SUBCONVEX BOUNDS FOR $U_{n+1} \times U_n$ IN HORIZONTAL ASPECTS

YUEKE HU AND PAUL D. NELSON

ABSTRACT. For L-functions attached to automorphic representations of unitary groups $U_{n+1} \times U_n$, we establish a subconvex bound valid in certain horizontal aspects, where the set of ramified places is allowed to vary.

CONTENTS

1. INTRODUCTION

1.1. **Overview.** This paper concerns the quantitative analysis of automorphic L-functions on higher rank classical groups. Recent works on this topic include

- an approach to the analytic test vector problem based on the orbit method, applied to the asymptotic evaluations of moments (see [20]), and
- subconvex bounds on $U_{n+1} \times U_n$ and GL_n (see [19, 18] and announcement talks of Marshall).

These works study sequences of automorphic forms whose ramification increases inside some *fixed* finite set of places, a setup known as the *depth* aspect.

By a *horizontal* aspect, we mean one where the set of ramified places is itself allowed to vary. For example, one can study Dirichlet characters of conductor p^m for fixed p as $m \to \infty$ (depth) and for fixed m as $p \to \infty$ (horizontal).

This paper establishes subconvex bounds on $U_{n+1} \times U_n$ in certain horizontal aspects. Our results are uniform enough to apply also in the depth aspect (indeed, as "$p^m \to \infty$"), but this is not the main novelty.

The depth aspect often has a Lie-algebraic flavor, involving tools such as stationary phase analysis and Taylor approximation. For instance, the works noted above make heavy use of the exponential map and Lie algebra for groups such as $GL_n(F)$, with F a fixed

local field. These techniques are applied to spectral problems (construction and analysis of test vectors) and geometric problems (the "volume bound").

In horizontal aspects, Lie-algebraic techniques are often less relevant. For this paper, we restrict to "even depth" cases ("p^{2m}") where the spectral analysis can still be carried out using Lie-algebraic techniques, but the geometric analysis requires new arguments, of algebro-geometric rather than Lie-algebraic flavor.

We now describe our main result (Theorem 1.1) and the main ideas of its proof, emphasizing new features encountered in horizontal aspects.

1.2. **The refined Gan–Gross–Prasad conjectures.** These conjectures, now known in many cases, provide a link between values of L-functions and integrals of automorphic forms. The body of this paper addresses the local problems that arise in estimating such integrals. The link described here is not otherwise applied in this paper, but provides context for interpreting our results.

Let F be a number field with adele ring $\mathbb{A}$, let E/F be a quadratic extension, let V be an $(n+1)$-dimensional hermitian space over E, and let W be an n-dimensional nondegenerate subspace of V. Define the pair of unitary groups $(G, H) := (\mathrm{U}(V), \mathrm{U}(W))$ over F. Given a pair of cuspidal automorphic representations π and σ of G and H that are locally distinguished, one may attach a branching coefficient $\mathcal{L}(\pi, \sigma)$ quantifying how automorphic forms in π correlate against those in σ. The definition depends upon the choice of a finite set S of places of F, taken large enough to contain every place that is archimedean or at which π or σ is ramified, and requires π and σ to be (nearly) tempered inside S. It is characterized by the following family of identities: for all $v \in \pi$ and $u \in \sigma$ that are unramified outside S, and with suitable normalization of measure,

$$\left| \int_{H(F) \backslash H(\mathbb{A})} v\bar{u} \right|^2 = \mathcal{L}(\pi, \sigma) \int_{H(F_S)} \langle hv, v \rangle \langle u, hu \rangle \, dh. \tag{1.1}$$

We refer to [19, §3.7] for further details. It has been conjectured by Ichino–Ikeda [11] and N. Harris [8] that if S is large enough in the sense recorded in [11, §1], then $\mathcal{L}(\pi, \sigma)$ is given by a ratio of L-values, namely, with notation as in [3],

$$\mathcal{L}(\pi, \sigma) = 2^{-\beta} \frac{L^{(S)}(\pi_E \otimes \sigma_E^{\vee}, 1/2)}{L^{(S)}(\mathrm{Ad}, \pi \boxtimes \sigma^{\vee}, 1)} \Delta_G^{(S)}. \tag{1.2}$$

This expectation has been proved at least when π and σ are tempered at all places [3].

1.3. **Main result.** We establish a subconvex bound on $\mathrm{U}_{n+1} \times \mathrm{U}_n$, with U_n anisotropic, for pairs (π, σ) whose ramification concentrates at some (varying) finite place $\mathfrak{p}$, provided that the conductor does not drop, under some local assumptions (e.g., principal series of even depth). In more detail:

Theorem 1.1. *Fix F, E, G, H as in §1.2, as well as a finite set S of of places of F, large enough in the sense specified in* [19, §3.6]. *Let $\mathcal{F}$ be a family of tuples $(\pi, \sigma, \mathfrak{p}, \mathfrak{q})$ satisfying the following conditions:*

(a) π and σ are cuspidal automorphic representations of G and H, respectively, having unitary central characters.

(b) $\mathfrak{p} \notin S$ is a non-archimedean place (or prime ideal) of F at which E/F splits, so that (see [19, §3.3]*)*

$$(G(F_{\mathfrak{p}}), H(F_{\mathfrak{p}})) \cong (\mathrm{GL}_{n+1}(F_{\mathfrak{p}}), \mathrm{GL}_n(F_{\mathfrak{p}})).$$

(c) (π, σ) is locally distinguished: there is a nonzero $H(\mathbb{A})$-invariant functional $\pi \to \sigma$.

(d) *Inside S, the representations π and σ are tempered, and their depth is uniformly bounded as (π, σ) traverses $\mathcal{F}$ (see [19, §1.3]).*

(e) *Outside $S \cup \{\mathfrak{p}\}$, the representations π and σ are unramified, and σ satisfies a uniform bound towards Ramanujan at places where E/F splits: it is ϑ-tempered at such places for some fixed $0 \leq \vartheta < 1/2$ (see [19, §5.2.1]).*

(f) *$\mathfrak{q}$ is a positive power of the prime ideal $\mathfrak{p}$.*

(g) *At $\mathfrak{p}$, the representations π and σ belong to the principal series and are induced by characters of conductor dividing $\mathfrak{q}^2$ for which the local analytic conductor for $\mathcal{L}(\pi, \sigma)$ at $\mathfrak{p}$ is as large as possible (see (1.6) for details, and Remark 1.6 for weaker allowable hypotheses).*

Assume that V (hence also W) is positive-definite. In particular, H is anisotropic, and at each archimedean place of F, the groups G and H are compact.

Then the branching coefficients $\mathcal{L}(\pi, \sigma)$, defined relative to the set $S \sqcup \{\mathfrak{p}\}$, satisfy a subconvex bound (see (1.8) and (1.9)) as (π, σ) varies over the family $\mathcal{F}$.

1.4. **Some details.** Here we clarify the statement of Theorem 1.1. We may write the local components at the "interesting" place $\mathfrak{p}$ as the normalized inductions

$$\pi_{\mathfrak{p}} = \chi_1 \boxplus \cdots \boxplus \chi_{n+1}, \qquad \sigma_{\mathfrak{p}} = \eta_1 \boxplus \cdots \boxplus \eta_n \qquad (1.3)$$

for some characters χ_i and η_j of $F_{\mathfrak{p}}^{\times}$. For a character ω of $F_{\mathfrak{p}}^{\times}$, we denote by $C(\omega)$ the analytic conductor, i.e., the absolute norm of the largest integral ideal $\mathfrak{a}$ in $F_{\mathfrak{p}}$ such that $\omega(x) = 1$ whenever $x - 1 \in \mathfrak{a}$. We set

$$T := \text{the absolute norm of } \mathfrak{q}^2. \qquad (1.4)$$

With this notation, the "depth $\mathfrak{q}^2$" assumption means

$$C(\chi_i) \leq T, \qquad C(\eta_j) \leq T, \qquad (1.5)$$

while "conductor as large as possible" means

$$\prod_{i,j} C(\chi_i/\eta_j)^2 = T^{2n(n+1)}, \qquad (1.6)$$

which is equivalent to requiring that for all i and j,

$$C(\chi_i/\eta_j) = T. \qquad (1.7)$$

We note that this forces T to be the maximum of the $C(\chi_i)$ and $C(\eta_j)$. "Subconvex bound" means that

$$\mathcal{L}(\pi, \sigma) \leq c_{\mathcal{F}} T^{2n(n+1)(1/4-\delta_n)} \qquad (1.8)$$

for some $c_{\mathcal{F}} \geq 0$ (resp. $\delta_n > 0$) depending only upon $\mathcal{F}$ (resp. n). In cases where $\mathcal{L}(\pi, \sigma)$ is known to coincide with an L-value (e.g., the case that π and σ are everywhere tempered), the estimate (1.8) improves upon the convexity bound, which would assert the same for some $\delta_n \leq 0$ (see [19, Proof of Cor 1.2] for details). In the case $n + 1 = 2$, such an estimate has been known for a while [17]. When $n + 1 \geq 3$, such an estimate is new; we will show then that (1.8) holds for any

$$\delta_n < \frac{1 - 2\vartheta}{4n(n+1)(A+1-2\vartheta)}, \quad A := (2(n+1)^2 - n)(n+1), \qquad (1.9)$$

with ϑ as in assumption (e) above (and $c_{\mathcal{F}}$ allowed to depend upon δ_n).

Remark 1.2. We do not address the interesting challenge of improving the numerical strength of the exponent (1.9). Several avenues for doing so were mentioned in [19, Rmk 1.4], many of which could be pursued in the present context.

Remark 1.3. The "large conductor" assumption (1.6) is the most serious one — it is an open problem to give any genuine subconvex bound for GL_3 or higher in any case where the conductor drops (see the final paragraph of [18, §1.4] and references).

1.5. **Related results.** In talks starting in March 2018, Simon Marshall tentatively announced results in the direction of Theorem 1.1, introducing important and fundamental ideas, under some additional assumptions:

 (i) (Wall-avoidance) For all $i \neq j$, one has
$$C(\chi_i/\chi_j) = T, \qquad C(\eta_i/\eta_j) = T.$$

 (ii) (Depth aspect) $\mathfrak{p}$ is a *fixed* (i.e., independent of $\mathcal{F}$) non-archimedean place.

An important input to his arguments is a certain *volume bound* (see §1.7 and [19, §1.5]), which Marshall announces he can establish assuming (i) and (ii).

The paper [19] gave analogues of Theorem 1.1 at an archimedean place $\mathfrak{p}$ (without assuming "principal series" or "E/F split at $\mathfrak{p}$"). Parts of the proof apply to the non-archimedean "$\mathfrak{p}$ fixed" case of Theorem 1.1; for instance, the volume bound [19, Thm 15.2] was established over any fixed local field, without assuming wall-avoidance.

In summary, the "$\mathfrak{p}$ fixed" case of Theorem 1.1 — an estimate like (1.8), but with $c_{\mathcal{F}}$ allowed to depend also upon $\mathfrak{p}$ — is closely related to existing results. The main novelty here is to allow $\mathfrak{p}$ to vary.

Remark 1.4. Strictly speaking, one could quantify the available arguments to obtain *some* uniformity, e.g., (1.8) with $c_{\mathcal{F}}$ depending polynomially upon $\mathfrak{p}$. This yields subconvexity under a "sufficient depth" restriction, namely, that $\mathfrak{q} = \mathfrak{p}^m$ with m large enough in terms of the rank n. Our main novelty is thus to allow $\mathfrak{p}$ to vary while $\mathfrak{q}$ is a small power of $\mathfrak{p}$, e.g., $\mathfrak{q} = \mathfrak{p}$. Such "horizontal" cases do not follow from direct quantification.

1.6. **Division of the proof.** Like in previous works, the basic object of study is the integral
$$\int_{[H]} \pi(\omega)v \cdot \overline{u} \tag{1.10}$$
for suitable vectors $v \in \pi$ and $u \in \sigma$ and a "convolution kernel" or "amplifier" $\omega \in C_c^\infty(G(\mathbb{A}))$, chosen so that $\pi(\omega)v$ approximates v. One seeks to bound the integrals (1.10) simultaneously

- from below, using (1.1), in terms of $\mathcal{L}(\pi, \sigma)$, and
- from above, by their second moment over "all" π and v, using its "relative trace formula" expansion

$$\int_{x,y \in [H]} \overline{u}(x)u(y) \sum_{\gamma \in G(F)} \omega(x^{-1}\gamma y) \, dx \, dy. \tag{1.11}$$

We work with factorizable vectors v and u and a factorizable test function ω. Their local components at the "uninteresting" places are chosen in a soft and general way [19, §5]. The key point is to choose the components $v_{\mathfrak{p}}, u_{\mathfrak{p}}$ and $\omega_{\mathfrak{p}}$ at the "interesting" place and then to establish the required estimates. The proof may be divided into roughly the following steps (see [19, §2] for a more leisurely overview).

 (1) For an individual representation π of G, a construction of vectors $v \in \pi$ that are *microlocalized* with respect to some parameter τ [20, §1.7].

 (2) For a pair (π, σ) as above, a study of the pairs of parameters (τ, τ_H), at which microlocalized vectors $v \in \pi$ and $u \in \sigma$ exist, that are *compatible* in the sense that τ restricts to τ_H [20, §13-14].

(3) Spectral estimates: the estimation (in particular, lower bound) of integrals of matrix coefficients of microlocalized vectors [20, §18].
(4) Geometric estimates, namely the "volume bound", which is the key technical problem that arises when attempting to estimate (1.11).

The first three of these steps adapt readily to the setting of this paper:

(1) We are content to impose local conditions that make the construction of microlocalized vectors particularly concrete (§8).
(2) The study of compatible parameters was addressed in [20, §13-14] over any base field of characteristic zero; we have found it convenient here to extend that study to a general base ring (§4), using arguments quite similar to those in *loc. cit.*
(3) The spectral estimates proceed in our context as in [20, §18], with simplification.

The focus of this paper is thus on the geometric estimates, namely, in establishing a volume bound that is uniform in $\mathfrak{p}$. We discuss this further starting in §1.7.

Remark 1.5. The method and ideas employed here have a long history. We refer to [19, §2.6] for an overview, but emphasize the use of amplification following Duke–Friedlander–Iwaniec [12, 5, 6], the systematic study of period integrals and test vectors following Bernstein–Reznikov [2], Venkatesh [23] and Michel–Venkatesh [17], and the influential papers of Sarnak [22] and Iwaniec–Sarnak [13].

Remark 1.6. To simplify the statement of Theorem 1.1 and focus on the primary "horizontal" novelty, we assumed that the local components $\pi_\mathfrak{p}$ and $\sigma_\mathfrak{p}$ belong to the principal series. We will actually establish a more general result, where that assumption is weakened to a more complicated condition: that $\pi_\mathfrak{p}$ and $\sigma_\mathfrak{p}$ are "regular at depth $\mathfrak{q}^2$" (Definition 8.1), with the analogue of the "large conductor" assumption (1.7) being that the pair $(\pi_\mathfrak{p}, \sigma_\mathfrak{p})$ is "stable at depth $\mathfrak{q}^2$" (Definition 8.6). We formulate these definitions in terms of what is needed by our argument (the existence of suitable vectors), and verify (§8) that they are closed under parabolic induction and apply to certain supercuspidals (e.g., characters). We do not exhaustively classify the representations to which they apply. Giving such a classification is an interesting problem in the representation theory of supercuspidals that belongs to step (1) in the proof strategy outlined above. For the other steps (2), (3) and (4), our treatment is general.

