Canad. J. Math. Vol. **59** (5), 2007 pp. 1050–1068

On the Restriction to $\mathcal{D}^* \times \mathcal{D}^*$ of Representations of p-adic $GL_2(\mathcal{D})$

In memory of my mother Shantha Anantharam

A. Raghuram

Abstract. Let $\mathcal{D}$ be a division algebra over a nonarchimedean local field. Given an irreducible representation π of $GL_2(\mathcal{D})$, we describe its restriction to the diagonal subgroup $\mathcal{D}^* \times \mathcal{D}^*$. The description is in terms of the structure of the twisted Jacquet module of the representation π. The proof involves Kirillov theory that we have developed earlier in joint work with Dipendra Prasad. The main result on restriction also shows that π is $\mathcal{D}^* \times \mathcal{D}^*$-distinguished if and only if π admits a Shalika model. We further prove that if $\mathcal{D}$ is a quaternion division algebra then the twisted Jacquet module is multiplicity-free by proving an appropriate theorem on invariant distributions; this then proves a multiplicity-one theorem on the restriction to $\mathcal{D}^* \times \mathcal{D}^*$ in the quaternionic case.

1 Introduction and Statements of Theorems

Let F denote a nonarchimedean local field and let $\mathcal{D}$ stand for a central division algebra over F. This is the third article in our study [15, 17] of representations of the group $G = GL_2(\mathcal{D})$. Let π be an irreducible admissible infinite dimensional representation of G. The main aim of this paper is to describe the restriction of π to the diagonal subgroup $M = \mathcal{D}^* \times \mathcal{D}^*$ of G.

To state the main theorem, which describes the restriction to the subgroup M, we need to introduce some notations. Let P denote the standard minimal parabolic subgroup of upper triangular matrices in G. Let N be the unipotent radical of P. Then N is the subgroup of upper triangular matrices with 1's on the diagonal and $N \simeq \mathcal{D}^+$. Fix a nontrivial additive character ψ_F of the base field F. Let ψ be the character of $\mathcal{D}$, defined as $\psi(x) = \psi_F(\mathrm{Trd}_{\mathcal{D}/F}(x))$ for all $x \in \mathcal{D}$, and where $\mathrm{Trd}_{\mathcal{D}/F}$ is the reduced trace map from $\mathcal{D}$ to F. We let ψ also denote the corresponding character of N. If (π, V) is an irreducible admissible infinite dimensional representation of G, then let $V_{N,\psi}$ denote the maximal quotient of V on which N acts via ψ. This space $V_{N,\psi}$ is naturally a representation of $\mathcal{D}^* \simeq \mathrm{stab}_M(\psi)$. This representation is denoted $\pi_{N,\psi}$, and is called the *twisted Jacquet module* of π relative to ψ. In the literature this module is also called the space of degenerate Whittaker models [11].

The main theorem of this paper is the following.

Received by the editors January 21, 2005; revised April 18, 2005.
AMS subject classification: Primary: 22E50; secondary: 22E35, 11S37.

Theorem 1.1 *Let $G = GL_2(\mathcal{D})$. Let π be an irreducible admissible infinite dimensional representation of G. Let τ_1 and τ_2 be two smooth irreducible representations of $\mathcal{D}^*$. Assume that either*

(i) *$\tau = \tau_1 \otimes \tau_2$ does not intertwine with the usual Jacquet module π_N (this includes the case when π is supercuspidal); or*

(ii) *π is irreducibly parabolically induced and π_N is semisimple as an M-module.*

Then the multiplicity with which $\tau_1 \otimes \tau_2$ occurs as a quotient of π restricted to the diagonal subgroup $\mathcal{D}^* \times \mathcal{D}^*$ is equal to the dimension of the space of intertwining operators between $\tau_1 \otimes \tau_2$, now as a representation of $\mathcal{D}^*$, and the twisted Jacquet module $\pi_{N,\psi}$, i.e.,

$$\dim_{\mathbb{C}}(\mathrm{Hom}_{\mathcal{D}^* \times \mathcal{D}^*}(\pi, \tau_1 \otimes \tau_2)) = \dim_{\mathbb{C}}(\mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \tau_1 \otimes \tau_2)).$$

Regarding the assumptions on π and τ, we certainly believe the statement to be true as stated for all π and τ, however, we have only been able to prove it in the cases (i) and (ii) above.

For the remaining cases, we point out at appropriate places what is lacking in the theory developed so far for the group $GL_2(\mathcal{D})$.

Note that for $GL_2(F)$, it has been observed by Waldspurger [21, Lemmas 8, 9] that a character $\chi_1 \otimes \chi_2$ of $F^* \times F^*$ occurs as a quotient of π if and only if $\chi_1\chi_2$ is the central character of π, and in this case, it occurs with multiplicity one. Observe that if $\mathcal{D} = F$, then by multiplicity one for Whittaker models of $GL_2(F)$ we have that $\pi_{N,\psi}$ is one dimensional, and as a module for F^* it is ω_π, the central character of π. Hence the theorem specializes to the above-mentioned result of Waldspurger. Indeed, this may be regarded as a different proof of Waldspurger's results.

In general, the structure of $\pi_{N,\psi}$ as a $\mathcal{D}^*$-module is rather mysterious. If $\mathcal{D}$ is quaternion, then there is a conjectural description of this module [13]. This paper along with the results of [15–17] add to the heuristic that the structure of the twisted Jacquet module $\pi_{N,\psi}$ substantially governs the structure of π.

The first ingredient of the proof is Kirillov theory for $G = GL_2(\mathcal{D})$, as developed in an earlier paper [15]. We need only part of the main theorem of that paper which gives a short exact sequence of P-modules for any irreducible representation π of G. (See Theorem 2.1 below.) The main idea is to apply the functor $\mathrm{Hom}_M(-, \tau_1 \otimes \tau_2)$ to this short exact sequence from Kirillov theory to get a certain long exact sequence; the hard work is to analyze the relevant part of this long exact sequence.

This brings us to the second ingredient in our proof, namely certain Ext computations. We need, in particular, an Ext^1 calculation for certain representations of $M = \mathcal{D}^* \times \mathcal{D}^*$. This is done in Section 3.

With these inputs in place we get the proof of Theorem 1.1 when π is supercuspidal or more generally, when π is arbitrary and $\tau_1 \otimes \tau_2$ does not intertwine with the Jacquet module π_N of π. To handle the remaining cases and for applications later in this paper, we need a third ingredient, which is a theorem due to Tadić [20], on reducibility for $GL_2(\mathcal{D})$ and explicit Jacquet module calculations. This is recalled in Theorem 2.2. The proof of Theorem 1.1 is taken up in §4.

We now consider some applications of Theorem 1.1. The first application is toward Shalika models for representations of $GL_2(\mathcal{D})$, which is taken up in §5. It has been shown by Jacquet and Rallis [9] that an irreducible representation π of $GL_{2n}(F)$ has, up to scalars, at most one $GL_n(F) \times GL_n(F)$-distinguishing functional, and as a consequence they show that there is, up to scalars, at most one Shalika functional. (Indeed, this paper and our earlier papers [15–17] stem from a paper of Dipendra Prasad [12] which proves a division algebra version of this theorem of Jacquet and Rallis.) In §5 we prove the following.

Theorem 5.3 *Let $G = GL_2(\mathcal{D})$ and let $M = \mathcal{D}^* \times \mathcal{D}^*$ be the diagonal subgroup of G. Let π be an irreducible admissible infinite dimensional representation of G. Then π is M-distinguished if and only if π admits a Shalika model.*

To put this theorem into perspective, see [9, 15]. The proof follows from applying Theorem 1.1 with τ the trivial representation of M. However, since we do not have a proof of Theorem 1.1 in all cases, we need to finesse this proof in the *bad cases*, and Theorem 5.3 is true unconditionally.

The second application is toward a multiplicity-one result, in the special case when $\mathcal{D}$ is the quaternion division algebra over F; this is taken up in §6. For the rest of the introduction we assume that $\mathcal{D}$ is quaternion. A result of Dipendra Prasad [13, 14] says that $\pi_{N,\psi}$ is multiplicity-free as a $\mathcal{D}^*$-module. We state this as Theorem 6.4 and for the reader's convenience sketch a proof of this theorem. The proof boils down to proving a certain result on invariant distributions which is stated as Theorem 6.5. The proof of this result on invariant distributions heavily uses Bernstein's localization technique. We would like to emphasize that the proof also heavily uses the fact that $\mathcal{D}$ is indeed a quaternion division algebra. Once one has that $\pi_{N,\psi}$ is multiplicity-free, then Theorem 1.1 can be used to prove the following multiplicity-one theorem.

