$ double cosets. Using (2.7), we see that

$$(3.3)\qquad \dim \mathrm{Hom}_G(\rho_A, \rho_{A'}) = \#(W_A\backslash W/W_{A'}).$$

If $A = A'$, then $\mathrm{Hom}(\rho_A, \rho_A) = \mathrm{End}_G(\rho_A)$ is an algebra, and, under the above identification, the algebra structure is convolution of functions on G.

In the particular case $P_A = B$ $(A = N)$, the algebra $\mathrm{End}_G(\rho_B)$ may be identified with the algebra $\mathscr{H}(G//B)$ of B bi-invariant functions on G. In this section, we give a fairly complete description of $\mathscr{H}(G//B)$. This description is adapted from [**I**]. For a general semisimple group $\underline{G}$ over a finite field, with Borel subgroup $\underline{B}$ and Weyl group $\underline{W}$, it is known that $\dim \mathscr{H}(\underline{G}//\underline{B}) = \#\underline{W}$, and, in fact, $\mathscr{H}(\underline{G}//\underline{B})$ is isomorphic to the group algebra of $\underline{W}$ [**I**]. For our group $G = \mathrm{GL}_n$, this is proved in Appendix 2.

Now, let s_i be the elementary transposition in W $(\simeq S_n)$ which exchanges i and $i+1$. Let $l(\cdot)$ denote the standard length function on W. Thus, for $w \in W$, the length $l(w)$ is the number of pairs (a, b) in $N \times N$ whose natural order is reversed by w. Alternatively, $l(w)$ is the smallest length of a product of elementary transpositions expressing w.

We have the formula

$$(3.4)\qquad \#(BwB) = \#(B)q^{l(w)}, \qquad w \in W,$$

for the cardinality of a cell of the Bruhat decomposition of G ((2.5)(b)). Let $f_w = 1/\#(B)$ times the characteristic function of BwB.

If we take counting measure on G so that each group element has mass equal to 1, then the functions f_w form an orthogonal basis for $\mathscr{H}(G//B)$ considered as a subspace of $L^2(G)$. Moreover, f_1 is idempotent and is the identity for the convolution structure on $\mathscr{H}(G//B)$.

A basic fact about multiplication in $\mathscr{H}(G//B)$ is expressed by the relation

(3.5) $$f_{w_1} * f_{w_2} = f_{w_1 w_2} \quad \text{if } l(w_1) + l(w_2) = l(w_1 w_2).$$

In particular, if $w = s_{i_1} s_{i_2} \cdots s_{i_l}$, where $l = l(w)$ and the product is a shortest possible expression for w in terms of the elementary transpositions s_i, then

(3.6) $$f_w = f_{s_{i_1}} * f_{s_{i_2}} * \cdots * f_{s_{i_l}}.$$

Thus, the functions f_{s_i} generate the Hecke algebra $\mathscr{H}(G//B)$.

Equation (3.6) also implies some relations among the f_{s_i} stemming from relations among the elementary transpositions s_i. For example, if $|i - j| > 1$, then $s_i s_j = s_j s_i$, whence

(3.7) $$f_{s_i} * f_{s_j} = f_{s_j} * f_{s_i}, \qquad |i - j| > 1.$$

The transpositions s_i and s_{i+1} generate an S_3 (symmetric group on 3 letters) in W, and it is easy to see that $s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1}$ is the third element of order 2 in this S_3. Hence,

(3.8) $$f_{s_i} * f_{s_{i+1}} * f_{s_i} = f_{s_{i+1}} * f_{s_i} * f_{s_{i+1}}.$$

To fully explicate the multiplication in $\mathscr{H}(G//B)$, we must also consider $f_{w_1} * f_{w_2}$ when $l(w_1 w_2) < l(w_1) + l(w_2)$. To compute this, it is enough to consider $f_{s_i} * f_{s_i}$. It is easy to show that

(3.9) $$f_{s_i} * f_{s_i} = q f_1 + (q - 1) f_{s_i}.$$

It is demonstrated in [I] (as a direct consequence of the analogous fact for W) that $\mathscr{H}(G//B)$ is the algebra generated by the functions f_{s_i} subject to relations (3.7)–(3.9).