Remark 1.7. It would be desirable to generalize Theorem 1.1 by removing the parity condition on the conductor exponent, i.e., by replacing "$\mathfrak{q}^2$" with $\mathfrak{q}$ in the statement, or equivalently, by allowing T to be the norm of any ideal (rather than the square an ideal) in the discussion following Theorem 1.1. Such a generalization seems accessible in the depth aspect, but would require new ideas in horizontal aspects. The issues touch upon all steps (1), (2), (3) and (4) in the proof strategy outlined above; for instance, in place of the volume bound, one might need to estimate character sums (see the third paragraph of [18, §1.4] and references). This seems to us an interesting direction for future work.

Remark 1.8. Using the methods of [18], it should be possible to extend Theorem 1.1 to the split case $E = F \times F$ and to Eisenstein series. This would yield subconvex bounds in horizontal aspects for standard L-functions away from conductor dropping. Such bounds should apply to character twists $L(\pi \otimes \chi, \frac{1}{2} + it)$, with π fixed and χ of *square* conductor. More generally, it should be possible to establish subconvex bounds for standard L-functions $L(\pi, \frac{1}{2})$ when π is induced at each finite place from characters all having the same square conductor. To remove the squareness assumption would require extending our methods as in Remark 1.7.

Remark 1.9. One motivation for studying horizontal aspects comes from conjectures and results of Lapid–Mao [14, 16, 15], which relate quadratic twists of self-dual standard L-functions on GL_n to Fourier coefficients of automorphic forms on Mp_{2n}. Estimates for such coefficients could have applications to representation problems concerning quadratic forms. The present work does not apply to such twists due to the squareness restriction in Remark 1.8, but may be understood as a step in that direction.

1.7. **Geometric estimates: the uniform volume bound.** Continuing the discussion of §1.6, we formulate the volume bound below as Problem 1.10. The asymptotic determination of (1.11) reduces to the volume bound as in previous work (see [19, §1.5.3] or §7). Cases of the volume bound sufficient for the depth aspect were established in [19], using reductions specific to that aspect (see §1.9 for details). We develop a different approach (§1.8) that gives the volume bound in general, uniformly in $\mathfrak{p}$.

The volume bound is a local assertion concerning the "interesting" place $\mathfrak{p}$. To simplify notation, we write simply G, H, etc., for points over $F_\mathfrak{p}$, and drop the subscripts $\mathfrak{p}$. Thus, F now denotes a non-archimedean local field, arising as the local component "$F_\mathfrak{p}$" of the global field considered above. We write simply $\mathfrak{p}$ for its maximal ideal, $\mathfrak{o}$ for its maximal order, and $q := |\mathfrak{o}/\mathfrak{p}|$ for the residue field cardinality. We set

$$(G, H, M, M_H) = (\mathrm{GL}_{n+1}(F), \mathrm{GL}_n(F), \mathrm{Mat}_{n+1}(F), \mathrm{Mat}_n(F)),$$

where Mat_n means "$n \times n$ matrices".

Let $K = \mathrm{GL}_{n+1}(\mathfrak{o}) < G$ denote the standard maximal compact subgroup. Recall from Theorem 1.1 the positive power $\mathfrak{q}$ of $\mathfrak{p}$, with square $\mathfrak{q}^2$ of norm T, that controls the depths of the representations π of G and σ of H that we consider.

Let $\tau \in M$, with upper-left block $\tau_H \in M_H$. Let G_τ and H_{τ_H} denote the respective centralizers of τ in G and of τ_H in H. We say that τ is *stable* if it has no eigenvalues in common with τ_H over an algebraic closure, or equivalently, if the characteristic polynomials of τ and τ_H generate the unit ideal. This is an avatar for the "large conductor" assumption (1.7) (see [20, §15]) and is equivalent to the geometric invariant theory notion of stability (see [20, §14]).

Problem 1.10. Fix a stable element $\tau \in M$. Fix $a \in G - HZ$, where $Z < G$ denotes the center. Give a nontrivial bound for the volume of the set of all $y \in H_{\tau_H} \cap K$ for which ay is congruent modulo $\mathfrak{q}$ to an element of HG_τ.

Here "nontrivial" means a power saving in T over the trivial bound, $\mathrm{vol}(H_{\tau_H} \cap K)$. Strictly speaking, we need mild refinements of the problem statement depending quantitatively upon a.

A key result of this paper is a solution to this problem (Theorem 6.7) that is uniform with respect to variation of the underlying local field F. In §1.8, we summarize the proof. In §1.9, we explain why the depth aspect treatment given in [19] is not uniform.

1.8. **The approach of this paper.** The volume bound controls elements of H_{τ_H} lying close to a certain subvariety, with "close" quantified by the ideal $\mathfrak{q}$. We deduce it from a purely algebraic statement, Theorem 1.11, whose informal content is that, among the equations defining that subvariety, there is at least one whose coefficients are at least as large as the distance from a to HZ.

To formulate that algebraic statement, we denote now by $(\mathbf{G}, \mathbf{H}, \mathbf{M}, \mathbf{M}_H)$ the group schemes $(\mathrm{GL}_{n+1}, \mathrm{GL}_n, \mathrm{Mat}_{n+1}, \mathrm{Mat}_n)$ of invertible and all square matrices of the indicated dimensions, and write $\mathbf{Z}$ for the center of $\mathbf{G}$. Let R be a ring, and let $\tau \in \mathbf{M}(R)$ satisfy the following "stability" hypothesis (§4): the characteristic polynomials of τ and

its upper-left block τ_H generate the unit ideal. Let $\mathbf{G}_\tau$ and $\mathbf{H}_{\tau_H}$ denote the centralizers of τ and τ_H, regarded as subgroup schemes of $\mathbf{G}$ and $\mathbf{H}$ defined over R, and let $a \in \mathbf{G}(R)$. For each R-algebra R', we define the following set (compare with Problem 1.10):

$$\mathbf{X}_{\tau,a}(R') := \{y \in \mathbf{H}_{\tau_H}(R') : ay \in \mathbf{H}(R')\mathbf{G}_\tau(R')\}.$$

The stability hypothesis on τ turns out to imply that $\mathbf{X}_{\tau,a}$ is naturally a closed subscheme of $\mathbf{H}_{\tau_H}$ over R, defined by finitely many polynomials equations (see Lemma 5.8).

It is easy to see that if a lies in $\mathbf{H}(R)\mathbf{Z}(R)$, then $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$ as schemes over R (i.e., their point sets coincide for all R'). We establish a converse:

Theorem 1.11 (Theorem 5.1). *Assume that $n + 1 \geq 3$, that 2 is a unit in R, and that $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$ as schemes over R. Then $a \in \mathbf{H}(R)\mathbf{Z}(R)$.*

We refer to §5.1 for refinements and discussion of why the hypotheses are necessary. Informally, the condition $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$ says that the defining equations for $\mathbf{X}_{\tau,a}$ inside $\mathbf{H}_{\tau_H}$ are tautological, i.e., their coefficients are all zero. The informal content of Theorem 1.11 is thus that if $a \notin \mathbf{H}(R)\mathbf{Z}(R)$, then we can find a nonzero (bounded degree) polynomial on $\mathbf{H}_{\tau_H}$ over R whose zero locus contains $\mathbf{X}_{\tau,a}$. This result, applied in the context of Problem 1.10 with R a suitable quotient of $\mathfrak{o}/\mathfrak{q}$, implies that the set whose volume we must bound is contained in the locus of a polynomial whose coefficients are not all too small. The required uniform volume bound (Theorem 6.7) then follows via general bounds for solutions to polynomial congruences (§6.1).

To prove Theorem 1.11, we apply the assumed equality between $\mathbf{X}_{\tau,a}$ and $\mathbf{H}_{\tau_H}$ first over the ring of dual numbers $R' = R[\varepsilon]/(\varepsilon^2)$, then over $R'' = R[\varepsilon_1, \varepsilon_2]/(\varepsilon_1, \varepsilon_2)$. We refer to these steps as linear and quadratic analysis, respectively. They may be understood as studying the consequences of the vanishing of linear and quadratic coefficients of the defining equations for $\mathbf{X}_{\tau,a}$ inside $\mathbf{H}_{\tau_H}$. The arguments sketched below were found after extensive numerical study of these coefficients,[1] and represent a main novelty of this paper.

The key case is when $a \in \mathbf{G}_\tau(R)$; we must show then that $a \in \mathbf{Z}(R)$.

From the linear analysis, we deduce that $a^2 \in \mathbf{Z}(R)$. To do so, we determine the linear coefficients with respect to the basis for $\mathrm{Lie}(\mathbf{H}_{\tau_H})$ given by powers of τ_H, and observe that they may be related via an upper-triangular substitution to some invariants of τ and a whose vanishing forces a^2 to be central. We refer to §5.4 for details. This step of the argument generalizes the approach of [19], recalled below in §1.9, which amounts to studying just the linear coefficient for the central direction in $\mathrm{Lie}(\mathbf{H}_{\tau_H})$.

From the quadratic analysis, we construct (Lemma 5.17) a multiplicative linear map $\mathrm{Lie}(\mathbf{H}_{\tau_H}) \to \mathrm{Lie}(\mathbf{G}_\tau)$. Our argument then divides according to whether this map preserves unit elements (residually). If it does not, then we may apply multiplicativity in the central direction to see that a is central. If it does, then we show by two applications of the linear analysis — first for a over R, then for suitable "nearby" $b \in \mathbf{G}_\tau(R')$ over $R' := R[\varepsilon_2]/(\varepsilon_2^2)$, with $a(1 + \varepsilon_1\tau_H) \in \mathbf{H}(R')b$ — that τ satisfies a monic polynomial of degree 2, contrary our assumption $n + 1 \geq 3$. We refer to §5.5 for details.

Remark 1.12. In the depth aspect, we may assume via the "near identity" reduction (§1.9.3) that a is close to the identity. Then $a^2 \in \mathbf{Z}(R) \implies a \in \mathbf{Z}(R)$, so the "linear" part of above argument suffices, yielding a simpler proof than in [19] of the volume bound required in the depth aspect (see Remark 5.14).

---

[1]using SAGE [**?**] (and its components Singular [**?**] and GiNaC) and Emacs Calc

1.9. **Available treatment in the depth aspect.** We indicate here why the approach to the volume bound given in [19] does not apply in horizontal aspects, motivating the approach described above and pursued in this paper.

1.9.1. *The short spectral projector ω.* We construct in §9.3 a convolution kernel $\omega \in C_c^\infty(K)$, the local component of the global amplifier discussed after (1.10). It is supported on a certain open subgroup $J_\tau < K$, and given there by a multiple of a certain character of J_τ, depending upon τ and some additional data. Roughly, J_τ is the inverse image modulo $\mathfrak{q}$ of the image of $G_\tau \cap K$. For example, when τ is a (regular) diagonal matrix, J_τ consists of elements of K that are congruent modulo $\mathfrak{q}$ to the diagonal.

1.9.2. *Reductions.* The depth aspect case of the volume bound (i.e., $\mathfrak{p}$ fixed) was addressed in [19] following a series of reductions:

(i) **Reduction to the near-identity case**. By [19, §15], it is not necessary to solve Problem 1.10 in general — it suffices to treat the "near-identity" case, where a lies in an arbitrarily small (but fixed) neighborhood of the identity (see §1.9.3).

(ii) **Reduction to the Lie algebra**. The near-identity case reduces further, via Lie-theoretic arguments [19, §16], to a Lie algebra problem, where $\mathrm{Lie}(H_{\tau_H})$ plays the role of H_{τ_H}.

(iii) **Reduction to central directions**. The Lie algebra problem concerns certain sub-spaces of $\mathrm{Lie}(H_{\tau_H})$, whose definition we omit here. The problem is to show that the subspaces are not the entire space. This problem was addressed in [19, §17] by showing that the subspaces do not contain the one-dimensional central subspace; in effect, this shows that central cosets in $H_{\tau_H} \cap K$ suffice to solve Problem 1.10.

These reductions fail in horizontal aspects, for independent reasons:

(i) The near-identity reduction in the depth aspect comes from the freedom to take the convolution kernel ω supported close to the identity, which is unavailable in horizontal aspects (see §1.9.3).

(ii) The reduction to the Lie algebra may be adapted beyond the near-identity case, but the resulting Lie algebra problem is unsolvable: there are natural families of counterexamples (see Theorem 5.5 and Remark 5.16).

(iii) Central directions do not suffice to solve the general case of Problem 1.10; we have again identified counterexamples (see Remark 5.20).

1.9.3. *The near-identity reduction.* We conclude §1.9 by explaining the "near-identity" reduction mentioned above, whose failure in horizontal aspects motivates why the geometric estimates in this paper are of "algebro-geometric" rather than "Lie-algebraic" flavor compared to those in [19].

Fix a small natural number d. Let $K[d]$ denote the dth principal congruence subgroup of K. The convolution kernel ω is supported in K, but not in $K[1]$. It projects onto a "short" family: the integral operator $\pi(\omega)$ vanishes unless the irreducible representation π lies in such a family. The normalized restriction $\omega^{[d]}$ of ω to $K[d]$ projects onto a larger family; in global settings, it is larger by a factor of roughly

$$q^{(n+1)^2 d}. \tag{1.12}$$

Example 1.13. If τ is diagonal, then the operator $\pi(\omega)$ (resp. $\pi(\omega^{[d]})$) vanishes unless π is a principal series representation $\chi_1 \boxplus \cdots \boxplus \chi_{n+1}$ induced by characters χ_j of $F^\times$ having prescribed restrictions to $\mathfrak{o}^\times$ (resp. to $1 + \mathfrak{p}^d$).

The near-identity reduction arises from the freedom to replace ω, supported on $G_\tau \cap K$, with $\omega^{(d)}$, supported on the smaller group $G_\tau \cap K[d]$. We can do so in the depth aspect, where q is fixed, because the factors (1.12) are then harmless.

In horizontal aspects — where q is large — such factors ruin the near-identity reduction: any small power of q saved by amplification is swamped by the large power of q in (1.12), yielding estimates that fail even to recover convexity. We must thus work with projectors ω supported on the full maximal compact subgroup K.

1.10. **Organization of this paper.**

This paper consists almost exclusively of local analysis at the "interesting" place $\mathfrak{p}$. It culminates in our main local result, Theorem 9.1, which is a direct analogue of its archimedean counterpart, [19, Theorem 4.2]. The auxiliary arguments of [19, §4–6] (concerning "uninteresting" places, amplification, counting, etc.) combine with our main local result to yield a subconvex bound, exactly as in [19]. While our local analysis is logically self-contained, its global motivation might be clarified by skimming [19, §6].

We conclude with the detailed breakdown. §2–§3 contain preliminaries. §4 treats "stability" for $(\mathrm{GL}_{n+1}, \mathrm{GL}_n)$ over a general ring. §5 establishes our key algebraic result (Theorem 5.1, or Theorem 1.11 above). §6 applies that result to derive the uniform volume bound (Theorem 6.7), which we further apply in §7 to estimate bilinear forms relevant for (1.11). §8 concerns "microlocal analysis" involving representations π, σ and parameters τ, τ_H. §9 establishes our main local result (Theorem 9.1), a uniform non-archimedean analogue of [19, Theorem 4.2]. §10 combines our main local result with the auxiliary arguments of [19] to complete the proof of Theorem 1.1.

1.11. **Acknowledgment.**

2. NOTATION

We record general notation and conventions, used throughout the paper.