Theorem 6.2 *Let $G = GL_2(\mathcal{D})$ where $\mathcal{D}$ is the quaternion division algebra with center F. Let $M = \mathcal{D}^* \times \mathcal{D}^*$ be the diagonal subgroup. Let π be an irreducible admissible representation of G. Let τ be any irreducible representation of M whose restriction to the diagonal $\mathcal{D}^*$ is irreducible. Then τ occurs as a quotient of the restriction of π to M with multiplicity at most one.*

Again, since Theorem 1.1 is not available in all cases, we need to finesse this proof, and Theorem 6.2 is unconditionally true . It would be interesting to see if the above is true for representations of $GL_4(F)$ restricted to $GL_2(F) \times GL_2(F)$.

2 Preliminaries and Notation

We continue with the notation in the introduction. We also use the notation from [15, §1.2]. The following theorem is one of the main results proved in [15].

Theorem 2.1 (Kirillov Theory) *Let π be an irreducible admissible infinite dimensional representation of G. Let $\pi_{N,\psi}$ be the twisted Jacquet module of π,* i.e., *the maximal quotient of π on which N acts via ψ. It is a module for $\mathcal{D}^*$ embedded diagonally in M. Let π_N denote the usual Jacquet module of π,* i.e., *the maximal quotient of π on which N acts trivially; it is an M-module. We have an exact sequence of P-modules:*

$$0 \to C_c^\infty(\mathcal{D}^*, \pi_{N,\psi}) \to \pi \to \pi_N \to 0.$$

Some remarks are in order. The fact that $\pi_{N,\psi}$ is a module for the diagonal $\mathcal{D}^*$ follows from the observation that the stabilizer inside M of the character ψ is this $\mathcal{D}^*$. It is known that $\pi_{N,\psi}$ is finite dimensional (see [11, 16]). The action of P on $C_c^\infty(\mathcal{D}^*, \pi_{N,\psi})$ is given as in [15, p. 21]. We let S stand for the *Shalika subgroup* of G. The P-module $C_c^\infty(\mathcal{D}^*, \pi_{N,\psi})$ is naturally isomorphic to $\mathrm{ind}_S^P(\pi_{N,\psi} \otimes \psi)$, where ind_S^P denotes compact and un-normalized induction from S to P.

We will need precise information about the Jacquet module of a representation π, especially when π is a subquotient of a parabolically induced representation. Toward this, let (π, V) now be any smooth representation of G. Let $V(N)$ denote the span of $\{v - \pi(n)v \mid n \in N,\ v \in V\}$. Then $V(N)$ is stable under M. Let $V_N = V/V(N)$. The natural action of M on V_N has been denoted π_N and is the usual un-normalized Jacquet module of π. We let $r_N(\pi)$ denote the normalized Jacquet module, defined as $r_N(\pi) = (|\cdot|^{-1/2} \otimes |\cdot|^{1/2}) \otimes \pi_N$. Let σ_1 and σ_2 be two irreducible representations of $\mathcal{D}^*$. Let $\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)$ denote the normalized parabolically induced representation of G. It is convenient to introduce the following notation. For a representation π of $\mathcal{D}^*$ or $GL_d(F)$ we let $\pi(s)$ stand for $\pi \otimes |\cdot|_F^s$, with the understanding that a character of F^* (such as $|\cdot|_F$) gives a character of $GL_d(F)$ (resp., $\mathcal{D}^*$) via the determinant (resp., the reduced norm). Also from the above normalizations [15, §1.2] we have for all $x \in \mathcal{D}^*$, $|x| = |\mathrm{Nrd}_{\mathcal{D}/F}(x)|_F^d$. Hence, if σ is a representation of $\mathcal{D}^*$, then $\sigma \otimes |\cdot|^{1/2} = \sigma(d/2)$. We have the following sequence of M-modules (see [15]):

$$0 \to \sigma_2(d/2) \otimes \sigma_1(-d/2) \to \mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_N \to \sigma_1(d/2) \otimes \sigma_2(-d/2) \to 0.$$

The normalized version of this exact sequence is

$$0 \to \sigma_2 \otimes \sigma_1 \to r_N(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)) \to \sigma_1 \otimes \sigma_2 \to 0.$$

We also record that the twisted Jacquet module is given by $\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_{N,\psi} = \sigma_1 \otimes \sigma_2$ as $\mathcal{D}^*$-modules. (See [15, Theorem 2.1].)

For Jacquet modules of subquotients of a parabolically induced representation, we record the following theorem due to Tadić [20]. We need some notations to state this theorem. Let σ denote an irreducible representation of $\mathcal{D}^*$. Recall that d denotes the index of $\mathcal{D}$. Let Σ denote the irreducible essentially square integrable representation of $GL_d(F)$ that corresponds to σ. (When F is of characteristic zero, this correspondence is due to Jacquet–Langlands [8] for $d = 2$; it is due to Deligne–Kazhdan–Vignèras [7] and also Rogawski [19] for $d > 2$. When F is of positive characteristic, it is due to Badulescu [1].) Any essentially square integrable Σ, in the notations of Kudla's article [10], is of the form $Q(\Delta)$ for a segment $\Delta = [\rho, \rho(1), \dots, \rho(a-1)]$, where ρ is an irreducible supercuspidal representation of $GL_b(F)$ and $d = ab$. We let $a(\sigma)$ denote this integer a, *i.e.,* it is the length of the segment which determines the Jacquet–Langlands lift of σ. Note that $a(\sigma \otimes \chi) = a(\sigma)$ for any character χ.

Theorem 2.2 (Tadić) *Let σ_1, σ_2 and σ be irreducible representations of $\mathcal{D}^*$. For brevity, let $\sigma_1 \times \sigma_2$ stand for the representation $\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)$. We have*

(i) *$\sigma_1 \times \sigma_2$ is reducible if and only if $\sigma_2 \simeq \sigma_1(\pm a(\sigma_1))$.*

(ii) *The representation* $\sigma(-a(\sigma)/2) \times \sigma(a(\sigma)/2)$ *has a unique irreducible quotient, which we denote by* $\mathrm{St}(\sigma)$*, which is also the unique irreducible essentially square integrable subquotient, whose normalized Jacquet module is given by*

$$r_N(\mathrm{St}(\sigma)) \simeq \sigma(a(\sigma)/2) \otimes \sigma(-a(\sigma)/2).$$

(iii) *The representation* $\sigma(a(\sigma)/2) \times \sigma(-a(\sigma)/2)$ *has a unique irreducible quotient, which we denote by* $\mathrm{Sp}(\sigma)$*, whose normalized Jacquet module is given by*

$$r_N(\mathrm{Sp}(\sigma)) \simeq \sigma(-a(\sigma)/2) \otimes \sigma(a(\sigma)/2).$$

(iv) *The representation* $\sigma \times \sigma(a(\sigma))$ *has two and only two irreducible subquotients both of which occur with multiplicity one.*

In the above theorem, (i) is contained in [20, Lemmas 2.5, 4.2]; (ii) and (iii) are in [20, Proposition 2.7] and (iv) is in [20, Proposition 4.3]. To compare our notations with the notations of Tadić [20], consider the segments $\Delta_1 = \{\sigma(-a(\sigma)/2)\}$ and $\Delta_2 = \{\sigma(a(\sigma)/2)\}$ and $\Delta = \Delta_1 \cup \Delta_2 = \{\sigma(-a(\sigma)/2), \sigma(a(\sigma))\}$. Our *generalized Steinberg* representation $\mathrm{St}(\sigma)$ is $L(\Delta)$ of [20] and similarly our *generalized Speh* representation $\mathrm{Sp}(\sigma)$ is $L(\Delta_1, \Delta_2)$ of [20]. We end this section by adding two remarks based on the above theorem of Tadić. The first remark is regarding functoriality of reducibility points. The second remark is about when an induced representation has a finite (and hence one) dimensional subquotient.

Remark 2.3 Let σ_1 and σ_2 be irreducible representations of $\mathcal{D}^*$ and let $\Sigma_1 = \mathrm{JL}(\sigma_1)$ and $\Sigma_2 = \mathrm{JL}(\sigma_2)$ be the corresponding irreducible representations of $GL_d(F)$. Then we obtain the following from Theorem 2.2 and well-known reducibility theorems of Bernstein and Zelevinskii for GL_n (see [10] for instance).