By taking advantage of relations (3.3) and the representation theory of the symmetric group, one can completely decompose the representation $\operatorname{Ind}_B^G 1$ and show that $\mathscr{H}(G//B)$ is isomorphic to the group algebra of W. Since this analysis takes us slightly outside our main story line, we present it in Appendix 2.

4. The philosophy of cusp forms. Let $\sigma \in \hat{G}$ and consider a standard parabolic subgroup P_A of G. Suppose that σ contains a vector invariant under U_A, the unipotent radical of P_A. Then Frobenius reciprocity implies that there is a representation σ_1 of P_A, trivial on U_A, such that σ is a subrepresentation of $\operatorname{Ind}_{P_A}^G \sigma_1$.

Since a representation of P_A, trivial on U_A, is a representation of M_A, and M_A is a product of GL_m's for $m < n$ (2.4), the problem of determining the possibilities for σ_1 (i.e., determining $(M_A)^\wedge$) is presumably easier than that of determining $\hat{G}$. Thus, the problem of determining all $\sigma \in \hat{G}$ with U_A fixed vectors is reduced to the problems of determining $(M_A)^\wedge$ and decomposing induced representations.

Evidently, we have here an inductive procedure for determining $\hat{G}$, the building blocks of which are those representations for which no such reduction is possible —that is, those representations σ which contain no U_A-invariant vectors for any A. Following Harish-Chandra, we call such representations *cuspidal*.

Now, the problem of determining $\hat{G}$ may be broken into two parts:

(4.1) (i) Determine the cuspidal representations of $\mathrm{GL}_m(\mathbf{F}_q)$ for any $m \leqslant n$.

(ii) Decompose the representations $\mathrm{Ind}_{P_A}^G \sigma_1$, where σ_1 is a cuspidal representation of $M_A \simeq P_A/U_A$.

This program is known as "the philosophy of cusp forms". Problem (4.1)(ii) has a quite satisfactory answer which was described by Harish-Chandra (for general semisimple groups) in **[HC1]** and which we shall review here. Problem (4.1)(i) has proved more difficult. The solution is provided in the papers of Green (for GL_n), Deligne-Lusztig, and Lusztig mentioned in the introduction to this chapter.

In approaching (4.1)(ii), we first recall some general facts about induced representations **[CR]**. Let G be a group with subgroups H_1 and H_2. For $i = 1, 2$, let τ_i be a representation of H_i on V_i. Set $\rho_i = \mathrm{Ind}_{H_i}^G \tau_i$, $i = 1, 2$. Then **[CR]** the space $\mathrm{End}_G(\rho_1, \rho_2)$ of intertwining operators between ρ_1 and ρ_2 can be canonically identified with the space of functions $f\colon G \to \mathrm{Hom}(V_1, V_2)$ satisfying

$$f(h_2 g h_1) = \tau_2(h_2) f(g) \tau_1(h_1), \qquad h_i \in H_i,\ g \in G. \tag{4.2}$$

Evidently $f(g)$ determines f on the whole double coset $H_2 g H_1$. We denote by $I(g, \rho_1, \rho_2)$ the dimension of the space of functions satisfying (4.2) and supported on $H_2 g H_1$. We say that g *intertwines* ρ_1 and ρ_2 if $I(g, \rho_1, \rho_2) > 0$.

THEOREM 4.1 (HARISH-CHANDRA). *Let P_A and $P_{A'}$ be standard parabolic subgroups of G, and let σ_1 and σ_1' be cuspidal representations of P_A/U_A and $P_{A'}/U_{A'}$ respectively.*

(i) *An element $w \in W$ intertwines σ_1 and σ_1' only if $w M_A w^{-1} = M_{A'}$.*

(ii) *Let w satisfy* (i) *and, for $m \in M_A$, write* $\mathrm{Ad}^* w(\sigma_1')(n) = \sigma_1'(wmw^{-1})$. *Then w intertwines σ_1 and σ_1' if and only if* $\mathrm{Ad}^* w(\sigma_1')$ *is equivalent to* σ_1.

(iii) *If we intertwines σ_1 and σ_1', then* $I(w, \sigma_1, \sigma_1') = 1$.