2.1. **Vectors and matrices.**

Let $\mathbf{V}$ be a finite free $\mathbb{Z}$-module, thus $\mathbf{V} \cong \mathbb{Z}^n$ for some n, the *rank* of $\mathbf{V}$. It will occasionally be convenient to denote the rank instead by $n + 1$, so that n refers instead to the rank of a codimension one submodule (see §2.2 below).

We denote by $\mathbf{V}^* = \mathrm{Hom}(\mathbf{V}, \mathbb{Z}) \cong \mathbb{Z}^n$ the dual module and by $\mathbf{M} := \mathrm{End}(\mathbf{V}) \cong \mathrm{Mat}_n(\mathbb{Z})$ the endomorphism ring. We denote by juxtaposition the natural pairing

$$\mathbf{V}^* \otimes \mathbf{V} \to \mathbb{Z}, \quad \ell \otimes v \mapsto \ell v := \ell(v),$$

as well as the left action, right action and outer product, respectively:

$$\begin{aligned}
\mathbf{M} \otimes \mathbf{V} \to \mathbf{V}, && a \otimes v \mapsto av := a(v), \\
\mathbf{V}^* \otimes \mathbf{M} \to \mathbf{V}^*, && \ell \otimes a \mapsto \ell a := [u \mapsto \ell(a(u))] \\
\mathbf{V} \otimes \mathbf{V}^* \to \mathbf{M}, && v \otimes \ell \mapsto v\ell := [u \mapsto \ell(u)v].
\end{aligned}$$

The motivation for the notation is that, by choosing a basis, we may regard $\mathbf{V}$ (resp. $\mathbf{V}^*$) as spaces of column (resp. row) vectors and $\mathbf{M}$ as a space of matrices, in which case the above pairings are given by matrix multiplication. We use the same notation for the extensions of these pairings to more general rings given below.

For us, a *ring* R is a commutative ring with identity, an *algebra* or *ring extension* R' of R is a ring map $R \to R'$, and an *affine scheme* $\mathbf{X}$ is a functor that assigns to each ring R a set $\mathbf{X}(R)$ that functorially identifies with the set of ring maps $\mathbb{Z}[\mathbf{X}] \to R$ for some ring $\mathbb{Z}[\mathbf{X}]$, the *coordinate ring* of $\mathbf{X}$. Affine schemes over R are defined similarly, but restricting to R-algebras; the above discussion then applies upon replacing $\mathbb{Z}$ with R.

By abuse of notation, we denote also by $\mathbf{V}, \mathbf{V}^*$ and $\mathbf{M}$ the affine group schemes over $\mathbb{Z}$ given by

$$\mathbf{V}(R) := \mathbf{V} \otimes R \cong R^n,$$
$$\mathbf{V}^*(R) := \mathbf{V}^* \otimes R \cong R^n,$$
$$\mathbf{M}(R) := \mathrm{End}_R(\mathbf{V}(R)) \cong \mathrm{Mat}_n(R).$$

The coordinate rings of $\mathbf{V}, \mathbf{V}^*$ and $\mathbf{M}$ identify with the polynomial rings over $\mathbb{Z}$ in n, n and n^2 variables, respectively, where n denotes the rank of $\mathbf{V}$. We denote by $\mathbf{G}$ the affine group scheme

$$\mathbf{G}(R) := \mathrm{Aut}_R(\mathbf{V}(R)) \cong \mathrm{GL}_n(R),$$

whose coordinate ring is obtained from that of $\mathbf{M}$ by inverting the determinant.

We denote by $\mathbf{Z}$ the center of $\mathbf{G}$, consisting of scalar matrices:

$$\mathbf{Z}(R) = R^\times \hookrightarrow \mathbf{G}(R).$$

For $g \in \mathbf{G}(R)$, we denote by $\mathrm{Ad}(g) : \mathbf{M}(R) \to \mathbf{M}(R)$ the conjugation map

$$\mathrm{Ad}(g)x := gxg^{-1}.$$

2.2. General linear GGP pairs.

In some parts of this paper, we consider just one general linear group $\mathbf{G}$. In others, we work with inclusions $\mathbf{H} \leq \mathbf{G}$ of general linear groups of neighboring rank. We then use the following notation.

Suppose given $e \in \mathbf{V}$ and $e^* \in \mathbf{V}^*$ such that $e^*e = 1$. We then define the submodule

$$\mathbf{V}_H := \{v \in \mathbf{V} : e^*v = 0\},$$

which participates in the direct sum decomposition

$$\mathbf{V} = \mathbf{V}_H \oplus \mathbb{Z}e, \tag{2.1}$$

$$v = (1 - ee^*)v + (e^*v)e.$$

This extends to any ring. We define affine group schemes

$$\mathbf{V}_H, \quad \mathbf{V}_H^*, \quad \mathbf{M}_H, \quad \mathbf{H}$$

in terms of $\mathbf{V}_H$, by analogy to the definitions of §2.1. These define closed subgroup schemes of $\mathbf{V}, \mathbf{V}^*, \mathbf{M}, \mathbf{G}$, respectively.

For any ring R, we denote by $1_H \in \mathbf{M}_H(R)$ the identity operator. It is given in terms of the identity operator $1 \in \mathbf{M}(R)$ by

$$1_H = 1 - ee^*.$$

Given $\tau \in \mathbf{H}(R)$, we denote by

$$\tau_H := 1_H \tau 1_H \in \mathbf{M}_H(R)$$

the element induced by the decomposition (2.1).

Example 2.1. Suppose $\mathbf{V}$ has rank three. Let e_1, e_2 be a basis for $\mathbf{V}_H$. Then e_1, e_2, e is a basis for $\mathbf{V}$. Using this basis to identify $\mathbf{M}$ with the space of 3×3 matrices, we have

$$\mathbf{M}_H = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad 1_H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\tau = \begin{pmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix} \implies \tau_H = \begin{pmatrix} \tau_{11} & \tau_{12} & 0 \\ \tau_{21} & \tau_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

2.3. **Centralizers.** For $\tau \in \mathbf{M}(R)$ or $\sigma \in \mathbf{M}_H(R)$, we denote by

$$\mathbf{G}_\tau \leq \mathbf{G}, \quad \mathbf{M}_\tau \leq \mathbf{M}, \quad \mathbf{H}_\sigma \leq \mathbf{H}, \quad \mathbf{M}_{H,\sigma} \leq \mathbf{M}_H$$

the centralizers, regarded as affine group schemes over R.

2.4. **Local fields and congruence subgroups.** Let F be a non-archimedean local field. We denote then by

$$\mathfrak{o}, \qquad \mathfrak{p}, \qquad \varpi \in \mathfrak{p}, \qquad q = |\mathfrak{o}/\mathfrak{p}|$$

its ring of integers, its maximal ideal, a uniformizer, and the cardinality of the residue field. We denote by $|.|_F$ the normalized valuation.

For each ideal $\mathfrak{a} \subseteq \mathfrak{o}$, we set

$$K(\mathfrak{a}) := \ker(\mathbf{G}(\mathfrak{o}) \to \mathbf{G}(\mathfrak{o}/\mathfrak{a})), \quad K_H(\mathfrak{a}) := \ker(\mathbf{H}(\mathfrak{o}) \to \mathbf{H}(\mathfrak{o}/\mathfrak{a})), \qquad (2.2)$$

We sometimes write simply

$$K := K(\mathfrak{o}) = \mathbf{G}(\mathfrak{o}), \quad K_H := K_H(\mathfrak{o}) = \mathbf{H}(\mathfrak{o}), \quad K_Z := \mathbf{Z}(\mathfrak{o})$$

for the standard maximal compact subgroups of $\mathbf{G}(F)$, $\mathbf{H}(F)$ and $\mathbf{Z}(F)$, respectively. We equip the latter groups with the Haar measure that assigns volume one to their maximal compact subgroups K_G, K_H and K_Z.

3. CYCLIC MATRICES

Let R be a ring. In this section, we write V, M, H, etc., for the set of R-points of the corresponding bold-faced group schemes.

Let $\tau \in M$. We denote by

$$P_\tau(X) = \det(X - \tau) \in R[X]$$

its characteristic polynomial. It is a monic polynomial whose degree is the rank of $\mathbf{V}$. We recall the Cayley–Hamilton theorem (see [7, Theorem 4.3]): $P_\tau(\tau) = 0$.

Definition 3.1. Let $\tau \in M$. We say that $v \in V$ is τ-*cyclic* if $R[\tau]v = V$. We say that V is τ-*cyclic* (or simply that τ *is cyclic* when V is understood, or that V *is cyclic* when τ is understood) if there exists a τ-cyclic vector $v \in V$.

Example 3.2. It is clear that any Jordan block is cyclic.

Example 3.3. It follows from the Vandermonde determinant calculation that a diagonal matrix is cyclic when the pairwise differences between its diagonal entries are invertible.

Example 3.4. If v is τ-cyclic, then it remains so upon passing to any ring extension R' of R. In particular, the cyclicity of τ is preserved under base extension.

Example 3.5. Let R be a local ring. By Nakayama's lemma, we can test whether τ is cyclic over the residue field. When R is a field, the structure theorem for modules over the principal ideal domain $R[X]$ says that V is isomorphic to a direct sum $\oplus_i R/(p_i)$, where p_i are monic primary polynomials (i.e., powers of irreducible polynomials). Then τ is cyclic if and only if the p_i are pairwise relatively prime. For instance, when R is algebraically closed, this says that no two Jordan blocks for τ have the same eigenvalue.

Lemma 3.6. *Write $n = \mathrm{rank}(\mathbf{V})$. Then v is τ-cyclic if and only if the map of modules*

$$R^n \to V \tag{3.1}$$

$$(c_0, \ldots, c_{n-1}) \mapsto \sum_{j<n} c_j \tau^j v \tag{3.2}$$

is surjective, in which case it is an isomorphism.

Proof. The first statement follows from the Cayley–Hamilton theorem, the second from the fact that any surjective map of free modules of the same finite rank is an isomorphism [7, Corollary 4.4a]. □

In particular, each cyclic element τ admits a *cyclic basis* $e_1, \ldots, e_n$ such that

$$\tau e_j = e_{j+1} \text{ for } j < n, \quad \text{e.g., } \tau = \begin{pmatrix} 0 & 0 & * \\ 1 & 0 & * \\ 0 & 1 & * \end{pmatrix}. \tag{3.3}$$

Lemma 3.7. *Let $\tau \in M$, and suppose that $v \in V$ is τ-cyclic. Then the map*

$$M_\tau \to V, \qquad x \mapsto xv$$

is an isomorphism of R-modules. Moreover,

$$M_\tau = R[\tau] = \left\{ \sum_{j=0}^{n-1} c_j \tau^j : (c_0, \ldots, c_{n-1}) \in R^n \right\}.$$

In particular, M_τ is a commutative R-algebra, and G_τ is an abelian group.

Proof. Since V is spanned by the $\tau^j v$ and $x\tau^j v = \tau^j xv$, we see that x is determined by xv. The map $M_\tau \to V$ is thus injective.

We next verify that $M_\tau = R[\tau]$. Let $x \in M_\tau$. By Lemma 3.6, we may choose $(c_0, \ldots, c_{n-1}) \in R^n$ so that $xv = \sum_{j=0}^{n-1} c_j \tau^j v$. By the noted injectivity of $M_\tau \to V$, it follows that $x = \sum_{j=0}^{n-1} c_j \tau^j$. Thus $M_\tau \subseteq R[\tau]$. The reverse containment is evident.

The surjectivity of $M_\tau \to V$ follows from the identity $M_\tau = R[\tau]$ and Lemma 3.6. □

Lemma 3.8. *Suppose that $\tau_1, \tau_2 \in M$ are cyclic elements with the same characteristic polynomial. Then they are conjugate under G.*

Proof. Let $e_1^{(j)}, \ldots, e_n^{(j)}$ be a cyclic basis for τ_1 (resp. τ_2). By Cayley–Hamilton, we have $\tau_j e_n^{(j)} = \sum_{i=1}^n c_i e_i^{(j)}$, where the coefficients $c_i \in R$ depend only upon the characteristic polynomial, hence are the same for τ_1 and τ_2. Let $g \in G$ be the unique map that sends $e_i^{(1)}$ to $e_i^{(2)}$. Then $g\tau_1 = \tau_2 g$, so τ_1 and τ_2 are conjugate. □

4. STABILITY

Definition 4.1. Let R be a ring and $\tau \in \mathbf{M}(R)$. We say that τ is *stable* if the vectors e and e^* are both τ-cyclic. We denote by $\mathbf{M}_{\mathrm{stab}}(R)$ the subset of stable elements.

Remark 4.2. This condition was studied by Rallis–Schiffmann [21] under the name "regular", and also in [20, §14], where it was related (for R a field of characteristic zero) to the geometric invariant theory notion of stability. See especially [21, Thm 6.1] and [20, Lem 14.8]. Many of the results of this section were established in [20, §13-14] when R is a field of characteristic zero, and in some respects, more systematically and generally (e.g., also for orthogonal GGP pairs). We have chosen to give short proofs of what we require, repeating or adapting the arguments of the cited reference. The adapted arguments could likely be extended to the orthogonal case along the same lines as in [20, §14].

Example 4.3. Suppose R is a local ring, with maximal ideal $\mathfrak{p}$. Let $\tau \in \mathbf{M}(R)$, with image $\bar{\tau} \in \mathbf{M}(R/\mathfrak{p})$. Nakayama's lemma implies that τ is stable if and only if $\bar{\tau}$ is stable.

We denote by $n+1$ the rank of $\mathbf{V}$, so that n is the rank of $\mathbf{V}_H$. For a ring R, we define

$$\Delta : \mathbf{M}(R) \to R$$

$$\Delta(\tau) := \det(e^*, e^*\tau, \ldots, e^*\tau^n) \det(e, \tau e, \ldots, \tau^n e),$$

where $\det(\cdots)$ denotes the determinant of the matrix having the indicated row vectors $e^*\tau^i \in \mathbf{V}^*(R)$ or column vectors $\tau^j e \in \mathbf{V}(R)$, with such vectors realized as $(n+1)$-tuples using some chosen basis for the free $\mathbb{Z}$-module $\mathbf{V}$. Choosing a different basis has the effect of multiplying the two determinants by mutually inverse factors, hence has no effect on $\Delta(\tau)$. This family of maps Δ defines a regular function on the scheme $\mathbf{M}$. It follows from Lemma 3.6 that τ is stable if and only if $\Delta(\tau)$ is a unit in R. In particular, $\mathbf{M}_{\mathrm{stab}}$ defines an open affine subscheme of $\mathbf{M}$.

In the remainder of this section, we focus on an individual (arbitrary) ring R and abbreviate $V := \mathbf{V}(R), G := \mathbf{G}(R)$, etc.

For $\tau \in M$, we may form the characteristic polynomials

$$P_\tau \in R[X] \quad \text{and} \quad P_{\tau_H} \in R[X].$$

These are monic polynomials of degrees $n+1$ and n, respectively.

Lemma 4.4. Let $P, P_H \in R[X]$ be monic polynomials of degrees $n+1$ and n, respectively. There exists $\tau \in M$ such that $P_\tau = P$ and $P_{\tau_H} = P_H$.