(i) If $\sigma_1 \times \sigma_2$ is reducible as a representation of $GL_2(\mathcal{D})$, then $\Sigma_1 \times \Sigma_2$ is reducible as a representation of $GL_{2d}(F)$.

(ii) The converse of (i) is not true in general. For example, take σ_1 the trivial character and $\sigma_2 = |\cdot|_F$ for a quaternion division algebra. Then $\Sigma_1 = \mathrm{St}_{GL_2}$ and $\Sigma_2 = \mathrm{St}_{GL_2}(1)$. Then $\sigma_1 \times \sigma_2$ is irreducible by the above theorem of Tadić, since $a(\sigma_1) = 2$. However, $\Sigma_1 \times \Sigma_2$ is reducible [10].

(iii) If Σ_1 and Σ_2 are both supercuspidal, then it is true that $\sigma_1 \times \sigma_2$ is reducible if and only if $\Sigma_1 \times \Sigma_2$ is reducible.

Remark 2.4 Let σ_1 and σ_2 be two irreducible representations of $\mathcal{D}^*$. Then the induced representation $\sigma_1 \times \sigma_2$ admits a one dimensional subquotient if and only if σ_1 and σ_2 are one dimensional and $\sigma_2 = \sigma_1(\pm d)$. This may be seen as follows. If $\sigma_1 \times \sigma_2$ has a one dimensional subquotient, then it must be of the form $\mathrm{Sp}(\sigma)$ (where σ is an appropriate twist of σ_1). By (iii) of the theorem above, both σ_1 and σ_2 have to be one dimensional and $a(\sigma) = d$. Conversely, if σ_1 and σ_2 are one dimensional, then, up to twisting and dualizing, we may assume that $\sigma_1 = |\cdot|_F^{-d/2}$ and $\sigma_2 = |\cdot|_F^{d/2}$. It is easy to see then that the space of constant functions is a one dimensional invariant subspace of the induced representation $\sigma_1 \times \sigma_2$.

3 An Ext^1-Calculation

For the proof of Theorem 1.1, we will need an Ext^1-calculation for the group $M = \mathcal{D}^* \times \mathcal{D}^*$. Our original approach toward this was to prove the much more general Künneth theorem for extensions between representations for a product of two arbitrary p-adic groups from which the required calculation follows as an easy special case. However, for the proof of Theorem 1.1, the referee has sketched a very simple argument, and we elaborate on that in this section. (Our Künneth theorem will appear elsewhere [18].)

The following lemma calculates Ext^1 for just one division algebra and the general case of a product of division algebras, stated as a corollary to the proof of the lemma, follows basically the same argument. If (π, V) is an irreducible representation of a group G, then by the adjoint of π, denoted $\mathrm{Ad}(\pi)$, we mean the representation of G on $\mathrm{End}_{\mathbb{C}}(V)$ given by $g \cdot \phi = \pi(g) \circ \phi \circ \pi(g)^{-1}$ for all $g \in G$ and all $\phi \in \mathrm{End}_{\mathbb{C}}(V)$. It is easy to see that $\mathrm{Ad}(\pi) \simeq \pi^\vee \otimes \pi$ where $\pi^\vee$ is the contragredient of π.

Lemma 3.1 *Let π be an irreducible representation of $\mathcal{D}^*$ for a p-adic division algebra $\mathcal{D}$. Then $\mathrm{Ext}^1_{\mathcal{D}^*}(\pi, \pi) = H^1(\mathcal{D}^*, \mathrm{Ad}(\pi))$ is a one dimensional space. (The right-hand side is continuous group cohomology.)*

Proof We use the usual identification of $\mathrm{Ext}^1_G(\pi, \pi)$ with the set of all short exact sequences

$$0 \to \pi \to \rho \to \pi \to 0$$

modulo Yoneda equivalence. We now analyze what representations ρ appear as above. In terms of block matrices, we can represent $\rho(x)$ for any $x \in D^*$ as $\rho(x) = \begin{bmatrix} \pi(x) & f(x) \\ 0 & \pi(x) \end{bmatrix}$ for some function $f\colon D^* \to \mathrm{End}_{\mathbb{C}}(V)$. (Here V is the representation space of π.) Using $\rho(xy) = \rho(x)\rho(y)$, we get $f(xy) = \pi(x)f(y) + f(x)\pi(y)$. Let $g(x) = f(x)\pi(x)^{-1}$. Then we have $g(xy) = g(x) + \pi(x)g(y)\pi(x)^{-1}$, this being an equation in $\mathrm{End}_{\mathbb{C}}(V)$. This also tells us that $g \in Z^1(D^*, \mathrm{Ad}(\pi))$, *i.e.*, g is a 1-cocycle on D^* with values in $\mathrm{Ad}(\pi)$. It is clear that g is a continuous cocycle.

Now suppose ρ_1 and ρ_2 are two such extensions of π by π. Let g_1 and g_2, respectively, be the associated 1-cocyles. It is easy to see that ρ_1 is Yoneda equivalent to ρ_2 if and only if g_1 and g_2 differ by a 1-coboundary. It is also standard to check that the map which associates to an extension ρ the function g, as above, is a vector space isomorphism.

Let $U = \mathcal{O}^\times$ be the group of units of $\mathcal{D}^*$. Since U is compact, it has vanishing cohomology in nonzero degree. Using the inflation-restriction sequence we get

$$H^1(G, A) \simeq H^1(G/U, A^U)$$

for any G-module A, where A^U is the U-invariants of A. Applying this to the case at hand, we get

$$H^1(\mathcal{D}^*, \mathrm{Ad}(\pi)) \simeq H^1(\mathbb{Z}, (\pi^\vee \otimes \pi)^{\mathcal{O}^\times}).$$

Observe that $(\pi^\vee \otimes \pi)^{\mathcal{O}^\times}$ is a sum of characters for $\mathbb{Z}$, with the trivial character $\mathbf{1}$ showing up exactly once. Noting that $\mathbb{Z}$ has cohomology only with the trivial coefficients, and that $H^1(\mathbb{Z}, \mathbf{1})$ is one dimensional, finishes the proof. ■

Corollary 3.2 *Let F be a p-adic field. Let $\mathcal{D}_1, \ldots, \mathcal{D}_r$ be central division algebras over F. For $1 \leq i \leq r$, let π_i and π_i' be smooth irreducible representations of $\mathcal{D}_i^*$. Let $\pi = \pi_1 \otimes \cdots \otimes \pi_r$ and $\pi' = \pi_1' \otimes \cdots \otimes \pi_r'$ be the corresponding smooth irreducible representations of $G = \mathcal{D}_1^* \times \cdots \times \mathcal{D}_r^*$. Then*

$$\dim_{\mathbb{C}}(\mathrm{Ext}^1_G(\pi, \pi')) = \begin{cases} 0 & \text{if } \pi \neq \pi' \\ r & \text{if } \pi = \pi'. \end{cases}$$

Proof Same proof as Lemma 3.1. Observe that the trivial character occurs in $\pi^\vee \otimes \pi'$ if and only if $\pi = \pi'$. Observe also that $H^1(\mathbb{Z}^r, \mathbf{1})$ has dimension r. ■

Corollary 3.3 *Let $\mathcal{D}$ be a division algebra over F. Let (π_1, W_1) and (π_2, W_2) be two irreducible representations of $\mathcal{D}^*$. Then $\mathrm{Ext}^1_{\mathcal{D}^* \times \mathcal{D}^*}(\pi_1 \otimes \pi_2, \pi_1 \otimes \pi_2)$ is a two dimensional vector space and may be realized as*

$$0 \to \pi_1 \otimes \pi_2 \xrightarrow{i} \begin{bmatrix} \pi_1 \otimes \pi_2 & f \\ 0 & \pi_1 \otimes \pi_2 \end{bmatrix} \xrightarrow{j} \pi_1 \otimes \pi_2 \to 0,$$

where $f\colon \mathcal{D}^* \times \mathcal{D}^* \to \mathrm{End}(W_1 \otimes W_2)$ is given by $f(x_1, x_2) = (a_1 \mathfrak{v}(x_1) + a_2 \mathfrak{v}(x_2))\mathbf{1}_{W_1 \otimes W_2}$ for two arbitrary complex numbers a_1 and a_2.

Proof Thinking of Ext in terms of Yoneda extensions, it is easy to see that each pair $(a_1, a_2) \in \mathbb{C}^2$ gives a short exact sequence, and distinct pairs give distinct Yoneda extensions, *i.e.,* are Yoneda inequivalent. ■

For notational convenience in the above corollary, we will denote the module in the middle by $E_{(a_1, a_2)}$, suppressing the dependence on π_1 and π_2, since in the applications they will be clear from the context.