PROOF. Let $A = \{0 = a_0 < a_1 < \cdots < a_l = n\}$ and $A' = \{0 = a_0' < a_1' < \cdots < a_k' = n\}$. Set $Y_i = V_{a_i} \cap V'_{n-a_{i-1}}$, and $Y_j' = V_{a_j'} \cap V'_{n-a_{j-1}'}$. Then $wM_Aw^{-1} = M_{A'}$ if and only if the collections $\{w(Y_i)\}$ and $\{Y_j'\}$ of subspaces of V_n are the same. Suppose this is not the case. Then we can find i and j such that

$$w(Y_i) = \left(w(Y_i) \cap V_{a_j'}\right) \oplus \left(w(Y_i) \cap V'_{a_j'}\right),$$

where neither summand is trivial. In this case, $\mathrm{GL}(Y_i) \cap (w^{-1} U_{a_j} w)$ is the (nontrivial) unipotent radical of the parabolic subgroup of $\mathrm{GL}(Y_i)$ which preserves $w^{-1}(w(Y_i) \cap V_{a_j'})$. Since σ_1' is trivial on $U_{a_j'}$, but σ_1 contains no fixed vectors for $\mathrm{GL}(Y_i) \cap (w^{-1} U_{a_j} w)$, it is not possible for w to intertwine σ_1 and σ_1'. This proves (i), which is the main statement of the theorem.

Statement (ii) now follows more or less immediately from equation (4.2). Thus, for $m \in M_A$ and f satisfying (4.2) with $\tau_1 = \sigma_1$ and $\tau_2 = \sigma_1'$, we have

$$f(w) = f(wmw^{-1}wm^{-1}) = \mathrm{Ad}^* w(\sigma_1')(m) f(w) \sigma_1(m^{-1}).$$

This implies that $f(w) \in \mathrm{Hom}_{M_A}(\sigma_1, \mathrm{Ad}^* w(\sigma_1'))$, and, if $f(w)$ is nonzero, it must be an isomorphism by Schur's lemma, since σ_1 and $\mathrm{Ad}^* w(\sigma_1')$ are irreducible representations of M_A.

Statement (iii) also follows from Schur's lemma. □

Corollary 4.2. *If P_A and $P_{A'}$ are not associate, then $\mathrm{Ind}_{P_A}^G \sigma_1$ and $\mathrm{Ind}_{P_{A'}}^G \sigma_2$ have no components in common.*

Proof. This is immediate from Theorem 4.1(i). □

Corollary 4.3. *If P_A and $P_{A'}$ are associate, then $\mathrm{Ind}_{P_A}^G \sigma_1$ and $\mathrm{Ind}_{P_{A'}}^G \sigma_1'$ are equivalent, or they have no components in common.*

Proof. Let $A = \{0 = a_0 < a_1 < \cdots < a_l = n\}$. Fix an index j and define $A' = \{0 = a_0' < a_1' < \cdots < a_l' = n\}$ by

$$a_i' = a_i \quad \text{if } i \neq j; \qquad a_j' = a_{j-1} + a_{j+1} - a_j.$$

Choose $w \in W$ such that w acts as the identity on $V_{a_{j-1}}$ and on $V'_{a_{j+1}}$, and such that $w(Y_j) = Y'_{j+1}$ and $w(Y_{j+1}) = Y_j'$. This is possible by the construction of A'. Now, given a cuspidal representation σ_1 of $P_A/U_A = M_A$, define σ_1' on $M_{A'}$ by

$$\sigma_1'(wmw^{-1}) = \sigma_1(m), \qquad m \in M_A.$$

Extend σ_1' to $P_{A'}$ by letting it be trivial on $U_{A'}$. To prove the corollary, it will suffice to prove that $\mathrm{Ind}_{P_A}^G \sigma_1$ and $\mathrm{Ind}_{P_{A'}}^G \sigma_1'$ are equivalent.