Proof. One can deduce the existence of τ directly from the spherical property of $H \hookrightarrow G \times H$ (see [20, §14.6]). For variety of exposition, we record an "explicit" construction, following [24, §6.2]. Supposing for instance that $H = \mathrm{GL}_2(R) \hookrightarrow G = \mathrm{GL}_3(R)$, embedded as the upper-left block, we take

$$\tau = \begin{pmatrix} a_1 & 1 & 0 \\ a_2 & 0 & 1 \\ b_2 & b_1 & b_0 \end{pmatrix}.$$

The coefficients of $\mathrm{charpoly}(\tau_H)$ are then $\pm a_j$, while the coefficients of $\mathrm{charpoly}(\tau)$ are of the form $\pm(b_j + c_j)$, where c_j depends upon the a_k and b_l for $l < j$. By choosing suitable a_j and b_j inductively, we may thus arrange that τ and τ_H have prescribed characteristic polynomials. $\square$

Lemma 4.5. For $\tau \in M$, the following are equivalent.

(i) P_τ and P_{τ_H} generate the unit ideal.

(ii) There are no nonzero τ-invariant submodules of V_H or of V_H^*.
(iii) $\tau \in M_{\text{stab}}$.

Proof. We show first that (i) implies (ii). By assumption, we may write $aP_\tau + bP_{\tau_H} = 1$ for some $a, b \in R[X]$. Since $P_\tau(\tau) = 0$, it follows that $b(\tau)P_{\tau_H}(\tau) = 1$. On the other hand, if U is a τ-invariant submodule of V_H and $u \in U$, then $f(\tau)u = f(\tau_H)u$ for each $f \in R[x]$, hence $P_{\tau_H}(\tau)u = P_{\tau_H}(\tau_H)u = 0$. Therefore $u = 0$, as required.

We show next that (ii) implies (iii). We must check that e and e^* are τ-cyclic. Suppose otherwise that $R[\tau]e \neq V$ or that $e^*R[\tau] \neq V^*$. In the first case, we set $U := (R[\tau]e)^\perp \subseteq V^*$; in the second, $U := (e^*R[\tau])^\perp \subseteq V$. In either case, we see that U is a non-zero τ-invariant submodule of V_H^* or V_H, contrary to (ii).

We show finally that (iii) implies (i). Suppose that τ is stable, but that the ideal $\mathfrak{a} \subseteq R[X]$ generated by P_τ and P_{τ_H} is not the unit ideal (1). Let $\mathfrak{m}$ be a maximal ideal that contains $\mathfrak{a}$, and set $\mathfrak{p} := R \cap \mathfrak{m}$. The stability of τ and the fact that $\mathfrak{a} \neq (1)$ are unaffected by modding out by $\mathfrak{p}$, passing to the field of fractions of $R/\mathfrak{p}$, and then passing to the algebraic closure of that field. We thereby reduce to the case that R is an algebraically closed field. Then P_τ and P_{τ_H} share a common root c. The remainder of the proof closely follows that of [20, Lemma 14.4]. We may find

- an eigenvector $v \in V_H$ for τ_H with eigenvalue c, and also
- an eigenvector $\ell \in V^*$ for τ with eigenvalue c.

Then
$$\ell \cdot (\tau - c) \cdot v = 0. \tag{4.1}$$
We consider two cases.

- $(\tau - c)v = 0$, so that $v \in V_H$ is an eigenvector for τ.
- $(\tau - c)v \neq 0$. Since $(\tau_H - c)v = 0$, it follows that e is a multiple of $(\tau - c)v$, hence by (4.1) that $\ell e = 0$, so that $\ell \in V_H^*$ is an eigenvector for τ.

In either case, we have produced an eigenvector for τ in V_H or in V_H^*. Since R is a field, the orthogonal complement of that eigenvector defines a proper τ-invariant subspace of V^* (resp. V) that contains the vector e^* (resp. e), contrary to our assumption that that vector is cyclic. $\square$

Lemma 4.6. Let $\tau \in M_{\text{stab}}$. Then $H \cap G_\tau = \{1\}$.

Proof. This follows from the fact that e is τ-cyclic and fixed by H, which shows that if $h \in H$ commutes with τ, then it fixes all of V, hence $h = 1$. $\square$

Lemma 4.7. Let $\tau \in M_{\text{stab}}$. Then τ_H is cyclic, that is to say, V_H and, equivalently, V_H^*, are τ_H-cyclic.

Proof. We will show that V_H^* is cyclic; a similar argument applies to V_H. We will verify more precisely that
$$\ell : V_H \to R, \qquad \ell(v) := e^*\tau v$$
defines a τ_H-cyclic vector $\ell \in V_H^*$, that is to say,
$$\ell R[\tau_H] = V_H^*. \tag{4.2}$$
In verifying this, we may assume that R is a local ring, and then (by Nakayama's lemma) that R is a field. In that case, if (4.2) fails, then the orthogonal complement of $\ell R[\tau_H]$ is a nonzero τ_H-invariant subspace U of V_H. By construction, $0 = \ell(U) = e^*\tau U$, so $\tau|_U = \tau_H|_U$. Thus U is in fact a τ-invariant subspace of V_H, which by part (ii) of Lemma 4.5 yields $U = 0$, giving the required contradiction. $\square$

5. TRANSVERSALITY

5.1. **Statement of results.** For $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$, $a \in \mathbf{G}(R)$, we define a subfunctor $\mathbf{X}_{\tau,a}$ of $\mathbf{H}_{\tau_H}$ over R, as follows: for each ring extension R' of R,

$$\mathbf{X}_{\tau,a}(R') := \{y \in \mathbf{H}_{\tau_H}(R') : ay \in \mathbf{H}(R')\mathbf{G}_\tau(R')\}.$$

As we explain below (Lemma 5.8), it defines a closed subscheme.

We observe that if a lies in $\mathbf{H}(R)\mathbf{Z}(R)$, then $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$, that is, their point sets coincide for all ring extensions R' of R. Indeed, if $a = hz$ with $(h, z) \in \mathbf{H}(R)\mathbf{Z}(R)$, then for each $y \in \mathbf{H}_{\tau_H}(R')$, we have

$$ay = hzy = (hy)z \in \mathbf{H}(R')\mathbf{G}_\tau(R').$$

The main result of this section is the following converse:

Theorem 5.1. *Assume that* $\mathrm{rank}(\mathbf{V}) \geq 3$. *Let R be a ring in which 2 is a unit. Let* $(\tau, a) \in \mathbf{M}_{\mathrm{stab}}(R) \times \mathbf{G}(R)$. *Suppose that* $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$ *(as schemes over R). Then* $a \in \mathbf{H}(R)\mathbf{Z}(R)$.

Remark 5.2 (Sharpness with respect to the rank)**.** The conclusion of Theorem 5.1 clearly fails when $\mathrm{rank}(\mathbf{V}) = 1$, since in that case, $\mathbf{H}_{\tau_H}$ is trivial. The conclusion may fail also when $\mathrm{rank}(\mathbf{V}) = 2$. For example, take $\mathbf{V} = \mathbb{Z}^2$ with standard basis e_1, e_2. Let e_1^*, e_2^* denote the corresponding dual basis of $\mathbf{V}^*$. Take $e := e_2$ and $e^* := e_2^*$, so that

$$\mathbf{H} = \begin{pmatrix} * & 0 \\ 0 & 1 \end{pmatrix} \leq \mathbf{G}.$$

Take

$$\tau := a := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then for any ring R, we see that a does not lie in $\mathbf{H}(R)\mathbf{Z}(R)$ (although it does normalize it). On the other hand, for any

$$y = \begin{pmatrix} y_1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbf{H}_{\tau_H}(R) = \mathbf{H}(R),$$

we have

$$ay = hb \quad \text{with } h := \begin{pmatrix} y_1^{-1} & 0 \\ 0 & 1 \end{pmatrix} \in \mathbf{H}(R), \quad b := \begin{pmatrix} 0 & y_1 \\ y_1 & 0 \end{pmatrix} \in \mathbf{G}_\tau(R).$$

Thus $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$. The proof of Theorem 5.1 can be adapted to show that these are essentially the only counterexamples when $\mathrm{rank}(\mathbf{V}) = 2$. This discussion suggests that Theorem 5.1 should be related to the fact that $\mathbf{HZ}$ has trivial normalizer in $\mathbf{G}$ when $\mathrm{rank}(\mathbf{V}) \geq 3$, although we did not spot a direct way to relate the two conditions.

Remark 5.3 (Sharpness with respect to the ring)**.** We have seen by computer calculation with Gröbner bases that when $\mathrm{rank}(\mathbf{V}) \in \{3, 4\}$, there exist fields R of characteristic 2 over which there are counterexamples to the conclusion of Theorem 5.1. We see no reason why such examples should not exist in any rank.

Theorem 5.1 is a consequence of some more precise assertions that we now formulate.

Definition 5.4. Let R be a ring. We define the extension rings

$$R' := R[\varepsilon]/(\varepsilon^2), \quad R'' := R[\varepsilon_1, \varepsilon_2]/(\varepsilon_1^2, \varepsilon_2^2),$$

which come with natural maps $R' \to R$ and $R'' \to R$ obtained by sending $\varepsilon, \varepsilon_1, \varepsilon_2$ to zero. For a scheme $\mathbf{X}$, we say that $y' \in \mathbf{X}(R')$ lies over $y \in \mathbf{X}(R)$ if $y' \mapsto y$ under the induced map $\mathbf{X}(R') \to \mathbf{X}(R)$, and similarly for R''.

Let $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$, $a \in \mathbf{G}(R)$ and $y \in \mathbf{X}_{\tau,a}(R)$.

(1) We say that $\mathbf{X}_{\tau,a}$ is *tangential at y over R* if for all $y' \in \mathbf{H}_{\tau_H}(R')$ lying over y, we have $y' \in \mathbf{X}_{\tau,a}(R')$.
(2) We say that $\mathbf{X}_{\tau,a}$ is *doubly-tangential at y over R* if for all $y'' \in \mathbf{H}_{\tau_H}(R'')$ lying over y, we have $y'' \in \mathbf{X}_{\tau,a}(R'')$.

The informal content is that $\mathbf{X}_{\tau,a}$ is tangential (resp. doubly-tangential) at y over R if the polynomials whose vanishing defines $\mathbf{X}_{\tau,a}$ over R have the property that their Taylor series at y vanish to order at least 2 (resp. 3).

Theorem 5.5. *Retain the setting of Definition 5.4. The following are equivalent:*

(i) $\mathbf{X}_{\tau,a}$ *is tangential at y over R.*
(ii) *For the (unique) elements $(h, b) \in \mathbf{H}(R) \times \mathbf{G}_\tau(R)$ defined by writing $ay = hb$, we have $b^2 \in \mathbf{Z}(R)$.*

Theorem 5.6. *Retain the setting of Definition 5.4. Assume that*

- *2 is a unit in R, and*
- $\mathrm{rank}(\mathbf{V}) \geq 3$.

Then the following are equivalent:

(i) $\mathbf{X}_{\tau,a}$ *is doubly-tangential at y over R.*
(ii) *We have $a \in \mathbf{H}(R)\mathbf{Z}(R)$.*

Theorem 5.6 implies in particular that, under the hypotheses of Theorem 5.1, if $a \notin \mathbf{H}(R)\mathbf{Z}(R)$, then $\mathbf{X}_{\tau,a}$ fails to be doubly-tangential at every point $y \in \mathbf{X}_{\tau,a}(R)$. In particular, $\mathbf{X}_{\tau,a} \neq \mathbf{H}_{\tau_H}$, so the conclusion of Theorem 5.1 holds. The remainder of this section is devoted to the proofs of Theorems 5.5 and 5.6.

5.2. **Defining $\mathbf{X}_{\tau,a}$ by polynomial equations.** Let R be a ring, and let $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$. By Lemma 3.7, the maps

$$\bullet e : \mathbf{M}_\tau(R) \to \mathbf{V}(R), \qquad\qquad x \mapsto xe$$
$$e^* \bullet : \mathbf{M}_\tau(R) \to \mathbf{V}^*(R), \qquad\qquad x \mapsto e^* x$$

are linear isomorphisms (i.e., isomorphisms of R-modules). We denote by

$$[\bullet e]^{-1} : \mathbf{V}(R) \to \mathbf{M}_\tau(R),$$
$$[e^* \bullet]^{-1} : \mathbf{V}^*(R) \to \mathbf{M}_\tau(R)$$

their inverses. These extend to linear isomorphisms over any ring extension R' of R.

We note the following equivariance property:

Lemma 5.7. *For any $v \in \mathbf{V}(R)$ (resp. $v^* \in \mathbf{V}^*(R)$) and $a \in \mathbf{G}_\tau(R)$, we have*

$$[\bullet e]^{-1}(av) = a[\bullet e]^{-1}(v), \qquad [e^* \bullet]^{-1}(v^* a) = [e^* \bullet]^{-1}(v^*)a.$$

Proof. By definition, $[\bullet e]^{-1}(v)$ is the unique element of $\mathbf{M}_\tau(R)$ with $[\bullet e]^{-1}(v)e = v$, while $[\bullet e]^{-1}(av)$ is the unique element with $[\bullet e]^{-1}(av)e = av$. It is clear then that $a[\bullet e]^{-1}(v) = [\bullet e]^{-1}(av)$. A similar argument gives the second identity. $\square$

Lemma 5.8. *For each ring extension R' of R, we have*

$$\mathbf{H}(R')\mathbf{G}_\tau(R') = \left\{ g \in \mathbf{G}(R') : [e^*\bullet]^{-1}(e^*g)[\bullet e]^{-1}(g^{-1}e) = 1 \right\}. \qquad (5.1)$$

In particular, $R' \mapsto \mathbf{H}(R')\mathbf{G}_\tau(R')$ defines a closed subscheme of $\mathbf{G}$ over R.

Proof. Suppose $g = hb$ with $(h, b) \in \mathbf{H}(R') \times \mathbf{G}_\tau(R')$. Then $g^{-1}e = b^{-1}e$ and $e^*g = e^*b$, so

$$[e^*\bullet]^{-1}(e^*g)[\bullet e]^{-1}(g^{-1}e) = bb^{-1} = 1.$$

Conversely, suppose g belongs to the right hand side of (5.1), Then, defining $b, b' \in \mathbf{M}_\tau(R')$ by $b' := [\bullet e]^{-1}(g^{-1}e)$ and $b := [e^*\bullet]^{-1}(e^*g)$, we have $b'b = 1$. It follows that $b \in \mathbf{G}_\tau(R')$. Moreover, $h := gb^{-1}$ satisfies $he = e$ and $e^*h = e^*$, hence $h \in \mathbf{H}(R')$. $\qquad\square$

Corollary 5.9. *Let*

$$f : \mathbf{H}_{\tau_H} \to \mathbf{M}_\tau$$

denote the map of schemes over R defined by

$$f(y) := [e^*\bullet]^{-1}(e^*ay)[\bullet e]^{-1}((ay)^{-1}e).$$

Then

$$\mathbf{X}_{\tau,a}(R') = \{ y \in \mathbf{H}_{\tau_H}(R') : f(y) = 1 \}.$$

In particular, $\mathbf{X}_{\tau,a}$ defines a closed subscheme of $\mathbf{H}_{\tau_H}$ over R.