4 Proof of Theorem 1.1

In this section we give a proof of Theorem 1.1. Consider the short exact sequence of P modules given by Kirillov theory (Theorem 2.1):

$$0 \to C_c^\infty(\mathcal{D}^*, \pi_{N,\psi}) \to \pi \to \pi_N \to 0.$$

Now we apply the functor $\mathrm{Hom}_M(-, \tau)$ to this short exact sequence to get the following long exact sequence:

$$\begin{aligned} 0 \to \mathrm{Hom}_M(\pi_N, \tau) \to \mathrm{Hom}_M(\pi, \tau) \to \mathrm{Hom}_M(C_c^\infty(\mathcal{D}^*, \pi_{N,\psi}), \tau) \\ \to \mathrm{Ext}^1_M(\pi_N, \tau) \to \mathrm{Ext}^1_M(\pi, \tau) \cdots . \end{aligned}$$

The heart of the matter is to analyze this sequence thoroughly.

Recall that as a P module we have $C_c^\infty(\mathcal{D}^*, \pi_{N,\psi}) \simeq \mathrm{ind}_S^P(\pi_{N,\psi} \otimes \psi)$ where S is the Shalika subgroup of G (see §2). For any irreducible representation τ of M we have, using Frobenius reciprocity, the following isomorphisms:

$$\begin{aligned}\mathrm{Hom}_M(C_c^\infty(\mathcal{D}^*, \pi_{N,\psi}), \tau) &\simeq \mathrm{Hom}_M(\mathrm{ind}_S^P(\pi_{N,\psi} \otimes \psi), \tau)\\ &\simeq \mathrm{Hom}_M(\mathrm{ind}_{\mathcal{D}^*}^M(\pi_{N,\psi}), \tau)\\ &\simeq \mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \tau).\end{aligned}$$

The middle isomorphism may be justfied by the fact that the restriction to M of $\mathrm{ind}_S^P(\pi_{N,\psi} \otimes \psi)$ is $\mathrm{ind}_{\mathcal{D}^*}^M(\pi_{N,\psi})$. The fact that τ is finite dimensional is needed for Frobenius reciprocity [4, 2.29] in the last isomorphism.

Using this isomorphism, the long exact sequence may be written as

$$\begin{aligned}0 \to \mathrm{Hom}_M(\pi_N, \tau) \to \mathrm{Hom}_M(\pi, \tau) \to \mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \tau)&\\ \to \mathrm{Ext}^1_M(\pi_N, \tau) \to \mathrm{Ext}^1_M(\pi, \tau) \cdots .&\end{aligned}$$

It is convenient to consider the following exhaustive list of cases.

(1) τ does not intertwine with the Jacquet module π_N. This includes the case when π is supercuspidal.
(2) π is an irreducibly (parabolically) induced representation with π_N semisimple as an M-module and τ intertwines with π_N.
(3) π is an irreducibly (parabolically) induced representation with π_N not semisimple as an M-module and τ intertwines with π_N.
(4) π is a generalized Steinberg representation and $\tau = \pi_N$.
(5) π is a generalized Speh representation and $\tau = \pi_N$.

Case 1: In this case, we have $\mathrm{Hom}_M(\pi_N, \tau) = (0)$. Using Corollary 3.2 we get that $\mathrm{Ext}^1_M(\pi_N, \tau) = (0)$. Hence from the long exact sequence, we get the isomorphism $\mathrm{Hom}_M(\pi, \tau) \simeq \mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \tau)$. In particular, their dimensions are equal.

Case 2: Let $\pi = \mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)$ be an *irreducible* representation of G parabolically induced from $\sigma_1 \otimes \sigma_2$ and assume also that π_N is semisimple. Recall from §2 that the (unnormalized) Jacquet module of π is given by

$$0 \to \sigma_2(d/2) \otimes \sigma_1(-d/2) \to \pi_N \to \sigma_1(d/2) \otimes \sigma_2(-d/2) \to 0.$$

From semisimplicity of π_N, the above sequence splits, which is equivalent to $\sigma_1 \not\simeq \sigma_2$. (This equivalence follows from Corollary 3.2, Frobenius reciprocity, and a parabolically induced representation of G being always multiplicity-free.)

Let τ be an irreducible representation of M which intertwines with π_N. Since $\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2) \simeq \mathrm{Ind}_P^G(\sigma_2 \otimes \sigma_1)$, it suffices to consider $\tau = \sigma_2(d/2) \otimes \sigma_1(-d/2)$. We have $\dim(\mathrm{Hom}_M(\pi_N, \tau)) = 1$. Also, since π_N is semisimple, we have $\mathrm{Ext}^1_M(\pi_N, \tau) =$

$\mathrm{Ext}^1_M(\tau,\tau)$ and the latter is two dimensional. The theorem now follows from the long exact sequence if we show that

$$\dim\big(\,\mathrm{Ker}\big(\,\mathrm{Ext}^1_M(\pi_N,\tau)\to \mathrm{Ext}^1_M(\pi,\tau)\big)\,\big) = 1.$$

This will follow from Lemmas 4.1 and 4.3.

Lemma 4.1 $\dim(\mathrm{Ker}(\mathrm{Ext}^1_M(\pi_N,\tau)\to \mathrm{Ext}^1_M(\pi,\tau)))\geq 1.$

Proof From the long exact sequence, the lemma is equivalent to showing that the map $\mathrm{Hom}_M(\pi,\tau)\to \mathrm{Hom}_M(C_c^\infty(\mathcal{D}^*,\pi_{N,\psi}),\tau)$ is not surjective. To see this, we explicitly construct an element $\ell\in \mathrm{Hom}_M(C_c^\infty(\mathcal{D}^*,\pi_{N,\psi}),\tau)$ which does not extend to π.

Let W_i be the representation space of σ_i, $i=1,2$. The representation space of τ is $W_2\otimes W_1$. By [16], the twisted Jacquet module $\pi_{N,\psi}$ is $\sigma_1\otimes\sigma_2$ as a $\mathcal{D}^*$-module. Let $\iota\colon W_1\otimes W_2\to W_2\otimes W_1$ be the map $\iota(w_1\otimes w_2)=w_2\otimes w_1$ extended linearly. Now consider the map ℓ given by

$$\ell(\phi)=\iota\Big(\int_{\mathcal{D}^*}|x|^{-1/2}(1\otimes\sigma_2(x^{-1}))\phi(x)\,d^*x\Big)$$

for all $\phi\in C_c^\infty(\mathcal{D}^*,\pi_{N,\psi})$. Here d^*x is a Haar measure on $\mathcal{D}^*$. From the action of M on $C_c^\infty(\mathcal{D}^*,\pi_{N,\psi})$,

$$\ell\left(\begin{bmatrix}x&0\\0&y\end{bmatrix}\phi\right)=\sigma_2(d/2)(x)\otimes\sigma_1(-d/2)(y)\ell(\phi),$$

i.e., $\ell\in\mathrm{Hom}_M(C_c^\infty(\mathcal{D}^*,\pi_{N,\psi}),\tau)$.

To show that ℓ does not extend to π, we use our study [17] of the asymptotics in the Kirillov model for π. Theorem 2.1 of [17] can be rephrased to state that the representation space of π can be described as

$$\pi=C_c^\infty(\mathcal{D}^*,\pi_{N,\psi})\oplus\bigoplus_\alpha\mathbb{C}f_\alpha\oplus\bigoplus_\beta\mathbb{C}g_\beta,$$

where f_α and g_β are functions on $\mathcal{D}^*$ defined by

$$f_\alpha(x)=A(x)|x|^{1/2}(\sigma_1(x)\otimes 1)\chi_{\mathcal{O}^*}(x)\alpha,\quad g_\beta(x)=|x|^{1/2}(1\otimes\sigma_2(x))\chi_{\mathcal{O}^*}(x)\beta,$$

with α and β running over any basis for $W_1\otimes W_2$. Here the $A(x)$ is the enigmatic function of x which showed up in [17], and $\chi_{\mathcal{O}^*}$ is the characteristic function of $\mathcal{O}^*\subset\mathcal{D}^*$.