Define $\tilde{A} = \{0 = \tilde{a}_0 < \tilde{a}_1 < \cdots < \tilde{a}_{l-1} = n\}$ by the rule

$$\tilde{a}_i = a_i, \quad i < j; \qquad \tilde{a}_i = a_{i+1}, \quad i \geqslant j.$$

Then, $P_{\tilde{A}}$ is a parabolic subgroup containing P_A and $P_{A'}$, and, it is enough to show that $\mathrm{Ind}_{P_A}^{P_{\tilde{A}}} \sigma_1$ and $\mathrm{Ind}_{P_{A'}}^{P_{\tilde{A}}} \sigma_1'$ are equivalent. Thus, it suffices to consider the case $l = 2$; that is, P_A (and hence $P_{A'}$) are maximal parabolic subgroups.

There are two possibilities. Either $\mathrm{Ind}_{P_A}^G \sigma_1$ is irreducible, or $P_A = P_{A'}$ and $\mathrm{Ad}^* w(\sigma_1) = \sigma_1$. In the first case, $\mathrm{Ind}_{P_{A'}}^G \sigma_1'$ is also irreducible, and w intertwines the two representations, which are therefore equivalent. In the second case, Theorem 4.1(iii) tells us that

$$\begin{aligned} \dim \mathrm{End}_G\left(\mathrm{Ind}_{P_A}^G \sigma_1, \mathrm{Ind}_{P_A}^G \sigma_1\right) &= \dim \mathrm{Hom}_G\left(\mathrm{Ind}_{P_A}^G \sigma_1, \mathrm{Ind}_{P_{A'}}^G \sigma_1'\right) \\ &= \dim \mathrm{End}_G\left(\mathrm{Ind}_{P_{A'}}^G \sigma_1', \mathrm{Ind}_{P_{A'}}^G \sigma_1'\right) = 2. \end{aligned}$$

Hence, each of $\mathrm{Ind}_{P_A}^G \sigma_1$ and $\mathrm{Ind}_{P_{A'}}^G \sigma_1'$ consists of a direct sum of the same two irreducible representations, so the two induced representations are equivalent. □

From Theorem 4.1 and its corollaries, we see that the process of forming induced representations from cuspidal representations of parabolic subgroups divides $\hat{G}$ into disjoint subsets parametrized by (associate classes of) Levi components of parabolic subgroups. In the case of $G = \mathrm{GL}_n$, these associate classes of parabolics are parametrized in turn by partitions of n.

The next step in the philosophy of cusp forms is to determine the irreducible components of the induced representations.

Consider then a parabolic subgroup P_A and a cuspidal representation σ_1 of M_A. Since P_A may be replaced by any associate parabolic subgroup, we shall take $A = \{0 = a_0 < a_1 < a_2 < \cdots < a_l = n\}$ so that the members $\lambda_i = a_i - a_{i-1}$ of the associated partition form a nonincreasing sequence. Of course, we may have certain of the λ_i equal to each other. Suppose j_k are the indices where the λ_i change value. Thus

$$\lambda_{j_k} > \lambda_{j_k+1} = \lambda_{j_k+2} = \cdots = \lambda_{j_{k+1}} > \lambda_{j_{k+1}+1},$$

and so forth. Define a partition $\{\mu_k\}$ of n by

$$\mu_k = (j_{k+1} - j_k)\lambda_{j_k+1},$$

and define an increasing sequence

$$\tilde{A} = \{0 = \tilde{a}_0 < \tilde{a}_1 < \tilde{a}_2 < \cdots < \tilde{a}_m = n\} \quad \text{with } \tilde{a}_r = \sum_{k \leq r} \mu_k.$$

By Theorem 4.1, any $w \in W$ which intertwines σ_1 with itself must normalize M_A. By the construction of $\tilde{A}$, we see therefore that w must belong to $P_{\tilde{A}}$. It then follows from general theory that each irreducible component of the induced representation $\mathrm{Ind}_{P_A}^{P_{\tilde{A}}} \sigma_1$ induces irreducibly to G. Thus, it will suffice to describe $\mathrm{Ind}_{P_A}^{P_{\tilde{A}}} \sigma_1$.