Remark 5.10. We may rewrite the defining equation $f(y) = 1$ for $\mathbf{X}_{\tau,a}$ using Cramer's rule for y^{-1}, as follows:

$$[e^*\bullet]^{-1}(e^*ay)[\bullet e]^{-1}(y^{\mathrm{adj}}a^{-1}e) = \det(y).$$

Here y^{adj} denotes the adjugate or cofactor matrix, characterized by the identity $yy^{\mathrm{adj}} = \det(y)$. Writing $n + 1 = \mathrm{rank}(\mathbf{V})$, so that $\mathbf{M}_\tau$ has rank $n + 1$, one can see that $\mathbf{X}_{\tau,a}$ is defined by $n + 1$ polynomial equations of degree at most n in the entries of y. Indeed, one can parameterize $\mathbf{H}_{\tau_H}$ using Lemma 3.7 with n variables, say, $y_1, \cdots y_n$. Then

- each entry of the row vector e^*ay is an affine function of the y_i (i.e., a linear combination of the y_i and 1),
- the coefficients of $[e^*\bullet]^{-1}(e^*ay) \in \mathbf{M}_\tau$ are likewise affine functions of the y_i,
- the matrices $y^{\mathrm{adj}}, y^{\mathrm{adj}}a^{-1}e$ and $[\bullet e]^{-1}(y^{\mathrm{adj}}a^{-1}e)$ have entries given by polynomials in the y_i of total degree at most $n - 1$, and
- $\det(y)$ is a polynomial in the y_i of total degree n.

Finally, the condition

$$[e^*\bullet]^{-1}(e^*ay)[\bullet e]^{-1}(y^{\mathrm{adj}}a^{-1}e) = \det(y)$$

is an equality of elements of $\mathbf{G}_\tau$, whose $n + 1$ coefficients give rise to $n + 1$ polynomial equations of degree at most n in the y_i.

5.3. **Reduction to the centralizer.** The set $\mathbf{X}_{\tau,a}(R)$ is empty unless we can find $h \in \mathbf{H}(R)$ and $y \in \mathbf{H}_{\tau_H}(R)$ so that $b := hay \in \mathbf{G}_\tau(R)$. Then for all ring extensions R' of R, we have

$$\mathbf{X}_{\tau,a}(R') = y\mathbf{X}_{\tau,b}(R').$$

Furthermore, for each point $z \in \mathbf{X}_{\tau,b}(R)$, the following are equivalent:

(1) $\mathbf{X}_{\tau,a}$ is tangential (resp. doubly-tangential) at yz.
(2) $\mathbf{X}_{\tau,b}$ is tangential (resp. doubly-tangential) at z.

This last equivalence holds in particular for $z = 1$, which lies in $\mathbf{X}_{\tau,b}(R)$ because b lies in $\mathbf{G}_\tau(R)$. In this way, many questions concerning the $\mathbf{X}_{\tau,a}$ may be reduced to the case $a \in \mathbf{G}_\tau(R)$ and $y = 1$.

Definition 5.11. Given $a \in \mathbf{G}_\tau(R)$, we denote by

$$\mu, \nu : \mathbf{M}_{H,\tau_H}(R) \to \mathbf{M}_\tau(R)$$

the linear maps given by

$$\mu(u) := [e^* \bullet]^{-1}(e^* a u a^{-1}),$$
$$\nu(u) := [\bullet e]^{-1}(a u a^{-1} e).$$

We may use the equivariance property noted in Lemma 5.7 to rewrite the map f from Corollary 5.9 as

$$f(y) = \mu(y)\nu(y^{-1}).$$

5.4. **Linear analysis.** In this section, we prove Theorem 5.5. By replacing a with ay, we may reduce to the case $y = 1$. By then replacing a with $h^{-1}a$, we may reduce further to the case that $a = b \in \mathbf{G}_\tau(R)$. The required equivalence then follows from that between (i) and (iv) in Proposition 5.12, below.

Proposition 5.12. *Let R be a ring, $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$, $a \in \mathbf{G}_\tau(R)$. The following are equivalent:*

(i) *$\mathbf{X}_{\tau,a}$ is tangential at 1 over R.*
(ii) *The maps μ and ν attached to a as in §5.3 satisfy*

$$\mu = \nu. \tag{5.2}$$

(iii) *For $j \geq 0$, define*

$$A_j := e^* a \tau^j e, \quad B_j := e^* a^{-1} \tau^j e.$$

Then

$$A_j a^{-1} = B_j a. \tag{5.3}$$

(iv) *We have $a^2 \in \mathbf{Z}(R)$.*

The proof uses the following key lemma.

Lemma 5.13. There exist $c_{ij} \in M_\tau$ ($0 \leq i \leq j$), with $c_{ii} = 1$, so that for each $j \geq 0$,

$$\mu(\tau_H^j) - \nu(\tau_H^j) = \sum_{i \leq j} c_{ij} \left(B_i a - A_i a^{-1} \right). \tag{5.4}$$

Here and henceforth, we adopt the convention

$$\tau_H^0 := 1_H.$$

Proof. We induct on j. We consider first the case $j = 0$. By definition, $\nu(1_H)$ is the element of $\mathbf{M}_\tau(R)$ such that

$$\nu(1_H)e = a 1_H a^{-1} e.$$

Expanding out $1_H = 1 - ee^*$ gives

$$a 1_H a^{-1} e = e - a e e^* a^{-1} e.$$

Recognizing $e^* a^{-1} e$ as the scalar B_0, we deduce that

$$\nu(1_H)e = (1 - B_0 a)e.$$

Since $1 - B_0 a \in \mathbf{M}_\tau(R)$, it follows from the definition of ν that

$$\nu(1_H) = 1 - B_0 a.$$

A similar argument gives

$$\mu(1_H) = 1 - A_0 a^{-1}. \tag{5.5}$$

Thus the required identity (5.4) holds for $j = 0$ (with $c_{00} = 1$).

We now aim to verify (5.4) for given $j \geq 1$, assuming that it holds for all smaller values of j. We expand

$$\tau_H^j = (1 - ee^*)\tau(1 - ee^*)\tau \cdots (1 - ee^*)\tau(1 - ee^*)$$
$$= \tau^j - ee^*\tau^j - \tau^j ee^* + \cdots,$$

where $\cdots$ denotes a linear combination of terms of the form

$$\tau^{j_1} ee^* \tau^{j_2} \quad (0 \leq j_1, j_2 < j) \tag{5.6}$$

that is symmetric with respect to j_1 and j_2. For instance, when $j = 1$, we have

$$\cdots = ee^*\tau ee^* = e(e^*\tau e)e^* = (e^*\tau e)ee^*,$$

which is a multiple of the vector ee^* (i.e., (5.6) with $j_1 = j_2 = 0$). Arguing like in the $j = 0$ case, we obtain

$$\mu(\tau_H^j) = \tau^j - A_j a^{-1} - A_0 a^{-1}\tau^j + \cdots, \tag{5.7}$$
$$\nu(\tau_H^j) = \tau^j - B_j a - B_0 a\tau^j + \cdots,$$

where $\cdots$ denotes a linear combination of terms given in the first case by $A_{j_1}\tau^{j_2} a^{-1}$ and in the second by $B_{j_1} a\tau^{j_2}$. Taking differences, we obtain

$$\mu(\tau_H^j) - \nu(\tau_H^j) = (B_j a - A_j a^{-1}) + \tau^j(B_0 a - A_0 a^{-1}) + \cdots,$$

where $\cdots$ denotes a linear combination of differences

$$\tau^{j_2}(B_{j_1} a - A_{j_1} a^{-1}).$$

with $j_1, j_2 < j$. $\qquad\qquad\square$

Proof of Proposition 5.12. We check first that (i) is equivalent to (ii). For $u \in \mathbf{M}_{H,\tau_H}(R)$, set

$$y := 1 + \varepsilon u \in \mathbf{H}_{\tau_H}(R').$$

Then $y^{-1} = 1 - \varepsilon u$, so

$$[e^* \bullet]^{-1}(e^* a y) = a + \varepsilon[e^* \bullet]^{-1}(e^* a u),$$
$$[\bullet e]^{-1}((ay)^{-1} e) = a^{-1} - \varepsilon[\bullet e]^{-1}(u a^{-1} e).$$

Multiplying these together and applying the equivariance properties noted in Lemma 5.7, we obtain

$$f(y) = 1 + \varepsilon\left([e^* \bullet]^{-1}(e^* a u a^{-1}) - [\bullet e]^{-1}(a u a^{-1} e)\right)$$
$$= 1 + \varepsilon\left(\mu(u) - \nu(u)\right).$$

Thus $f(y) = 1$ for all u if and only if the identity (5.2) holds.

To verify the equivalence between (ii) and (iii), observe first that powers τ_H^j for $j \geq 0$ span $\mathbf{M}_{H,\tau_H}(R)$, so the identity (5.2) is equivalent to the identity

$$\mu(\tau_H^j) = \nu(\tau_H^j) \tag{5.8}$$

holding for all $j \geq 0$. By inductive application of Lemma 5.13, we see that (5.8) holds for all j if and only if (5.3) does. The required equivalence follows.

We verify finally that (iii) and (iv) are equivalent. The relation (5.3) maybe be rewritten

$$B_j a^2 = A_j 1_G.$$

For notational clarity in what follows, let us choose a basis of $\mathbf{V}(R)$, so that we may describe a^2 and 1_G by their matrix entries a_{ij} and δ_{ij}. Let us also define the row vectors $A = (A_0, \ldots, A_n)$ and $B = (B_0, \ldots, B_n)$. Then we further rewrite

$$(a^2)_{ij} B = \delta_{ij} A. \tag{5.9}$$

Let P denote the matrix with columns $e, \tau e, \ldots, \tau^n e$. Since $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$, the matrix P is invertible. We compute

$$A = e^* a P, \quad B = e^* a^{-1} P,$$
$$BP^{-1} ae = e^* e = 1,$$
$$AP^{-1} ae = e^* a^2 e.$$

Multiplying the relation (5.9) on the right by $P^{-1} ae$ gives

$$(a^2)_{ij} = \delta_{ij} e^* a^2 e,$$

hence

$$a^2 = (e^* a^2 e) 1_G \in \mathbf{Z}(R).$$

This shows that (iii) implies (iv). The converse may be verified by reversing the above steps, or directly from the definitions. $\qquad\square$

Remark 5.14. The arguments used to establish Theorem 5.5 lead to a much simpler proof of the following generalizations of Theorem [19, Theorem 16.3]: *for each ring R with $2 \in R^\times$, each $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$ and $x \in \mathbf{M}_\tau(R) - R$, we have*

$$[x, 1_H] \notin \mathbf{M}_H(R) + \mathbf{M}_\tau(R). \tag{5.10}$$

(The quoted result asserts the same conclusion when R is a field of characteristic zero.) To see this, suppose otherwise that $[x, 1_H] \in t + \mathbf{M}_H(R)$ for some $t \in \mathbf{M}_\tau(R)$. Then

$$e^* t = e^* [x, 1_H], \quad te = [x, 1_H] e. \tag{5.11}$$

Using that $1_H = 1 - ee^*$ and identities like $e^* xee^* = e^*(e^* xe)$, we compute that

$$e^* [x, 1_H] = e^*(x - e^* xe), \quad [x, 1_H]e = -(x - e^* xe)e. \tag{5.12}$$

The stability of τ implies that $t \in \mathbf{M}_\tau(R)$ is determined by either of the identities in (5.11). Since both factors in (5.12) lie in $\mathbf{M}_\tau(R)$, it follows that

$$x - e^* xe = t = -x + e^* xe,$$

which simplifies to $2x = 2e^* xe$. Since $2 \in R^\times$, we deduce that $x = e^* xe \in R$, contrary to assumption.

Remark 5.15. There are many examples of τ for which we have the implication

$$a \in \mathbf{G}_\tau(R), \, a^2 \in \mathbf{Z}(R) \implies a \in \mathbf{Z}(R).$$

For instance, this implication holds if R is a field of characteristic $\neq 2$ and a is a cyclic element that is not semisimple. Theorem 5.5 implies that the conclusion of Theorem 5.1 holds for such τ.

Remark 5.16. The basic cases where Theorem 5.5 on its own does not suffice to prove Theorem 5.1 are when, with respect to some basis, τ and a are both diagonal, the entries of a equal to ± 1 but not all the same (cf. Remark 5.20). When R is a field, the stability assumption then says that e and e^* are vectors (with inner product $e^*e = 1$) all of whose entries are nonzero.

5.5. **Quadratic analysis.** Here we prove Theorem 5.6, following several preparatory lemmas.

Lemma 5.17. Let R be ring, $\tau \in \mathbf{M}_{\mathrm{stab}}(R)$ and $a \in \mathbf{G}_\tau(R)$. Assume that $\mathbf{X}_{\tau,a}$ is doubly-tangential at 1 over R; in particular, it is tangential, so by Proposition 5.12, the linear maps μ and ν from §5.3 coincide. Then

$$2\left(\mu(u_1 u_2) - \mu(u_1)\mu(u_2)\right) = 0 \quad \text{for all } u_1, u_2 \in \mathbf{M}_{H,\tau_H}(R). \tag{5.13}$$

Proof. With $y := 1 + \varepsilon_1 u_1 + \varepsilon_2 u_2$, we see by a short calculation that

$$y^{-1} = 1 - \varepsilon_1 u_1 - \varepsilon_2 u_2 + 2\varepsilon_1\varepsilon_2 u_1 u_2,$$

hence

$$\mu(y) = 1 + \varepsilon_1 \mu(u_1) + \varepsilon_2 \mu(u_2).$$
$$\nu(y^{-1}) = 1 - \varepsilon_1 \nu(u_1) - \varepsilon_2 \nu(u_2) + 2\varepsilon_1\varepsilon_2 \nu(u_1 u_2),$$

By our doubly-tangential hypothesis, we have the relation $1 = f(y) = \mu(y)\nu(y^{-1})$. Expanding this out and using that $\mu = \nu$ gives

$$2\mu(u_1 u_2) = \mu(u_1)\mu(u_2) + \mu(u_2)\mu(u_1).$$

The claimed identity follows now from the commutativity of $\mathbf{M}_\tau(R)$ (Lemma 3.7). □

Lemma 5.18. Under the same hypotheses as Lemma 5.17, and with $A_0 := e^*ae$ as in Proposition 5.12, we have

$$2A_0(a - A_0) = 0. \tag{5.14}$$

Proof. We specialize (5.13) to $u_1 = u_2 = 1_H$. The identity (5.5) for $\mu = \nu$ gives

$$\mu(1_H) = 1 - A_0 a^{-1}.$$

A short calculation then gives

$$A_0(a - A_0) = a^2(\mu(1_H) - \mu(1_H)^2).$$

The conclusion now follows from (5.13). □

Remark 5.19. We pause to sketch a proof of Theorem 5.1 in the special case

$$A_0 = A_1 = 0 \tag{5.15}$$

by an argument that seems more illuminating than the general argument given below. (We note that the hypothesis (5.15) excludes the case $a \in \mathbf{H}(R)\mathbf{Z}(R)$, since then A_0 is a unit.) Since $A_0 = 0$, we have

$$\mu(1_H) = 1.$$

Since $A_1 = 0$, we see from the formula (5.7) (specialized to $j = 1$) that

$$\mu(\tau_H) = \tau.$$

Since $2 \notin R^\times$, we deduce from (5.13) the homomorphism property

$$\mu(u_1)\mu(u_2) = \mu(u_1 u_2).$$

By iterating this and applying the Cayley–Hamilton theorem to τ_H, we obtain
$$0 = \mu(P_{\tau_H}(\tau_H)) = P_{\tau_H}(\mu(\tau_H)) = P_{\tau_H}(\tau).$$
Since τ is stable, the characteristic polynomials P_τ and P_{τ_H} generate the unit ideal (Lemma 4.5), contradicting the Cayley–Hamilton theorem for τ.