Consider the function g_β. It is easy to see that

$$\left(\begin{bmatrix}t&0\\0&1\end{bmatrix}g_\beta\right)(x)=|t|^{1/2}(1\otimes\sigma_2(t))g_\beta(x)+\Lambda_t(x),$$

where $\Lambda_t(x) = |tx|^{1/2}(1 \otimes \sigma_2(tx))(\chi_{t^{-1}\mathcal{O}^*}(x) - \chi_{\mathcal{O}^*}(x))\beta$. Observe that for each t, $\Lambda_t(x)$ as a function of x is in $C_c^\infty(\mathcal{D}^*, \pi_{N,\psi})$. If ℓ extends to all of π as an element of $\mathrm{Hom}_M(\pi, \tau)$, then applying ℓ to the above equation, we get

$$\ell\left(\begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} g_\beta \right) = \ell(|t|^{1/2}(1 \otimes \sigma_2(t))g_\beta) + \ell(\Lambda_t).$$

Putting $t = \varpi_F$ and cancelling the left-hand side with the first term on the right-hand side, we get

$$\iota\Big(\int_{\mathcal{D}^*} (\chi_{\varpi_F^{-1}\mathcal{O}^*}(x) - \chi_{\mathcal{O}^*}(x))\beta \, d^*x \Big) = 0$$

for any β, which is absurd. Hence ℓ does not extend. ∎

Remark 4.2 Observe that Lemma 4.1, which says the map

$$\mathrm{Hom}_M(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2), \tau) \to \mathrm{Hom}_{\mathcal{D}^*}(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_{N,\psi}, \tau),$$

is not surjective, remains valid even if $\sigma_1 = \sigma_2$. (This will be relevant in §6.)

Lemma 4.3 $\dim(\mathrm{Ker}(\mathrm{Ext}_M^1(\pi_N, \tau) \to \mathrm{Ext}_M^1(\pi, \tau))) \leq 1$.

Proof Since π_N is semisimple and τ occurs in π_N with multiplicity one, we have $\mathrm{Ext}_M^1(\pi_N, \tau) \simeq \mathrm{Ext}_M^1(\tau, \tau)$. We identify the latter with $\mathbb{C}^2$ as in Corollary 3.3. For each $(a, b) \in \mathbb{C}^2$ we have an extension $0 \to \tau \to E_{(a,b)} \to \tau \to 0$. The image of the class $[E_{(a,b)}]$ under the map $\mathrm{Ext}_M^1(\pi_N, \tau) \to \mathrm{Ext}_M^1(\pi, \tau)$ is given by the following pullback diagram.

$$\begin{tikzcd}
0 \arrow[r] & \tau \arrow[r] \arrow[d, "\mathrm{Id}_\tau"] & E_{(a,b)} \times_\tau \pi \arrow[r] \arrow[d] & \pi \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \tau \arrow[r, "i"] & E_{(a,b)} \arrow[r, "j"] & \tau \arrow[r] & 0,
\end{tikzcd}$$

where the map from π to τ factors via π_N. To understand the kernel of the map $\mathrm{Ext}_M^1(\pi_N, \tau) \to \mathrm{Ext}_M^1(\pi, \tau)$, we need to analyze as to when we have a Yoneda equivalence.

$$\begin{tikzcd}
0 \arrow[r] & \tau \arrow[r] \arrow[d, "\mathrm{Id}_\tau"] & \tau \oplus \pi \arrow[r] \arrow[d, "f"] & \pi \arrow[r] \arrow[d, "\mathrm{Id}_\pi"] & 0 \\
0 \arrow[r] & \tau \arrow[r] & E_{(a,b)} \times_\tau \pi \arrow[r] & \pi \arrow[r] & 0.
\end{tikzcd}$$

We will leave it to the reader to check that one has a map f in the above diagram, and what is important is that it is a diagram *of M-modules*, only if $a + b = 0$. ∎

Since Cases 1 and 2 correspond to the hypotheses (i) and (ii) of Theorem 1.1, this completes the proof in those cases.

Case 3: π is irreducibly and parabolically induced, π_N not semisimple and τ intertwines with π_N. This necessarily implies that $\pi = \mathrm{Ind}_P^G(\sigma \otimes \sigma)$ and $\tau = \sigma(d/2) \otimes \sigma(-d/2)$. We have not been able to prove the theorem in this case. (What we believe is true, but do not have a proof if $\mathcal{D} \neq F$, is that $\dim(\mathrm{Ext}^1_M(\pi_N, \tau)) = 2$.)

Cases 4, 5: $\pi = \mathrm{St}(\sigma)$ or $\mathrm{Sp}(\sigma)$ and $\tau = \pi_N$. In both these cases, we have not been able to prove the theorem. The missing ingredient is that for these representations, we do not have precise information on the asymptotics in the Kirillov model. Our previous paper [17] falls short, especially because of the enigmatic function $A(x)$ which shows up in that paper and about which we have no control right now. Another stumbling block is that as of now, we do not know the twisted Jacquet module of these representations. This has been a hindrance in our earlier paper [15] and also in other papers of Dipendra Prasad; see for instance [13]. In the special case of σ being one dimensional, from Remark 2.4 we know that $\mathrm{Sp}(\sigma)$ is one dimensional and we believe that some of the arguments elsewhere in the paper can be used to finesse the proof for $\mathrm{St}(\sigma)$. We have not carried this out, because it is a very special case, and we believe there should be reasonably uniform proofs. Besides, for later applications, we have been able to finesse it anyway.

5 Shalika Models

Definition 5.1 (Shalika Models) Let (π, V) be an irreducible admissible infinite dimensional representation of $G = GL_2(\mathcal{D})$. A linear functional $\ell\colon V \to \mathbb{C}$ is said to be a *Shalika functional* if

$$\ell\left(\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} v \right) = \psi(x)\ell(v)$$

for all $x \in \mathcal{D}$, $a \in \mathcal{D}^*$ and all $v \in V$. We say that π admits a *Shalika model* if there is a nonzero Shalika functional.

Definition 5.2 (M-distinguished) Let (π, V) be an irreducible admissible representation of $G = GL_2(\mathcal{D})$. We say that π is *M-distinguished* if there is a nonzero linear functional $\ell\colon V \to \mathbb{C}$ such that $\ell(\pi(m)v) = \ell(v)$ for all $m \in M$ and all $v \in V$.

The above two notions are intimately linked. To put the following theorem into perspective, we refer the reader to [15, Theorems 6.1, 6.2]. What is proved there is that every nonzero Shalika functional can be *averaged over M* to give a nonzero M-distinguishing functional. The following theorem gives a converse. See also the paper of Jacquet and Rallis [9] which is the source of some of these ideas.

Theorem 5.3 *Let $G = GL_2(\mathcal{D})$ and let $M = \mathcal{D}^* \times \mathcal{D}^*$ be the diagonal subgroup of G. Let π be an irreducible admissible infinite dimensional representation of G. Then π is M-distinguished if and only if π admits a Shalika model.*

Proof Observe that the space of Shalika functionals may be identified with the space $\mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \mathbf{1})$. (Here and elsewhere, $\mathbf{1}$ will denote the trivial one dimensional representation of the group in context.) The essence of the proof is to apply Theorem 1.1 for $\tau_1 = \tau_2 = \mathbf{1}$. Since that result does not apply in all cases, we need to finesse this proof. It is convenient to break up the proof into the following five exhaustive cases.

Case 1: π is supercuspidal. Then Theorem 1.1 applies, and we have

$$\dim_{\mathbb{C}}(\mathrm{Hom}_{\mathcal{D}^*\times\mathcal{D}^*}(\pi, \mathbf{1})) = \dim_{\mathbb{C}}(\mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \mathbf{1})),$$

from which the result follows.

Case 2: π is irreducibly, parabolically induced and π_N is semisimple. The proof is exactly as in Case 1, since Theorem 1.1 applies.

Case 3: $\pi = \mathrm{Ind}_P^G(\sigma \otimes \sigma)$. Then π is necessarily irreducibly parabolically induced and its Jacquet module is not semisimple. Recall from §2 we have

$$0 \to \sigma(d/2) \otimes \sigma(-d/2) \to \pi_N \to \sigma(d/2) \otimes \sigma(-d/2) \to 0.$$

The trivial representation $\mathbf{1}$ of M intertwines with the Jacquet module if and only if $\sigma(d/2) \otimes \sigma(-d/2) = \mathbf{1}$, and the latter is impossible. Hence, Theorem 1.1 applies with $\tau = \mathbf{1}$, and the proof follows as in the previous cases.