Since σ_1 is trivial on $U_A \supseteq U_{\tilde{A}}$, the induced representation $\mathrm{Ind}_{P_A}^{P_{\tilde{A}}} \sigma_1$ will be trivial on $U_{\tilde{A}}$, and therefore will effectively be a representation of $M_{\tilde{A}}$. But $M_{\tilde{A}}$ is a product

$$M_{\tilde{A}} \simeq \prod_k \mathrm{GL}_{\mu_k}(\mathbf{F}_q).$$

Furthermore,

$$P_A \cap M_{\tilde{A}} \simeq \prod_k \left(\mathrm{GL}_{\mu_k}(\mathbf{F}_q) \cap P_A\right).$$

The representation σ_1 of $P_A \cap M_{\tilde{A}}$ will be a tensor product of representations $(\sigma_1)_k$ of the factors $\mathrm{GL}_{\mu_k}(\mathbf{F}_q) \cap P_A$, and the induced representation will be a tensor product of the induced representations

$$\mathrm{Ind}_{\mathrm{GL}_{\mu_k}(\mathbf{F}_q) \cap P_A}^{\mathrm{GL}_{\mu_k}(\mathbf{F}_q)} (\sigma_1)_k. \tag{4.3}$$

Hence, to determine the structure of $\mathrm{Ind}_{P_A}^{P_{\tilde{A}}} \sigma_1$, it suffices to determine the structure of representations of the type (4.3). In other words, it suffices to consider the case $A = \{0 = a_0 < a_1 < a_2 < \cdots < a_l = n\}$ in which all the steps $a_i - a_{i-1}$ have a common value, say λ, so that $n = l\lambda$.

For such an A as just specified, we have

$$M_A \simeq (\mathrm{GL}_\lambda(\mathbf{F}_q))^l,$$

and a cuspidal representation σ_1 of M_A will be a tensor product

$$\sigma_1 \simeq \bigotimes_{i=1}^{l} \tau_i,$$

where each τ_i is a cuspidal representation of $\mathrm{GL}_\lambda(\mathbf{F}_q)$. Arguing again as above, and appealing to Theorem 4.1(ii), we can further reduce to the case in which all the τ_i are equivalent to each other. This basic case will be investigated in the next section.

5. Intertwining algebras for induced representations. Suppose n (as in V_n and $\mathrm{GL}_n(\mathbf{F}_q) = G$) is composite, say $n = \lambda m$. We divide an $n \times n$ matrix into an $m \times m$ array of $\lambda \times \lambda$ blocks. Let $\tilde{G}$ denote the subgroup of elements of G which, on each $\lambda \times \lambda$ block, contain a multiple of the identity matrix. Then $\tilde{G}$ is isomorphic to $\mathrm{GL}_m(\mathbf{F}_q)$. Let $\tilde{W} = W \cap \tilde{G}$. Let P be the parabolic subgroup of G consisting of matrices whose entries in the subdiagonal blocks are all zero. Thus, $P = P_A$, where $A = \{0 = a_0 < a_1 = \lambda < a_2 = 2\lambda < \cdots < a_m = m\lambda = n\}$. If M is the Levi component of P, then

$$M \simeq (\mathrm{GL}_\lambda(\mathbf{F}_q))^m.$$

Let τ be an irreducible cuspidal representation of $\mathrm{GL}_\lambda(\mathbf{F}_q)$ on a Hilbert space $\mathcal{Y}$, and let σ_1 be the representation of M obtained by taking the m-fold tensor product of τ acting on $\otimes^m \mathcal{Y}$, the m-fold tensor power of $\mathcal{Y}$. According to Theorem 4.1(i), the elements of G which intertwine σ_1 with itself belong to the set $P\tilde{W}P$. In fact, we can write these elements explicitly.

We identify $\tilde{W}$ with the symmetric group S_m by identifying $w \in \tilde{W}$ with its action by conjugation on the diagonal blocks labeled according to their position along the diagonal. Let $J(w)$ be the operator on $\otimes^m \mathcal{Y}$ which permutes the factors of an m-fold tensor according to the permutation w. Thus,

$$J(w)(y_1 \otimes y_2 \otimes \cdots \otimes y_m) = y_{w(1)} \otimes y_{w(2)} \otimes \cdots \otimes y_{w(m)}.$$

Then the intertwining algebra $\mathrm{End}_G(\mathrm{Ind}_P^G \sigma_1)$ has a basis consisting of the functions f_w defined by

$$f(p_1 w p_2) = (\# P)^{-1} \sigma_1(p_1) J(w) \sigma_1(p_2), \qquad p_1, p_2 \in P; \tag{5.1}$$
$$f(g) = 0 \quad \text{if } g \notin PwP.$$