Remark 5.20. The argument sketched in 5.19, while aesthetically pleasing, does not suffice to establish the general case of Theorem 5.1. We record a counterexample with $\text{rank}(\mathbf{V}) = 6$. Take $R = \mathbb{Q}(\alpha)$ with $\alpha^2 = 2$,
$$\tau = \text{diag}(0, 1, 2, 2\alpha, 1 + 2\alpha, 2 + 2\alpha),$$
$$e^* = (1, 1, 1, 1, 1, 1), \quad e = \tfrac{1}{6}(1, 1, 1, 1, 1, 1)^t,$$
$$a = \text{diag}(1, 1, 1, -1, -1, -1).$$
One can verify that $A_0 = B_0 = 0$, $A_1 = -\alpha$,
$$\mu(1_H) = 1, \quad \mu(\tau_H) = \text{diag}\left(\alpha, \alpha + 1, \alpha + 2, \alpha, \alpha + 1, \alpha + 2\right),$$
and that $P_{\tau_H}(\mu(\tau_H)) = 0$. We have checked by computer calculation with Gröbner bases that no similar examples exist when $\text{rank}(\mathbf{V}) \in \{3, 4\}$.

One can also check in this example that $\mathbf{X}_{\tau,a}$ contains the center of $\mathbf{H}$. (Of course, once we have proved Theorem 5.1, we will know that $\mathbf{X}_{\tau,a}$ does not contain the full centralizer $\mathbf{H}_{\tau_H}$.)

The following lemma contains the final main idea for the proof of Theorem 5.6.

Lemma 5.21. *Let* $(R, \mathfrak{m})$ *be a local ring with* $2 \in R^\times$. *Let* $\tau \in \mathbf{M}_{\text{stab}}(R)$, $a \in \mathbf{G}_\tau(R)$. *Assume that* $\mathbf{X}_{\tau,a}$ *is doubly-tangential at* 1 *over* R. *Then one of the following is true:*

  (i) $a \in \mathbf{Z}(R)$.
 (ii) $a \notin \mathbf{Z}(R)$, $a^2 \in \mathbf{Z}(R)$ *and*
$$\tau \in Ra + R \subseteq \mathbf{M}(R).$$

Proof. The hypotheses of Lemma 5.18 apply, so the identity (5.14) holds. The argument divides according to whether A_0 lies in $\mathfrak{m}$ or not. If $A_0 \notin \mathfrak{m}$, so that A_0 is a unit, then from (5.14) and our assumption $2 \in R^\times$, we see that $a \in \mathbf{Z}(R)$. The remainder of the proof concerns the case $A_0 \in \mathfrak{m}$.

By Theorem 5.5 and our hypothesis that $\mathbf{X}_{\tau,a}$ is (doubly-)tangential at 1 over R, we have
$$a^2 \in \mathbf{Z}(R).$$
This hypothesis further implies that for
$$R' := R[\varepsilon]/(\varepsilon^2), \quad y := 1 + \varepsilon\tau_H \in \mathbf{H}_{\tau_H}(R'),$$
there exists $(h, b) \in \mathbf{H}(R') \times \mathbf{G}_\tau(R')$ such that $ay = hb$. It implies finally that $\mathbf{X}_{\tau,a}$ is tangential at y over R', hence by Theorem 5.5 that
$$b^2 \in \mathbf{Z}(R').$$
We retain the notation A_j, B_j from Proposition 5.12. From the calculation
$$e^* a \tau_H = e^* a(1 - ee^*)\tau(1 - ee^*) = e^* \left(a\tau - A_0\tau - A_1 + A_0(e^*\tau e)\right),$$
we see that
$$b = a + \varepsilon\left(a\tau - A_0\tau - A_1 + A_0(e^*\tau e)\right).$$
Squaring this relation gives
$$b^2 = a^2 + 2\varepsilon a\left(a\tau - A_0\tau - A_1 + A_0(e^*\tau e)\right), \tag{5.16}$$

using here that a and τ commute. Using now that both a^2 and b^2 lie in $\mathbf{Z}(R')$, together with our assumption $2 \in R^\times$, we deduce that

$$a(a - A_0)\tau \in Ra + R \subseteq \mathbf{M}(R).$$

Using next our assumption $A_0 \in \mathfrak{m}$, we see that $a(a - A_0)$ is invertible. Using again that $a^2 \in \mathbf{Z}(R)$, we conclude that $\tau \in Ra + R$. $\qquad\square$

Proof of Theorem 5.6. We have already noted, at the start of §5.1, that if $a \in \mathbf{H}(R)\mathbf{Z}(R)$, then $\mathbf{X}_{\tau,a} = \mathbf{H}_{\tau_H}$; in particular, $\mathbf{X}_{\tau,a}$ is doubly-tangential at y over R. The converse is the interesting direction.

Suppose, thus, that 2 is a unit in R, $\mathrm{rank}(\mathbf{V}) \geq 3$, $\tau \in \mathbf{M}_{\mathrm{stab}}$ and $a \notin \mathbf{H}(R)\mathbf{Z}(R)$, but $\mathbf{X}_{\tau,a}$ is doubly-tangential at some $y \in \mathbf{X}_{\tau,a}(R)$ over R. We aim to derive a contradiction.

By passing to the localization of R at a prime $\mathfrak{p}$ for which the image of a does not lie in $\mathbf{H}(R_\mathfrak{p})\mathbf{Z}(R_\mathfrak{p})$, we may assume that R is a local ring. We may reduce further to the case $a \in \mathbf{G}_\tau(R) - \mathbf{Z}(R)$ and $y = 1$, for the same reasons as noted at the beginning of §5.4.

Since $\mathbf{X}_{\tau,a}$ is doubly-tangential at 1 over R, we know from Theorem 5.5 and Lemma 5.21 that $a^2 \in \mathbf{Z}(R)$ and $\tau \in Ra + R$. In particular, we may find $c_0, c_1, c_2 \in R$ so that $\tau = c_0 + c_1 a$ and $a^2 = c_2$. Then $\tau^2 = c_0^2 + c_1^2 c_2 + 2c_0 c_1 a$, hence $\tau^2 - 2c_0\tau + c_0^2 - c_1^2 c_2 = 0$. Since τ is stable, the vector e is τ-cyclic, so

$$\mathbf{V}(R) = R[\tau]e = Re + R\tau e.$$

Since the rank of $\mathbf{V}$ is ≥ 3, we obtain a contradiction. $\qquad\square$

6. Volume bounds

We now apply the results of §5 to deduce a uniform solution to Problem 1.10.

6.1. **Bounds for polynomial congruences.** Let n and d be natural numbers. In §6.1, we use the equivalent notations $A \ll B$ and $A = \mathrm{O}(B)$ to denote that $|A| \leq C|B|$, where C depends at most upon n and d.

Lemma 6.1. Let F be a finite field, of cardinality q. Let $P \in F[x_1, \ldots, x_n]$ be a polynomial of degree at most d whose coefficients are not all zero. Then

$$|\{x \in F^n : P(x) = 0\}| \ll q^{n-1}.$$

Proof. By the fundamental theorem of algebra, we know that for all but $\mathrm{O}(1)$ many x_1, the polynomial that we get by specializing to that value is nonzero. By iterating this observation for x_2, x_3, and so on, we obtain the required estimate. $\qquad\square$

Lemma 6.2. Let m be a natural number. Suppose given a subset $\mathcal{D} \subseteq (\mathfrak{o}/\mathfrak{p})^n$ and a polynomial $P \in (\mathfrak{o}/\mathfrak{p}^m)[X_1, \ldots, X_n]$ of degree $\leq d$. Let $\mathcal{D}_m \subseteq (\mathfrak{o}/\mathfrak{p}^m)^n$ denote the inverse image of $\mathcal{D}$.

(i) *Assume that for each $y \in \mathcal{D}_m$, at least one of the linear Taylor coefficients for P at y is a unit. Then*

$$|\{y \in \mathcal{D}_m : P(y) \in \mathfrak{p}^m\}| \ll q^{mn-m}.$$

(ii) *Assume that for each $y \in \mathcal{D}_m$, at least one of the linear or quadratic Taylor coefficients for P at y is a unit. Then*

$$|\{y \in \mathcal{D}_m : P(y) \in \mathfrak{p}^m\}| \ll q^{mn-\lceil m/2 \rceil}.$$

Proof. In the case $m = 1$, either assertion reduces to the hypersurface bound (Lemma 6.1).

The case $m \geq 2$ of part (i) reduces via Hensel's lemma to the case $m = 1$. Indeed, under the stated assumptions, we have more precisely that each fiber of $\mathcal{D}_m \to \mathcal{D}$ contains at most $q^{(m-1)(n-1)}$ elements y for which $P(y) \in \mathfrak{p}^m$. To see this, suppose for instance that the coefficient of X_1 in the Taylor expansion of P at y is a unit. We may then use Hensel's lemma and the condition $P(y) \in \mathfrak{p}^m$ to determine $y_1 \in \mathfrak{o}/\mathfrak{p}^m$ in terms of the other coefficients. The number of possibilities for each of $y_2, \ldots, y_n$ is at most q^{m-1}, giving the required estimate.

We turn to part (ii). The case $m = 2$ follows formally from the case $m = 1$, using that solutions modulo $\mathfrak{p}^2$ map to solutions modulo $\mathfrak{p}$. We address the cases $m \geq 3$ by induction. We first observe that if $P(y) \in \mathfrak{p}^m$, then certainly $P(y) \in \mathfrak{p}$, and the latter condition depends only upon the class of y modulo $\mathfrak{p}$. By the case $m = 1$ that we have already addressed, the number of such classes is $\ll q^{n-1}$. By replacing P with a translate, we reduce to showing that when $P(0) \in \mathfrak{p}^m$, we have

$$|\{y \in (\mathfrak{p}/\mathfrak{p}^m)^n : P(y) \in \mathfrak{p}^m\}| \ll q^{mn - \lceil m/2 \rceil - (n-1)}.$$

We now consider two cases separately:

The first is when some linear Taylor coefficient of P is a unit, say the coefficient of X_1. Then, arguing as in part (i), we see that the cardinality in question is at most $q^{(m-1)(n-1)}$, which is better than the required estimate because $m - 1 \geq \lceil m/2 \rceil$ for $m \geq 3$.

The second is when all linear Taylor coefficients of P lie in $\mathfrak{p}$. We then form the polynomial

$$Q(y) := \varpi^{-2} P(\varpi y) \in (\mathfrak{o}/\mathfrak{p}^{m-2})[X_1, \ldots, X_n]$$

and we see that

$$|\{y \in (\mathfrak{p}/\mathfrak{p}^m)^n : P(y) \in \mathfrak{p}^m\}| = q^n |\{y \in (\mathfrak{o}/\mathfrak{p}^{m-2})^n : Q(y) \in \mathfrak{p}^{m-2}\}|.$$

By hypothesis, some quadratic Taylor coefficient of P is a unit, hence the same holds for Q. We may thus apply our inductive hypothesis to bound the cardinality on the right hand side, giving that the left hand side satisfies the estimate

$$\ll q^n q^{(m-2)n - \lceil (m-2)/2 \rceil} = q^{mn - \lceil m/2 \rceil - (n-1)},$$

as required. $\square$

6.2. Norms and distance functions.

For each $m \geq 0$, we equip the vector space F^m with the norm

$$|(v_1, \ldots, v_m)|_F := \max_{1 \leq j \leq m} |v_j|_F.$$

We apply this notation more generally to any vector space that comes with a natural basis, e.g., to the space of $m_1 \times m_2$ matrices over F.

We choose a basis $\{e_1, \ldots, e_n\}$ of $\mathbf{V}_H(\mathfrak{o})$, so that $\{e_1, \ldots, e_n, e\}$ is a basis of $\mathbf{V}(\mathfrak{o})$, and use these bases to identify $\mathbf{V}(F), \mathbf{V}^*(F)$ and $\mathbf{G}(F)$ with matrices over F. The above notation then applies.

For $h \in \mathbf{G}(F)$, we write

$$h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad h^{-1} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix},$$

where a and a' are $n \times n$ block matrices, and set

$$d_H(g) := \min\left(1, \max\left\{|b/d|_F, |b'/d'|_F, |c/d|_F, |c'/d'|_F\right\}\right).$$

This definition differs mildly from that in [19, §4.2], but agrees up to a constant factor, for reasons explained there.

Lemma 6.3. *Let $\ell \geq 0$ and $g \in K$ be such that*

$$g \in K_H K_Z K(\mathfrak{p}^\ell). \tag{6.1}$$

Then $d_H(g) \leq q^{-\ell}$, with equality if $g \notin K_H K_Z K(\mathfrak{p}^{\ell+1})$.

Proof. This follows readily from the following observation: for $\ell \geq 1$, the membership (6.1) says that $d, d' \in \mathfrak{o}^\times$ and that at least one b or c (and b' or c') lies in $\mathfrak{p}^\ell$. $\square$

In other words, for $g \in K = \mathbf{G}(\mathfrak{o})$, the quantity $d_H(g)$ is the infimum of $q^{-\ell}$ taken over all $\ell \geq 0$ for which the image of g in $\mathbf{G}(\mathfrak{o}/\mathfrak{p}^\ell)$ lies in $\mathbf{H}(\mathfrak{o}/\mathfrak{p}^\ell)\mathbf{Z}(\mathfrak{o}/\mathfrak{p}^\ell)$. For example, if the image of g does not lie in $\mathbf{H}(\mathfrak{o}/\mathfrak{p})$, then $d_H(g) = 1$, while if g lies in $\mathbf{Z}(\mathfrak{o})\mathbf{H}(\mathfrak{o})$, then $d_H(g) = 0$. We abbreviate

$$d_H(g)^\infty := \begin{cases} 1 & \text{if } d_H(a) = 1, \\ 0 & \text{if } d_H(a) < 1. \end{cases}$$

We note that for a nonzero ideal $\mathfrak{q} \subseteq \mathfrak{p}$ and $a \in \mathbf{G}(\mathfrak{o}/\mathfrak{q}) - \mathbf{H}(\mathfrak{o}/\mathfrak{q})\mathbf{Z}(\mathfrak{o}/\mathfrak{q})$, the quantity $d_H(a)$ is well-defined.

6.3. Miscellaneous lemmas.
Let $(F, \mathfrak{o}, \mathfrak{p}, q)$ be a non-archimedean local field, and let $\mathfrak{q} \subseteq \mathfrak{p}$ be a nonzero $\mathfrak{o}$-ideal.

Lemma 6.4. *Assume that q is odd. Let $a \in \mathbf{G}(\mathfrak{o}/\mathfrak{q})$ such that*

(i) *the image of a in $\mathbf{G}(\mathfrak{o}/\mathfrak{p})$ lies in $\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$, and*
(ii) *$a^2 \in \mathbf{Z}(\mathfrak{o}/\mathfrak{q})$.*

Then $a \in \mathbf{Z}(\mathfrak{o}/\mathfrak{q})$.

Proof. Write $\mathfrak{q} = \mathfrak{p}^m$. If the conclusion fails, then we may write $a = \lambda(1 + \varpi^\ell x)$, where $\lambda \in (\mathfrak{o}/\mathfrak{q})^\times$, $1 \leq \ell < m$ and $x \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ has the property that its image $\bar{x}$ in $\mathbf{M}(\mathfrak{o}/\mathfrak{p})$ does not lie in $\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$. Then

$$\mathbf{Z}(\mathfrak{o}/\mathfrak{q}) \ni a^2/\lambda^2 = 1 + 2\varpi^\ell x + \varpi^{2\ell} x^2.$$

Reducing this identity modulo $\mathfrak{p}^{\ell+1}$ and using that 2 is a unit, we deduce that $\bar{x}$ lies in $\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$, giving the required contradiction. $\square$

Recall from Definition 5.4 the meaning of "tangential".