Case 4: $\pi = \mathrm{St}(\sigma)$, the generalized Steinberg representation for an irreducible representation σ of $\mathcal{D}^*$. Recall from Theorem 2.2 that we have

$$r_N(\mathrm{St}(\sigma)) = \sigma(a(\sigma)/2) \otimes \sigma(-a(\sigma)/2),$$

or, what is more relevant to us,

$$\mathrm{St}(\sigma)_N = \sigma\Big(\frac{a(\sigma)+d}{2}\Big) \otimes \sigma\Big(\frac{-a(\sigma)-d}{2}\Big).$$

Hence the trivial representation $\mathbf{1}$ intertwines with $\mathrm{St}(\sigma)_N$ if and only if $\sigma = |\cdot|^{-1}$. Therefore if $\sigma \neq |\cdot|^{-1}$, then Theorem 1.1 applies, and as above, we are done.

Suppose now that $\sigma = |\cdot|^{-1}$. We argue that $\mathrm{St}(\sigma)$ is neither M-distinguished nor does it have a Shalika functional, because a necessary condition for both is that the representation should have trivial central character. The central character of $\mathrm{St}(|\cdot|^{-1})$ is $|\cdot|_F^{-2d}$, which is not trivial.

Case 5: $\pi = \mathrm{Sp}(\sigma)$, the generalized Speh representation for an irreducible representation σ of $\mathcal{D}^*$. From Theorem 2.2 we have

$$r_N(\mathrm{Sp}(\sigma)) = \sigma(-a(\sigma)/2) \otimes \sigma(a(\sigma)/2),$$

or, as above, what is more relevant to us,

$$\mathrm{Sp}(\sigma)_N = \sigma\Big(\frac{-a(\sigma)+d}{2}\Big) \otimes \sigma\Big(\frac{a(\sigma)-d}{2}\Big).$$

Hence the trivial representation $\mathbf{1}$ intertwines with $\mathrm{St}(\sigma)_N$ if and only if $\sigma = \mathbf{1}$. Therefore if $\sigma \neq \mathbf{1}$, then Theorem 1.1 is applicable, and as above, we are done. And if σ is trivial, then by Remark 2.4, $\mathrm{Sp}(\sigma)$ is one dimensional, and we are concerned only with infinite dimensional representations of G in this theorem. (The theorem obviously need not be true for one dimensional representations, since they do not admit Shalika models, however, they can be M-distinguished. Indeed, $\mathrm{Sp}(\mathbf{1})$ is the trivial representation of G, which is M-distinguished!) ■

6 A Multiplicity-One Theorem in the Quaternionic Case

From this point onwards we assume that $\mathcal{D}$ is the quaternion division algebra over F. For any $x \in \mathcal{D}$, we let $\overline{x} = \mathrm{Trd}_{\mathcal{D}/F}(x) - x$ be the canonical (anti-)involution on $\mathcal{D}$. For any $g \in GL_n(\mathcal{D})$, define $g^* = w({}^t\overline{g})w^{-1}$, where $w(i,j) = \delta_{i,n-j+1}$ and $({}^t\overline{g})(i,j) = \overline{g_{j,i}}$.

Theorem 6.1 *Let $G = GL_2(\mathcal{D})$ where $\mathcal{D}$ is the quaternion division algebra with center F. Let $M = \mathcal{D}^* \times \mathcal{D}^*$ be the diagonal subgroup. Let π be an irreducible admissible representation of G. Then any one dimensional representation of M occurs as a quotient of the restriction of π to M with multiplicity at most one.*

Proof of Theorem 6.1 assuming Theorem 6.4. (We would like to emphasize that the proofs of Theorems 6.4 and 6.5 are independent of the rest of the paper.) It is convenient to break up the proof into the following five exhuastive cases.

Case 1: π is supercuspidal.

Case 2: π is irreducibly parabolically induced with π_N semisimple.

In both of these cases, Theorem 1.1 applies. Consider the case when τ is one dimensional. From Theorem 6.4, τ occurs in $\pi_{N,\psi}$ at most once, and hence, τ occurs as a quotient of π at most once.

Case 3: $\pi = \mathrm{Ind}_P^G(\sigma \otimes \sigma)$. Then π is irreducibly parabolically induced with π_N not semisimple. We have from Theorem 1.1 (and Section 2) that

$$\dim_{\mathbb{C}}(\mathrm{Hom}_M(\mathrm{Ind}_P^G(\sigma \otimes \sigma), \tau)) = \dim_{\mathbb{C}}(\mathrm{Hom}_{\mathcal{D}^*}(\mathrm{Ind}_P^G(\sigma \otimes \sigma)_{N,\psi}, \tau))$$

as long as τ does not intertwine with π_N, *i.e.*, $\tau \neq \sigma(d/2) \otimes \sigma(-d/2)$. In this case, using Theorem 6.4, we are done. Now consider the following rather special case.

Subcase 3.5: $\pi = \mathrm{Ind}_P^G(\sigma \otimes \sigma)$, σ one dimensional and $\tau = \sigma(d/2) \otimes \sigma(-d/2)$. Going back to the basic long exact sequence of Section 4, we have

$$0 \to \mathrm{Hom}_M(\pi_N, \tau) \to \mathrm{Hom}_M(\pi, \tau) \to \mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi}, \tau) \\ \to \mathrm{Ext}^1_M(\pi_N, \tau) \to \mathrm{Ext}^1_M(\pi, \tau) \cdots .$$

We have from Section 2 for π_N and $\pi_{N,\psi}$ that

$$\dim(\mathrm{Hom}_M(\pi_N,\tau)) = \dim(\mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi},\tau)) = 1.$$

Hence $\dim(\mathrm{Hom}_M(\pi,\tau)) \leq 2$. If $\dim(\mathrm{Hom}_M(\pi,\tau)) = 2$, then the map

$$\mathrm{Hom}_M(\pi,\tau) \to \mathrm{Hom}_{\mathcal{D}^*}(\pi_{N,\psi},\tau)$$

is surjective, which contradicts Lemma 4.1. (See Remark 4.2.) Hence the required dimension is at most one.

Case 4: $\pi = \mathrm{St}(\sigma)$. Theorem 1.1 is applicable as long as $\tau \neq \pi_N$. As above, we are done in this case. The specific case when τ is one dimensional and equal to π_N is taken up as the following subcase.

Subcase 4.5: σ one dimensional, $\pi = \mathrm{St}(\sigma)$, $\tau = \sigma(d) \otimes \sigma(-d) = \pi_N$. Recall from Section 2 that we have

$$0 \to \mathrm{Sp}(\sigma) \to \mathrm{Ind}_P^G(\sigma(-d/2) \otimes \sigma(d/2)) \to \mathrm{St}(\sigma) \to 0.$$

It suffices to show that $\dim(\mathrm{Hom}(\mathrm{Ind}_P^G(\sigma(-d/2) \otimes \sigma(d/2)),\tau)) \leq 1$.

For this we use the results of [17, Theorem 2.1; Remark 2.2]. The point is that for a parabolically induced representation, irrespective of whether it is irreducible or not, one has a Kirillov theory, and in particular we have an exact sequence of P-modules for any two irreducible representations σ_1 and σ_2 of $\mathcal{D}^*$, given by

$$0 \to C_c^\infty(\mathcal{D}^*, \sigma_1 \otimes \sigma_2) \to \mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2) \to \mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_N \to 0.$$

Apply $\mathrm{Hom}_M(-,\tau)$ to this short exact sequence to get

$$\begin{aligned} 0 \to \mathrm{Hom}_M(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_N,\tau) &\to \mathrm{Hom}_M(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2),\tau) \\ &\to \mathrm{Hom}_{\mathcal{D}^*}(\sigma_1 \otimes \sigma_2,\tau) \to \mathrm{Ext}^1_M(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)_N,\tau) \\ &\to \mathrm{Ext}^1_M(\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2),\tau) \cdots . \end{aligned}$$

Specializing to the case at hand, *i.e.*, $\sigma_1 = \sigma(-d/2)$ and $\sigma_2 = \sigma(d/2)$ ($d = 2$ in this section), we can finish the argument exactly as in Subcase (3.5). (We remark that Lemma 4.1 which is used as in Subcase 3.5 does not need $\mathrm{Ind}_P^G(\sigma_1 \otimes \sigma_2)$ to be irreducible, by virtue of [17].)

Case 5: $\pi = \mathrm{Sp}(\sigma)$. Theorem 1.1 applies as long as τ does not intertwine with π_N. In this case, we are done, as in Case 3, for instance. Now consider the following.