We now compute the products of the functions f_w defined by (5.1). Let l be the length function on W defined in §3. Since each $w \in W$ normalizes the group D of diagonal matrices in G, equation (3.4) is equivalent to

$$\#(U_B/(U_B \cap wU_Bw^{-1})) = q^{l(w)}, \qquad w \in W. \tag{5.2}$$

If $w \in \tilde{W}$, then w normalizes M and $M \cap U_B$. Thus, we may conclude

$$\#(PwP) = q^{l(w)}\#(P), \qquad w \in \tilde{W}. \tag{5.3}$$

REMARK. Since $\tilde{W}$ has been identified with S_m, it too possesses a length function which we denote by $\tilde{l}$. It is easy to check that

$$l|\tilde{W} = \lambda^2\tilde{l}. \tag{5.4}$$

Take $w, w' \in \tilde{W}$. Suppose first that $l(ww') = l(w) + l(w')$, and consider the multiplication map

$$(PwP) \times (Pw'P) \mapsto Pww'P. \tag{5.5}$$

Clearly, if $xy = z$, then $(xp)(p^{-1}y) = z$ for any $p \in P$. Thus, the fibers of the map (5.5) are of cardinality a multiple of $\#(P)$. From (5.3), we see that, under our assumption, the fibers of the map (5.5) must have cardinality exactly $\#(P)$. Hence, in particular, if $xy = ww'$, then $x = wp$ and $y = p^{-1}w'$. It follows immediately that

$$f_w * f_{w'} = f_{ww'}, \quad w, w' \in \tilde{W}, \qquad l(w) + l(w') = l(ww'). \tag{5.6}$$

Thus, if the elementary transpositions of $\tilde{W}$ are denoted by $\tilde{s}_i$, we see that the $f_{\tilde{s}_i}$ generate $\mathrm{End}_G(\mathrm{Ind}_P^G \sigma_1)$, and further, we have the relations

$$\text{(5.7)} \qquad \begin{aligned} &\text{(i)} \quad f_{\tilde{s}_i} * f_{\tilde{s}_j} = f_{\tilde{s}_j} * f_{\tilde{s}_i}, \qquad |i-j| > 1, \\ &\text{(ii)} \quad f_{\tilde{s}_i} * f_{\tilde{s}_{i+1}} * f_{\tilde{s}_i} = f_{\tilde{s}_{i+1}} * f_{\tilde{s}_i} * f_{\tilde{s}_{i+1}}, \end{aligned}$$

analogous to relations (3.7) and (3.8) in $\mathscr{H}(G//B)$.

Next, we consider the convolution product $f_{\tilde{s}_i} * f_{\tilde{s}_i}$. The element $\tilde{s}_i$, together with P, will generate a parabolic subgroup P' which is minimal with respect to the property of strictly containing P. It is easy to see that among the double cosets PwP, $w \in W$, contained in P', the only ones with $w \in \tilde{W}$ are P and $P\tilde{s}_iP$. Hence, to determine $f_{\tilde{s}_i} * f_{\tilde{s}_i}$, it suffices to determine its value at 1 and at $\tilde{s}_i$.

If $x, y \in P\tilde{s}_iP$, and $xy = 1$, then clearly $y = x^{-1}$, and, if $x = p_1\tilde{s}_ip_2$, then $y = p_2^{-1}\tilde{s}_ip_i^{-1}$. From (5.1) and (5.3), we see that

$$f_{\tilde{s}_i} * f_{\tilde{s}_i}(1) = \#(P)^{-2}\#(P\tilde{s}_iP) = q^{l(\tilde{s}_i)}\#(P)^{-1}. \tag{5.8}$$