Lemma 6.5. *Assume that q is odd. Let $a \in \mathbf{G}(\mathfrak{o}/\mathfrak{q}) - \mathbf{H}(\mathfrak{o}/\mathfrak{q})\mathbf{Z}(\mathfrak{o}/\mathfrak{q})$ with $d_H(a) < 1$. Then for each $y \in \mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})$, we have that $\mathbf{X}_{\tau,a}$ is not tangential at y over $\mathfrak{o}/\mathfrak{q}$.*

Proof. Suppose otherwise that $\mathbf{X}_{\tau,a}$ is tangential at y over $\mathfrak{o}/\mathfrak{q}$. Writing $ay = hb$ with $(h, b) \in \mathbf{H}(\mathfrak{o}/\mathfrak{q}) \times \mathbf{G}_\tau(\mathfrak{o}/\mathfrak{q})$, it follows then by Theorem 5.5 that $b^2 \in \mathbf{Z}(\mathfrak{o}/\mathfrak{q})$. On the other hand, since $d_H(a) < 1$, the image $\bar{a}$ of a modulo $\mathfrak{p}$ lies in $\mathbf{H}(\mathfrak{o}/\mathfrak{p})\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$. By the uniqueness of the decomposition $\bar{a}\bar{y} = \bar{h}\bar{b}$, it follows that $\bar{b}$ lies in $\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$. By Lemma 6.4, it follows that $b \in \mathbf{Z}(\mathfrak{o}/\mathfrak{q})$. But then $a \in \mathbf{H}(\mathfrak{o}/\mathfrak{q})\mathbf{Z}(\mathfrak{o}/\mathfrak{q})$, contrary to assumption. $\square$

Lemma 6.6. *Let $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ be cyclic. Write n for the rank of $\mathbf{V}$. Then the group $\mathbf{G}_\tau(\mathfrak{o}/\mathfrak{q})$ has cardinality at least $(1 - q^{-1})^n [\mathfrak{o} : \mathfrak{q}]^n$.*

Proof. Since $\mathbf{G}_\tau(\mathfrak{o}/\mathfrak{q}) = (\mathbf{M}_\tau(\mathfrak{o}/\mathfrak{q}) \twoheadrightarrow \mathbf{M}_\tau(\mathfrak{o}/\mathfrak{p}))^{-1}(\mathbf{G}_\tau(\mathfrak{o}/\mathfrak{p}))$, we may reduce to the case $\mathfrak{q} = \mathfrak{p}$. Since $k := \mathfrak{o}/\mathfrak{p}$ is a field, we may appeal to Lemma 3.7 and the structure theorem for modules over the PID to see that the ring $\mathbf{M}_\tau(k)$ is isomorphic to a product of rings $k[X]/(p_i)$, where the p_i are primary polynomials whose degrees sum to n. Let d_i denote the degree of the irreducible polyomial of which p_i is a power. Taking unit groups, we obtain $|\mathbf{G}_\tau(k)| = q^n \prod_i(1 - q^{-d_i})$, and the claim follows from the fact that $\sum_i d_i \leq n$. $\qquad\square$

6.4. **Main result.**

Theorem 6.7. *Let $(F, \mathfrak{o}, \mathfrak{p}, q)$ be a non-archimedean local field with q odd. Let $\mathfrak{q} \subseteq \mathfrak{p}$ be a nonzero $\mathfrak{o}$-ideal. Denote by $Q := [\mathfrak{o} : \mathfrak{q}]$ its absolute norm. Set*

$$Q^* := q^{\lceil m/2 \rceil} \text{ if } Q = q^m.$$

Let $\tau \in \mathbf{M}_{\mathrm{stab}}(\mathfrak{o}/\mathfrak{q})$ and $a \in \mathbf{G}(\mathfrak{o}/\mathfrak{q}) - \mathbf{H}(\mathfrak{o}/\mathfrak{q})\mathbf{Z}(\mathfrak{o}/\mathfrak{q})$. Then

$$|\mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})| \leq C \left(\frac{1}{1 + Qd_H(a)} + \frac{d_H(a)^\infty}{Q^*} \right) |\mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{q})|,$$

where $C \geq 0$ depends at most upon the rank of $\mathbf{V}$.

Proof. By Lemma 6.6, the cardinality of $\mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{q})$ is comparable to Q^n, where we write

$$\mathrm{rank}(\mathbf{V}) = n + 1.$$

For this reason, it suffices to establish the modified estimate obtained by replacing $|\mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{q})|$ with Q^n. Moreover, in view of the trivial bound $|\mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})| \leq |\mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{q})|$, it suffices to show that

$$|\mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})| \ll \left(\frac{1}{Qd_H(a)} + \frac{d_H(a)^\infty}{Q^*} \right) Q^n,$$

where $\ll$ means "bounded in magnitude by a scalar depending at most upon the rank of $\mathbf{V}$". We will establish this separately when $d_H(a) < 1$ and when $d_H(a) = 1$.

Recall from Remark 5.10 that $\mathbf{X}_{\tau,a}$ is defined by a system of $n+1$ polynomial equations in the entries of $y \in \mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{q})$, each of degree at most n.

Consider first the case $d_H(a) < 1$. Write $Q = q^m$ and $d_H(a) = q^{-\ell}$, so that $1 \leq \ell < m$. The image of a lies in $\mathbf{H}(\mathfrak{o}/\mathfrak{p}^\ell)\mathbf{Z}(\mathfrak{o}/\mathfrak{p}^\ell)$, thus the defining polynomial equations are trivial mod $\mathfrak{p}^\ell$. The image of a does not lie in $\mathbf{H}(\mathfrak{o}/\mathfrak{p}^{\ell+1})\mathbf{Z}(\mathfrak{o}/\mathfrak{p}^{\ell+1})$, so by Lemma 6.5, we see that $\mathbf{X}_{\tau,a}$ is not tangential over $\mathfrak{o}/\mathfrak{p}^{\ell+1}$ at any $y \in \mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{p}^{\ell+1})$. This implies that at least one of the polynomial equations in our system has the property that at least one of the linear Taylor coefficients at y lies in $\mathfrak{p}^\ell - \mathfrak{p}^{\ell+1}$. Let us divide that polynomial congruence by ϖ^ℓ and view it now as a polynomial congruence taken modulo $\mathfrak{p}^{m-\ell}$. Each solution of the new congruence corresponds to exactly $q^{\ell n}$ solutions of the old congruence, but at least one linear Taylor coefficient for the new congruence is a unit, so by part (i) of Lemma 6.2 (applied with $\mathcal{D} = \mathbf{H}_{\tau_H}(\mathfrak{o}/\mathfrak{p})$), we obtain

$$|\mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})| \ll q^{\ell n} q^{(m-\ell)n - (m-\ell)} = Q^{n-1}q^\ell = \frac{1}{Qd_H(a)}Q^n,$$

as required in this case.

It remains to consider the case $d_H(a) = 1$, where our goal bound is now

$$|\mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{q})| \ll \frac{1}{Q^*}Q^n.$$

We now have that the image of a does not lie in $\mathbf{H}(\mathfrak{o}/\mathfrak{p})\mathbf{Z}(\mathfrak{o}/\mathfrak{p})$. We may thus apply Theorem 5.6 to see that for each $y \in \mathbf{X}_{\tau,a}(\mathfrak{o}/\mathfrak{p})$, at least one of the linear or quadratic

Taylor coefficients for one of the defining polynomials for $\mathbf{X}_{\tau,a}$ is a unit. By part (ii) of Lemma 6.2, we deduce the required bound. □

Remark 6.8. For applications to subconvexity involving a *fixed* local field of characteristic zero, the "near-identity reduction" (see §1.9.3 and [19, §15.6]) reduces our task, as far as volume bounds like in Theorem 6.7 are concerned, to the case that a is close to the identity element. In that case, an adequate volume bound follows from [19, Theorem 15.2]. The novelty of Theorem 6.7 relative to the cited result is its validity for all a in the maximal compact subgroup, uniformly in the local field. The restriction to characteristic $\neq 2$ in Theorem 6.7 is thus harmless in applications. With additional work, this restriction could be eliminated in the case of a non-archimedean local field of characteristic zero and residue characteristic 2, basically because the proof of Theorem 5.1 requires only a bounded number of divisions by 2.

There remains the case of a local field of characteristic 2, where the conclusion of Theorem 6.7 fails for the reasons indicated in Remark 5.3. That case is not relevant for the subconvexity problem over number fields, but would be relevant for studying moments of L-functions in horizontal aspects over a function field of characteristic 2. Estimating such moments thus presents an interesting challenge to which our results do not apply.

7. BILINEAR FORMS ESTIMATES

We now apply the volume bound to estimate bilinear forms relevant for (1.11).

Let $(F, \mathfrak{o}, \mathfrak{p}, \varpi, q)$ be a non-archimedean local field of odd residue characteristic. Let $\mathfrak{q} \subseteq \mathfrak{p}$ be a nonzero $\mathfrak{o}$-ideal. In this section, we abbreviate

$$G := \mathbf{G}(\mathfrak{o}/\mathfrak{q}), \qquad H := \mathbf{H}(\mathfrak{o}/\mathfrak{q}),$$

etc. Write $Q := [\mathfrak{o} : \mathfrak{q}]$, and define Q^* as in the statement of Theorem 6.7. Recall from §6.2 the "distance from HZ" function d_H.

Theorem 7.1. *Let $\tau \in M_{\mathrm{stab}}$. Let $u_1, u_2 : H \to \mathbb{R}_{\geq 0}$ be nonnegative functions that are right-invariant under H_{τ_H}. Let $\gamma \in G$. Then the quantity*

$$I := \frac{1}{|H|^2} \sum_{\substack{x,y \in H: \\ x^{-1}\gamma y \in G_\tau}} u_1(x)u_2(y) \tag{7.1}$$

satisfies the trivial bound

$$I \leq \frac{1}{|H|} \|u_1\|_{L^2} \|u_2\|_{L^2} \tag{7.2}$$

and, for $\gamma \notin HZ$, the refined bound

$$I \leq \frac{C}{|H|} \left(\frac{1}{1 + Q d_H(\gamma)} + \frac{d_H(\gamma)^\infty}{Q^*} \right) \|u_1\|_{L^2} \|u_2\|_{L^2}. \tag{7.3}$$

Here C depends at most upon $\mathrm{rank}(\mathbf{V})$, and may be taken to be the same constant as in Theorem 6.7. The L^2-norms are defined using the invariant probability measures.

Proof. We closely follow the proof of [19, Thm 15.1, Lem 15.3]. By Cauchy–Schwarz, we have $I \leq \sqrt{I_1 I_2}$, where

$$I_1 := \frac{1}{|H|^2} \sum_{\substack{x,y \in H: \\ x^{-1}\gamma y \in G_\tau}} u_1(x)^2,$$

$$I_2 := \frac{1}{|H|^2} \sum_{\substack{x,y \in H: \\ x^{-1}\gamma y \in G_\tau}} u_2(y)^2.$$

The set G_τ, the condition $\gamma \notin HZ$ and the quantity $d_H(\gamma)$ are invariant under inversion, so it will suffice to obtain a suitable estimate for I_1, as the same argument then applies to I_2. By definition,

$$I_1 = \frac{1}{|H|^2} \sum_{x \in H} u_1(x)^2 \nu(x), \quad \nu(x) := \left| \left\{ y \in H : x^{-1}\gamma y \in G_\tau \right\} \right|.$$

By Lemma 4.6, we have $H \cap G_\tau = \{1\}$, and so $\nu(x) \in \{0, 1\}$. Thus

$$I_1 \le \frac{1}{|H|} I_1', \quad I_1' := \frac{1}{|H|} \sum_{\substack{x \in H: \\ \gamma^{-1}x \in HG_\tau}} u_1(x)^2.$$

The trivial bound $I_1' \le \|u_1\|_{L^2}^2$, obtained by dropping the summation condition, yields the first required estimate (7.2).

For the second required estimate, we appeal to the right H_{τ_H}-invariance of u_1 to write

$$I_1' := \frac{1}{|H|} \sum_{x \in H} u_1(x)^2 \mu(x), \quad \mu(x) := \frac{\left| \left\{ u \in H_{\tau_H} : \gamma^{-1}xu \in HG_\tau \right\} \right|}{|H_{\tau_H}|}.$$

If $\gamma \notin HZ$, then we have $\gamma^{-1}x \notin HZ$ and $d_H(\gamma^{-1}x) = d_H(\gamma)$ for all $x \in H$. Estimating $\mu(x)$ via Theorem 6.7 yields the second estimate (7.3). $\qquad\square$

8. Construction of test vectors

The primary aim of this section is to make Theorem 1.1 precise by specifying the local conditions required at the distinguished place (Remark 1.6). To that end, we define "stable" pairs of representations of general linear groups (Definition 8.6); informally, such representations contain "microlocalized" test vectors whose localization parameters have no matching eigenvalues. To apply this definition, we need to check when it holds. Since representation-theoretic issues are orthogonal to the main novelty of this paper, we are content here to check that our definition is preserved under parabolic induction (Example 8.7) and to investigate fully the case of principal series (Example 8.8), leaving open the case of general supercuspidals (see Example 8.9 for some special cases).

We note that many of the ideas and results recorded in this section are well-known in the type theory literature, going back to work of Howe from the 1970s (see, e.g., [10, 9] and [1, §5]). The cited works would be most relevant to the special cases of our results where τ and τ_H are diagonalizable, a condition that corresponds to the "wall-avoidance" hypothesis (i) mentioned in §1.5. Since we do not wish to impose such hypotheses, we give short proofs of what we need here, without claiming particular novelty.

8.1. **Overview.** In §8.1, we summarize the main definitions and results of §8; the remaining subsections are then devoted to the proofs of these results.

Let F be a non-archimedean local field, with ring of integers $\mathfrak{o}$ and maximal ideal $\mathfrak{p}$. Let $\mathfrak{q} \subseteq \mathfrak{p}$ be an $\mathfrak{o}$-ideal. Let ψ be a unitary character of $\mathfrak{q}$ that is trivial on $\mathfrak{q}^2$, but not on $\mathfrak{p}^{-1}\mathfrak{q}^2$.

For $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$, we denote by χ_τ the character of the group $K(\mathfrak{q})/K(\mathfrak{q}^2)$ given by

$$\chi_\tau(1 + x) = \psi(\mathrm{trace}(x\tau)). \tag{8.1}$$

Every character of that group arises in this way.

In what follows, *representation* always means "complex representation". We introduce some terminology concerning representations π of $\mathbf{G}(\mathfrak{o})$. We will apply this terminology also to representations of $\mathbf{G}(F)$, through their restrictions.

Definition 8.1. Let π be a representation of $\mathbf{G}(\mathfrak{o})$. We say that π is *regular at depth* $\mathfrak{q}^2$ if there is a cyclic element $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ (see §3) such that π contains a nonzero vector v that transforms under $K(\mathfrak{q})$ via the character χ_τ, i.e.,

$$gv = \chi_\tau(g)v \quad \text{for all } g \in K(\mathfrak{q}). \tag{8.2}$$

In that case, we refer to τ as a *regular parameter for π at depth* $\mathfrak{q}^2$, to its characteristic polynomial $P \in (\mathfrak{o}/\mathfrak{q})[X]$ as a *polynomial for π at depth* $\mathfrak{q}^2$, and to the dimension of the space of v satisfying (8.2) as the *multiplicity* of τ or of P in π. (These notions depend, of course, upon the choice of ψ.)