Subcase 5.5: σ is one dimensional, $\pi = \mathrm{Sp}(\sigma)$ and $\tau = \sigma \otimes \sigma$. But, if σ is one dimensional, then by Remark 2.4, so is $\mathrm{Sp}(\sigma)$, and the theorem is trivially true in this case. ■

Theorem 6.2 *Let $G = GL_2(\mathcal{D})$ where $\mathcal{D}$ is the quaternion division algebra with center F. Let $M = \mathcal{D}^* \times \mathcal{D}^*$ be the diagonal subgroup. Let π be an irreducible admissible representation of G. Let τ be any irreducible representation of M whose restriction to the diagonal $\mathcal{D}^*$ is irreducible. Then τ occurs as a quotient of the restriction of π to M, with multiplicity at most one.*

Proof The proof of Theorem 6.1 goes through *mutatis mutandis*. For the half-integral subcases, use the fact that for an irreducible representation σ of $\mathcal{D}^*$ of dimension at least 2, the $\mathcal{D}^*$-representation $\sigma \otimes \sigma$ is never irreducible. Observe also that if $\tau = \tau_1 \otimes \tau_2$ and, say $\dim(\tau_1) = 1$ and $\dim(\tau_2) > 1$, then the half-integral subcases are vacuously true. ■

Proposition 6.3 *Let τ_1 and τ_2 be two irreducible representations of $\mathcal{D}^*$. The $\mathcal{D}^*$-representation $\tau_1 \otimes \tau_2$ is irreducible if and only if at least one of the τ_i's is one dimensional.*

Proof (Suggested by Dipendra Prasad) We may assume that both τ_i are minimal, *i.e.,* their conductor is not greater than that of any twist. The proof follows by noting that for a minimal irreducible representation of $\mathcal{D}^*$, the dimension depends only on the conductor. (See [6, Proposition 6.5] for instance.) ■

We now state and prove the theorem that the twisted Jacquet module is multiplicity-free as a $\mathcal{D}^*$-module. This result is due to Dipendra Prasad, although in [13] he attributes it to Rallis. He sketched out a proof [14], but, as has been pointed out to us, there is a minor snag in that proof. The theorem itself is by no means obvious. As is usual in proving such a theorem, it really depends on a theorem on invariant distributions, which we have stated as Theorem 6.5. For the reader's convenience we sketch a proof below, which is essentially the same as the proof in [14].

Theorem 6.4 (Dipendra Prasad) *Let $G = GL_2(\mathcal{D})$ where $\mathcal{D}$ is the quaternion division algebra with center F. Let π be an irreducible admissible representation of G. The twisted Jacquet module $\pi_{N,\psi}$ of π is multiplicity-free as a $\mathcal{D}^*$-module.*

Borrowing the terminology of [4] for an l-space X, we let $S(X) = C_c^\infty(X)$. We let $S^*(X) = \mathrm{Hom}_{\mathbb{C}}(S(X), \mathbb{C})$. If H is a subgroup of a group G, then the action of $h \in H$ on the left (resp. right) on G will be denoted λ_h (resp. ρ_h), *i.e.,* $\lambda_h \cdot g = hg$ (resp. $\rho_h \cdot g = gh^{-1}$). Any involution $*$ on G induces an involution $T \mapsto T^*$ on $S^*(G)$.

Theorem 6.5 *If $T \in S^*(G)$ is a distribution which satisfies*

(i) *T is invariant under conjugation by S, the Shalika subgroup of G;*
(ii) *$\lambda_n \cdot T = \psi(n)T$ and $\rho_n \cdot T = \psi^{-1}(n)T$ for all $n \in N$;*
(iii) *$T^* = -T$.*

Then $T = 0$.

The proof of Theorem 6.5 will require a few lemmas. Observe that $*$ is defined such that if T satisfies (i) and (ii), then so does T^*. Hence, the theorem may also be stated as: a distribution satisfying (i) and (ii) is invariant under $*$. The proof heavily uses Bernstein's localization principle, see [3, p. 58] or [5, Proposition 4.3.15]. To begin, consider the short exact sequences

$$0 \to S^*(P) \to S^*(G) \to S^*(PwP) \to 0,$$
$$0 \to S^*(S) \to S^*(P) \to S^*(P - S) \to 0.$$

Observe that all the spaces PwP, $P - S$ and S are preserved by inner conjugation by S, left and right translations by N and by $g \mapsto g^*$. It suffices to prove the theorem for $T \in S^*(PwP)$ and then for $T \in S^*(P - S)$ and finally for $T \in S^*(S)$. We consider these cases in Lemmas 6.6, 6.7 and 6.8.

Lemma 6.6 *If T is a distribution on PwP which satisfies hypothesis* (i)–(iii) *of Theorem* 6.5*, then* $T = 0$.

Proof Let $T \in S^*(PwP)$ be a distribution which satisfies hypothesis (i)–(iii) as in the statement of the theorem. (It will turn out that we need T to satisfy only (ii) and (iii) for this case.) We apply Bernstein's localization to PwP by considering the map $p_1 \colon PwP \to F^* \times F^*$ given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (\mathrm{Nrd}_{\mathcal{D}/F}(c), \mathrm{Nrd}_{\mathcal{D}/F}(b - ac^{-1}d)).$$

(Note that $b - ac^{-1}d \neq 0$.) Let $y = (\underline{c}, \underline{\delta}) \in F^* \times F^*$. Let $c_0, \delta_0 \in \mathcal{D}^*$ such that $\mathrm{Nrd}_{\mathcal{D}/F}(c_0) = \underline{c}$ and $\mathrm{Nrd}_{\mathcal{D}/F}(\delta_0) = \underline{\delta}$. The fiber $p_1^{-1}(y)$ may be described as:

$$p_1^{-1}(y) = \left\{ \begin{pmatrix} 1 & e \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_0 u & 0 \\ 0 & \delta_0 v \end{pmatrix} \begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} : e, f \in \mathcal{D}; u, v \in \mathcal{D}^{(1)} \right\},$$

where $\mathcal{D}^{(1)} = SL_1(\mathcal{D})$ which is the group of reduced norm-one elements in $\mathcal{D}$. We may therefore identify $p_1^{-1}(y)$ with $\mathcal{D} \times \mathcal{D}^{(1)} \times \mathcal{D}^{(1)} \times \mathcal{D}$ via the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (ac^{-1}, c_0^{-1}c, \delta_0^{-1}(b - ac^{-1}d), c^{-1}d).$$

The left and right N-action and the involution $*$ may be transferred to actions on $\mathcal{D} \times \mathcal{D}^{(1)} \times \mathcal{D}^{(1)} \times \mathcal{D}$ as follows:

- The left N-action is via left translations on the first factor $\mathcal{D}$ of $p_1^{-1}(y)$.
- The right N-action is via right translations on the last factor $\mathcal{D}$ of $p_1^{-1}(y)$.
- The involution $*$ acts via

$$(e, u, v, f) \mapsto (e, u, v, f)^* := (\overline{f}, c_0^{-1}\overline{u}\,\overline{c_0}, \delta_0^{-1}\overline{v}\overline{\delta_0}, \overline{e}).$$

It suffices to prove that a distribution $T \in S^*(\mathcal{D} \times \mathcal{D}^{(1)} \times \mathcal{D}^{(1)} \times \mathcal{D})$ which is left-(N, ψ) and right-(N, ψ^{-1}) invariant is also invariant under $*$.

To prove this we use Bernstein's localization again as follows. For brevity, let $U = \mathcal{D}^{(1)} \times \mathcal{D}^{(1)}$. If $\mathfrak{u} = (u, v) \in U$, then $\mathfrak{u}^* = (c_0^{-1}\overline{uc_0}, \delta_0^{-1}\overline{v\delta_0})$. Consider the map

$$p_2 \colon \mathcal{D} \times U \times \mathcal{D} \to \mathrm{Sym}^2(U),$$

where $\mathrm{Sym}^2(U) := (U \times U)/(\mathbb{Z}/2\mathbb{Z})$ with the action of $(\mathbb{Z}/2\mathbb{Z})$ being to switch the two factors. The map p_2 sends $(e, \mathfrak{u}, f)$ to the class of $(\mathfrak{u}, \mathfrak{u}^*)$, which can be identified with the set $\{\mathfrak{u}, \mathfrak{u}^*\}$. Having fixed the map p_2, it is relatively straightforward to check that any nonempty fiber $p_2^{-1}(y_2)$ with $y_2 = \{\mathfrak{u}, \mathfrak{v}\} \in \mathrm{Sym}^2(U)$ cannot support such a distribution. We leave the details to the reader. ■

Lemma 6.7 *If T is a distribution on $P - S$ which satisfies hypotheses* (i)–(iii) *of Theorem* 6.5*, then $T = 0$.*

Proof For this proof it will suffice to assume that T satisfies only hypothesis (ii) of Theorem 6.5. Here one can use Bernstein's localization by taking the map $p_3 \colon P - S \to \mathcal{D}^* \times \mathcal{D}^*$ given by $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \mapsto (a, d)$. We leave the details to the reader. ■

Lemma 6.8 *If T is a distribution on S which satisfies hypothesis* (i)–(iii) *of Theorem* 6.5*, then $T = 0$.*

Proof For this proof it suffices to assume that T satisfies (i) and (iii) in the hypothesis of Theorem 6.5. The lemma can then be restated as: *a conjugation invariant distribution T on S is also invariant under $*$.* This follows from well-known results of Bernstein and Zelevinskii (see [4, Theorems 6.13, 6.15] or [17, pp. 460–461]) once we establish the following claim.