Next, consider the equation

$$xy = \tilde{s}_i, \qquad x, y \in P\tilde{s}_iP. \tag{5.9}$$

In studying this equation, we may as well assume that $P' = G$, so that $m = 2$. In this case, we write $\tilde{s}_i = \tilde{s}$, and we may take x and y to have the form

$$x = u_1\tilde{s}p_1, \qquad y = p_2\tilde{s}u_2,$$

with $u_1, u_2 \in U$, the unipotent radical of P, and $p_1, p_2 \in P$. If we write $p = p_1p_2$, then equation (5.9) becomes

$$u_1\tilde{s}p\tilde{s}u_2 = \tilde{s}. \tag{5.10}$$

Equation (5.10) is equivalent to the matrix equation

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} b & d \\ 0 & c \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & e \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{5.11}$$

where the entries are λ by λ matrices. Straightforward multiplication, followed by a comparison of entries, leads to the conclusion that a must be invertible, and

$$b = a^{-1}, \quad c = -a, \quad d = 1, \quad e = -a^{-1}. \tag{5.12}$$

Since $m = 2$, the space of the representation σ_1 with which we are now dealing is $\mathscr{Y} \otimes \mathscr{Y}$, where $\mathscr{Y}$ is the space of the representation τ of $\mathrm{GL}_\lambda(\mathbf{F}_q) = G_1$. Consider the operator

$$T = \sum_{g \in G_1} \tau(g) \otimes \tau(-g^{-1}) \tag{5.13}$$

on $\mathscr{Y} \otimes \mathscr{Y}$. One sees easily that

$$(\tau(g_1) \otimes \tau(g_2))T = T(\tau(g_1) \otimes \tau(g_2)), \qquad g_1, g_2 \in G.$$

Thus, T accomplishes the same automorphism of $\sigma_1(M) = \tau \otimes \tau(G_1 \times G_1)$ as does the operator $J(w)$. Schur's lemma then guarantees that T is actually a multiple of $J(w)$. To compute this multiple, it suffices to compute the trace of $TJ(w)$.

Choose $g \in G_1$, and let $\{y_i\}$ be a basis for $\mathscr{Y}$ consisting of eigenvectors for $\tau(g)$, so that $\tau(g)y_i = \alpha_i y_i$. Then $\{y_i \otimes y_j\}$ is a basis for $\mathscr{Y} \otimes \mathscr{Y}$, and we have

$$(\tau(g) \otimes \tau(-g^{-1}))J(w)(y_i \otimes y_j) = \varepsilon(-1)\alpha_j \alpha_i^{-1}(y_j \otimes y_i),$$

where $\varepsilon(-1)$ is the multiple of the identity map of $\mathscr{Y}$ which is equal to $\tau(-1)$. Here, -1 denotes the negative of the identity in $\mathrm{GL}_\lambda(\mathbf{F}_q) = G_1$. If we use the above basis to compute the trace, we obtain

$$\mathrm{tr}\big((\tau(g) \otimes \tau(-g^{-1}))J(w)\big) = \varepsilon(-1)\dim\mathscr{Y}.$$

Thus,

$$\mathrm{tr}\Big(\sum_{g \in G_1} (\tau(g) \otimes \tau(-g^{-1}))J(w)\Big) = \varepsilon(-1)(\dim\mathscr{Y})\#(G_1),$$

and

$$\sum_{g \in G_1} (\tau(g) \otimes \tau(-g^{-1})) = \varepsilon(-1)(\dim\mathscr{Y})^{-1}\#(G_1)J(w). \tag{5.14}$$

Combining equations (5.8)–(5.14), we find that

$$f_{\tilde{s}_i} * f_{\tilde{s}_i} = q^{l(\tilde{s}_i)}f_1 + \varepsilon(-1)(\dim\mathscr{Y})^{-1}\#(G_1)f_{\tilde{s}_i}. \tag{5.15}$$

From Appendix 3, we see that

$$(\dim\mathscr{Y})^{-1}\#(G_1) = (q^\lambda - 1)q^{(\lambda(\lambda-1))/2}. \tag{5.16}$$

Now set

$$f^0_{\tilde{s}_i} = \varepsilon(-1)q^{-(\lambda(\lambda-1))/2}f_{\tilde{s}_i}; \qquad f^0_1 = f_1. \tag{5.17}$$

Then, using (5.4), (5.15), and the fact that $\tilde{l}(\tilde{s}_i) = 1$, we have

$$f^0_{\tilde{s}_i} * f^0_{\tilde{s}_i} = q^\lambda f^0_1 + (q^\lambda - 1)f^0_{\tilde{s}_i}. \tag{5.18}$$