Claim *In S, any element s is conjugate to s^*.*

To see this, let $s = \left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right) \in S$. If $a \in F$, then choose $t \in \mathcal{D}^*$ such that $tbt^{-1} = \overline{b}$. Then $\left(\begin{smallmatrix} t & 0 \\ 0 & t \end{smallmatrix}\right)$ conjugates s to s^*. If $b \in F$, then choose $t \in \mathcal{D}^*$ such that $tat^{-1} = \overline{a}$. Then $\left(\begin{smallmatrix} t & 0 \\ 0 & t \end{smallmatrix}\right)$ conjugates s to s^*.

Assume henceforth that $a \notin F$ and $b \notin F$. Consider the matrix equation

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \begin{pmatrix} t & x \\ 0 & t \end{pmatrix} = \begin{pmatrix} t & x \\ 0 & t \end{pmatrix} \begin{pmatrix} \overline{a} & \overline{b} \\ 0 & \overline{a} \end{pmatrix}.$$

The above matrix equation is also the following system of equations:

$$at = t\overline{a}, \quad ax + bt = t\overline{b} + x\overline{a}.$$

We need to show that we can solve these equations with $t \in \mathcal{D}^*$ and $x \in \mathcal{D}$. Let $W_1 = \{y \in \mathcal{D} : ay = y\overline{a}\}$ and let $W_2 = \{y \in \mathcal{D} : by = y\overline{b}\}$. It is clear that both W_1 and W_2 are two dimensional F-subspaces of $\mathcal{D}$. Since we assumed that neither a nor

b are central, we also have $W_1 \cap Z = W_2 \cap Z = (0)$ if $Z \simeq F$ is the center of $\mathcal{D}$. Hence W_1 and W_2 intersect non trivially mod-the-center, *i.e.,* there is a $t \in \mathcal{D}^*$ and $z \in Z$ such that $t \in W_1$ and $t{+}z \in W_2$. The latter condition implies $b(t{+}z) = (t{+}z)\overline{b}$, which gives $bt - t\overline{b} = z\overline{b} - bz = z(\overline{b} - b)$. Observe that $(\overline{b} - b)/(\overline{a} - a)$ is a nonzero element of the center. Let $x = z(\overline{b} - b)/(\overline{a} - a)$. Then we have $z(\overline{b} - b) = x(\overline{a} - a) = x\overline{a} - ax$. Hence we have solved the above equations for t and x. This establishes the claim and completes the proof of Lemma 6.8. ■

Proof of Theorem 6.5 Theorem 6.5 follows from Lemmas 6.6–6.8. ■

Proof of Theorem 6.4 The proof using Theorem 6.5 is entirely standard. One can argue as in the proof of multiplicity one for Whittaker models for $GL(n)$ [5, pp. 456-458]. (See also [2, Theorems 1.1, 1.2, 2.5, 2.6].) The involution used in [5] needs to be replaced by our involution $g \mapsto g^*$, while using an earlier theorem of ours [17, Theorem 3.1] that for $GL_n(\mathcal{D})$, given an irreducible representation π, its contragredient representation is equivalent to $g \mapsto \pi((g^*)^{-1})$. We leave the details to the reader. ■

Acknowledgments I thank Dipendra Prasad for his interest in this work; indeed this paper owes its existence to my joint work with him on the group $GL_2(\mathcal{D})$. It is a pleasure to thank Phil Kutzko for his enthusiasm and encouragement for all my p-adic endeavors. I thank U. K. Anandavaradhanan, Larry Morris, and members of the representation theory seminar at Iowa for hearing me ramble on this on many occasions. I thank the referee for many detailed suggestions which have enhanced the clarity of exposition.

References

[1] A.I. Badulescu, *Correspondance de Jacquet-Langlands pour les corps locaux de caractéristique non nulle.* Ann. Sci. École Norm. Sup. **35**(2002), no. 5, 695–747.

[2] E. M. Baruch and S. Rallis, *A uniqueness theorem of Fourier Jacobi models for representations of* Sp(4). J. London Math. Soc. **62**(2000), no. 1, 183–197.

[3] J. Bernstein, *P-invariant distributions on* GL(n) *and the classification of unitary representations of* GL(N), *(non-Archimedean case).* In: Lie Group Representations. II. Lecture Notes in Math. 1041, Springer- Verlag, 1984, pp. 141–184.

[4] I. N. Bernšteĭn and A. V. Zelevinskiĭ, *Representation theory of $GL(n, F)$ where F is a local non-Archimedean field.* Uspehi Mat. Nauk, **31**(1976), no. 3, 1–76.

[5] D. Bump, *Automorphic forms and representations.* Cambridge Studies in Advanced Mathematics 55, Cambridge University Press, Cambridge, 1997.

[6] H. Carayol, *Représentations cuspidales du groupe linéaire.* Ann. Sci. École Norm.Sup. **17**(1984), no. 2, 191–225.

[7] P. Deligne, D. Kazhdan and M.-F. Vignéras, *Représentations des algèbres centrales simples p-adiques*, In: Representations of reductive groups over a local field, Hermann, Paris, 1984, pp. 33–118.

[8] H. Jacquet and R. Langlands, *Automorphic Forms on* GL_2. Lecture Notes in Math. 114, Springer-Verlag, Berlin, 1970.

[9] H. Jacquet and S. Rallis, *Uniqueness of linear periods.* Compositio Math. **102**(1996), no. 1, 65–123.

[10] S. S. Kudla, *The local Langlands correspondence: The non-Archimedean case.* In Motives. Proc. Sympos. Pure Math. 55, American Mathematical Society, Providence, RI, 1994, pp. 365–391.

[11] C. Mœglin and J.-L. Waldspurger, *Modèles de Whittaker dégénérés pour des groupes p-adiques.* Math. Z. **196**(1987), no. 3, 427–452.

[12] D. Prasad, *Some remarks on representations of a division algebra and of the Galois group of a local field.* J. Number Theory, **74**(1999), no. 1, 73–97.

[13] ———, *The space of degenerate Whittaker models for $GL(4)$ over p-adic fields.* In: Cohomology of Arithmetic Groups, L-Functions and Automorphic Forms. Tata Inst. Fund. Res. Stud. Math. 15, Tata Inst. Fund. Res., Bombay, 2001, pp. 103–115.

[14] ———, *The space of degenerate Whittaker models for general linear groups over finite and p-adic fields.* www.math.tifr.res.in/~dprasad/tifr.pdf.

[15] D. Prasad and A. Raghuram, *Kirillov theory for $GL_2(\mathcal{D})$ where $\mathcal{D}$ is a division algebra over a non-Archimedean local field.* Duke Math. J. **104**(2000), no. 1, 19–44.

[16] A. Raghuram, *Some Topics in Algebraic Groups : Representation Theory of $GL_2(\mathcal{D})$ Where $\mathcal{D}$ Is a Division Algebra over a Non-Archimedean Local Field.* Ph.D. Thesis, TIFR, University of Mumbai, 1999.

[17] ———, *On representations of p-adic $GL_2(\mathcal{D})$.* Pacific J. Math. **206**(2002), no. 2, 451–464.

[18] ———, *A Künneth theorem for p-adic groups,* To appear, Canad. Math. Bull.

[19] J. Rogawski, *Representations of $GL(n)$ and division algebras over a p-adic field.* Duke Math. J. **50**(1983), no. 1, 161–196.

[20] M. Tadić, *Induced representations of $GL(n, A)$ for p-adic division algebras A.* J. Reine Angew. Math. **405**(1990), 48–77.

[21] J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie.* Compositio Math. **54**(1985), no. 2, 173–242.

Department of Mathematics
Oklahoma State University
Stillwater, OK 74078
U.S.A.
e-mail: araghur@math.okstate.edu