Example 8.2. Let χ be a character of $\mathrm{GL}_1(\mathfrak{o}) = \mathfrak{o}^\times$, regarded also as a one-dimensional representation. We denote by

$$c(\chi) \subseteq \mathfrak{o}$$

the largest ideal $\mathfrak{a}$ of $\mathfrak{o}$ for which χ has trivial restriction to $\mathfrak{o}^\times \cap (1 + \mathfrak{a})$. It is clear then that

$$c(\chi) \supseteq \mathfrak{q}^2 \iff \chi \text{ is regular at depth } \mathfrak{q}^2.$$

Example 8.3. We verify in Proposition 8.19 that being regular at depth $\mathfrak{q}^2$ is closed under parabolic induction. That is to say, suppose given a partition $n = m_1 + \cdots + m_k$, corresponding to a standard parabolic subgroup of $\mathrm{GL}_n(\mathfrak{o})$. For each $j \in \{1, \ldots, k\}$, let π_j be a representation of $\mathrm{GL}_{m_j}(\mathfrak{o})$ that is regular at depth $\mathfrak{q}^2$ with polynomial P_j. Then the same holds for the parabolic induction of $\pi_1 \otimes \cdots \otimes \pi_k$ to $\mathrm{GL}_n(\mathfrak{o})$, with polynomial at depth $\mathfrak{q}^2$ given by $P = \prod_i P_i$ (unique if the P_i are).

Example 8.4. By combining Examples 8.2 and 8.3, we deduce that a principal series representation induced by characters χ_i with $c(\chi_i) \supseteq \mathfrak{q}^2$ is regular at depth $\mathfrak{q}^2$. Its unique polynomial at depth $\mathfrak{q}^2$ is the product of linear factors $\prod_i (X - \xi_i)$, where ξ_i is characterized by the identity $\chi_i(1 + y) = \psi(y\xi_i)$ for all $y \in \mathfrak{q}$.

Example 8.5. We verify in §8.8 that certain supercuspidal representations are regular at depth $\mathfrak{q}^2$. We do not determine precisely which supercuspidal representations are regular at depth $\mathfrak{q}^2$, but it may be possible to do so using known classifications of the latter (see Remark 8.27).

Definition 8.6. Let m and n be natural numbers. Let π and σ be representations of $\mathrm{GL}_n(\mathfrak{o})$ and $\mathrm{GL}_m(\mathfrak{o})$. We say that (π, σ) is *stable at depth* $\mathfrak{q}^2$ if
- both π and σ are regular at depth $\mathfrak{q}^2$, and
- there are polynomials P_π (resp. P_σ) for π (resp. σ) at depth $\mathfrak{q}^2$ that generate the unit ideal of $(\mathfrak{o}/\mathfrak{q})[X]$.

Example 8.7. Being stable at depth $\mathfrak{q}^2$ is closed under parabolic induction (Proposition 8.21): if π (resp. σ) is induced by some collection of representations π_i (resp. σ_j) of $\mathrm{GL}_{n_i}(\mathfrak{o})$ (resp. $\mathrm{GL}_{m_j}(\mathfrak{o})$) such that each pair (π_i, σ_j) is stable at depth $\mathfrak{q}^2$, then the same holds for (π, σ).

Example 8.8. If π (resp. σ) is a principal series representation induced by characters χ_i (resp. η_j), then the pair (π, σ) is stable at depth $\mathfrak{q}^2$ provided that for all relevant indices,

$$c(\chi_i) \supseteq \mathfrak{q}^2, \qquad c(\eta_j) \supseteq \mathfrak{q}^2, \qquad c(\chi_i/\eta_j) = \mathfrak{q}^2. \tag{8.3}$$

Assuming the first two conditions in (8.3), the third condition is equivalent to

$$\prod_{i,j}[\mathfrak{o} : c(\chi_i/\eta_j)] = [\mathfrak{o} : \mathfrak{q}^2]^{n(n+1)}. \tag{8.4}$$

The left hand side may be understood as the conductor of the Rankin–Selberg convolution $\pi \times \sigma^\vee$, and the equality (8.4) may be understood as an analogue of the "no conductor dropping" or "uniform growth" conditions considered in [19] and [18].

Example 8.9. When (π, σ) is a pair of regular supercuspidal representations of $(\mathrm{GL}_m, \mathrm{GL}_n)$ as discussed in Example 8.5 or §8.8, it is automatically stable when $m \neq n$; when $m = n$, one needs to impose additional conditions similar to the case of principal series representations (see Lemma 8.28).

8.2. **Characters.** Let $\mathfrak{q} \subseteq \mathfrak{p}$ be an $\mathfrak{o}$-ideal. Let $\psi : \mathfrak{q}/\mathfrak{q}^2 \to \mathbb{C}^\times$ be a character that is nontrivial on $\mathfrak{p}^{-1}\mathfrak{q}^2/\mathfrak{q}^2$. The character ψ induces an identification

$$\mathfrak{o}/\mathfrak{q} \xrightarrow{\cong} \left\{ \text{characters } \mathfrak{q}/\mathfrak{q}^2 \to \mathbb{C}^\times \right\},$$

$$\xi \mapsto \psi(\xi\bullet).$$

For a character χ of $1 + \mathfrak{q}$ that is trivial on $1 + \mathfrak{q}^2$, we denote by $\xi_\chi \in \mathfrak{o}/\mathfrak{q}$ the unique element with

$$\chi(1 + x) = \psi(\xi_\chi x) \quad \text{for } x \in \mathfrak{q}.$$

8.3. **Centralizers of characters.** Given $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$, we denote by

$$J_\tau \leq \mathbf{G}(\mathfrak{o})$$

the inverse image of the mod-$\mathfrak{q}$ centralizer $\mathbf{G}_\tau(\mathfrak{o}/\mathfrak{q})$. It is a subgroup that contains $K(\mathfrak{q})$.

8.4. **Preliminary lemmas regarding regular parameters.**

Lemma 8.10. Let P be a polynomial for π at depth $\mathfrak{q}^2$. Then any cyclic element $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ whose characteristic polynomial is P is a regular parameter for π at depth $\mathfrak{q}^2$.

Proof. We observe first that the set of regular parameters for π at depth $\mathfrak{q}^2$ is invariant under conjugation by $\mathbf{G}(\mathfrak{o}/\mathfrak{q})$. The conclusion of the lemma then follows from the fact (Lemma 3.8) that any two cyclic elements having the same characteristic polynomial are conjugate. $\square$

Lemma 8.11. Let $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ be cyclic. Then $\ker(\chi_\tau)$ is a normal subgroup of J_τ, and the quotient group $J_\tau/\ker(\chi_\tau)$ is abelian.

Proof. We verify first that the kernel is normal. Let $x \in J_\tau$ and $y \in \ker(\chi_\tau)$. Then xyx^{-1} lies in $K(\mathfrak{q})$ (because y lies in $K(\mathfrak{q})$, which is normal in $\mathbf{G}(\mathfrak{o})$) and also in $\ker(\chi_\tau)$ (because x commutes with τ modulo $\mathfrak{q}$).

We verify next that the quotient is abelian. Choose an arbitrary lift $\tilde{\tau} \in \mathbf{G}(\mathfrak{o})$ of τ. Then, since $\mathbf{M}_\tau(\mathfrak{o}/\mathfrak{q})$ consists of polynomials in τ (Lemma 3.7), we see that any element of J_τ may be written as a product ab, where a is a polynomial in $\tilde{\tau}$ and b lies in $K(\mathfrak{q})$. The images of such elements generate the quotient group $J_\tau/\ker(\chi_\tau)$, so to verify that the latter is abelian, we reduce to checking that any commutator (a_1, a_2) or (a_1, b_1) or (b_1, b_2) lies in $\ker(\chi_\tau)$, where the a_i and b_j are as before. Indeed, we have $(a_1, a_2) = 1$ (because any two polynomials in $\tilde{\tau}$ commute with one another) and $\chi_\tau((a_1, b_1)) = 1$ (because a_1 commutes with τ modulo $\mathfrak{q}$) and $\chi_\tau((b_1, b_2)) = 1$ (because χ_τ is a character of $K(\mathfrak{q})$). $\square$

Lemma 8.12. Let $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ be a regular parameter for π at depth $\mathfrak{q}^2$. Then there is an extension of χ_τ to a character $\tilde{\chi}_\tau$ of J_τ and a nonzero vector $v \in \pi$ that transforms under J_τ via $\tilde{\chi}_\tau$. If τ occurs with multiplicity one, then v is unique up to scalar multiple.

Proof. Let v_0 be a nonzero vector in π that transforms under $K(\mathfrak{q})$ via χ_τ. Let V denote the span of v_0 under J_τ. This is a vector space of dimension at most $[J_\tau : K(\mathfrak{q})]$, on which $K(\mathfrak{q}^2)$ acts trivially. By Lemma 8.11, the action of J_τ on V factors through that of the quotient group $J_\tau/\ker(\chi_\tau)$. The latter is a finite abelian group, so its action on V decomposes as a direct sum of one-dimensional subspaces. Take v to lie in one of these subspaces, and $\tilde{\chi}_\tau$ to be the character describing the eigenvalues for J_τ acting on v. The final assertion is clear. $\square$

Example 8.13. Let $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ be cyclic. It defines a character χ_τ of the subgroup $K(\mathfrak{q})$ of J_τ. Suppose given an extension $\tilde{\chi}_\tau$ of χ_τ to a character of J_τ. Let π denote the representation of $\mathbf{G}(\mathfrak{o})$ induced from $\tilde{\chi}_\tau$. Then π is regular at depth $\mathfrak{q}^2$, with regular parameter τ. These elementary observations imply that certain supercuspidal representations satisfy the above definitions (specifically, those of "odd depth" and "generic inducing data" — see §8.8).

8.5. **Some linear algebra.** We record some general lemmas to be applied in the following subsection. Here we focus on an individual ring R and abbreviate $V := \mathbf{V}(R)$, $M := \mathbf{M}(R)$, etc.

Lemma 8.14. Let $\tau \in M$. Then V is τ-cyclic if and only if V^* is τ-cyclic. More precisely, if the vector $v \in V$ is τ-cyclic, then functional $v^* \in V^*$ given by $\sum c_j \tau^j v \mapsto c_{n-1}$ is τ-cyclic.

Proof. This is the observation that the transpose of a matrix like (3.3) is cyclic; we leave it to the reader. $\square$

In what follows, by a *flag* $V_0 \subset \cdots \subset V_k$ in V, we mean a sequence of free submodules, with $V_0 = \{0\}$ and $V_k = V$, such that each quotient V_j/V_{j-1} is also free. We say that $\tau \in M$ preserves this flag if $\tau V_j \subseteq V_j$ for each $j \in \{1, \ldots, k\}$.

Lemma 8.15. Let $\tau \in M$ be cyclic. Let $V_0 \subset \cdots \subset V_k$ be a flag. Suppose that τ preserves the flag, so that τ induces endomorphisms $\tau_j \in \mathrm{End}(V_j/V_{j-1})$. Then each τ_j is cyclic.

Proof. By induction, we may reduce to the case $k = 2$. It is clear that τ_2 is cyclic, because the image in V/V_1 of any cyclic vector for τ is a cyclic vector for τ_2. It remains to show that τ_1 is cyclic. By Lemma 8.14, we know that $\tau^* \in \mathrm{End}(V^*)$ is cyclic, and it suffices to show that $\tau_1^* \in \mathrm{End}(V_1^*)$ is cyclic. By identifying V_1^* with the quotient $V^*/V_1^\perp$, we reduce to the previous observation. $\square$

Lemma 8.16. If τ is cyclic, then its characteristic polynomial P_τ generates the annihilator ideal $\{f \in R[X] : f(\tau) = 0\}$.

Proof. Otherwise, let f be polynomial of minimal degree that is not divisible by P_τ and yet for which $f(\tau) = 0$. If the degree of f is less than the rank of $\mathbf{V}$, then we obtain a contradiction from Lemma 3.6. Otherwise, we can apply division with remainder using the monic polynomial P_τ to contradict the minimality of f. $\square$

Lemma 8.17. Let $V_1 \subset \cdots \subset V_k$ be a flag, and let $\tau_j \in \mathrm{End}(V_j/V_{j-1})$ be cyclic elements. Then there is a cyclic element $\tau \in \mathrm{End}(V)$ that preserves the flag and for which the induced action on V_j/V_{j-1} is τ_j.

Proof. By induction, we may reduce to the case $k = 2$ and $V_k = V$.

We may find a basis $e_1, \ldots, e_m$ for V_1 with respect to which τ_1 is cyclic and a basis $\bar{f}_1, \ldots, \bar{f}_n$ for V/V_1 with respect to which τ_2 is cyclic, i.e.,

$$\tau_1 e_i = e_{i+1} \quad (i < m), \qquad \tau_2 \bar{f}_j = \bar{f}_{j+1} \quad (j < n).$$

By choosing a splitting $V = V_1 \oplus V_2$, we identify τ_2 with an endomorphism of V_2, extended by zero to V. We then define $\tau \in \mathrm{End}(V)$ by taking

$$\tau e_i = \tau_1 e_i, \qquad \tau f_j = \tau_2 f_j + \begin{cases} e_1 & \text{if } j = n, \\ 0 & \text{otherwise.} \end{cases}$$

We claim then that f_1 is a cyclic vector for τ. Indeed, the set $\{f_1, \tau f_1, \ldots, \tau^{n-1} f_1\} = \{f_1, f_2, \ldots, f_n\}$ spans V_2, while the set $\{\tau^n f_1, \ldots, \tau^{n+m-1} f_1\}$ is congruent modulo V_2 to the spanning set $\{e_1, \ldots, e_m\}$ for V_1. $\qquad \square$

Lemma 8.18. *Let $V_1 \subset \cdots \subset V_k$ be a flag. Let $\tau \in M$ be a cyclic element that preserves this flag. Suppose that $g \in G$ has the following properties:*

(i) *$\mathrm{Ad}(g)\tau$ preserves the flag.*
(ii) *For each j, the induced actions of τ and $\mathrm{Ad}(g)\tau$ on V_j/V_{j-1} have the same characteristic polynomial.*

Then g preserves the flag.

Proof. By induction, we may reduce to the case $k = 2$. Let $v \in V$ be τ-cyclic. Then every element of V may be written $f(\tau)v$ for some polynomial $f \in R[X]$, so it suffices to show that if $f(\tau)v \in V_1$, then $gf(\tau)v \in V_1$. Suppose, thus, that $f(\tau)v \in V_1$. Let $v_2 \in V/V_1$ denote the image of v, and write $\tau_2 \in \mathrm{End}(V/V_1)$ for the induced action of τ on the quotient. Then v_2 is τ_2-cyclic and $f(\tau_2)v_2 = 0$, hence $f(\tau_2) = 0$. It follows that f is divisible by the characteristic polynomial of τ_2 (Lemma 8.16). Since the induced action of $\mathrm{Ad}(g)\tau$ on V/V_1 has the same characteristic polynomial as τ_2, it follows from the Cayley–Hamilton theorem that $f(\mathrm{Ad}(g)\tau)V \subseteq V_1$. But then $gf(\tau)v = f(\mathrm{Ad}(g)\tau)gv \in V_1$, as required. $\qquad \square$

8.6. Parabolic induction.

The following proposition shows that the existence and uniqueness of localized vectors is preserved by induction. (The "uniqueness" assertions will not be used here, but might be useful for certain extensions of the work of this paper.)

Proposition 8.19. *The class of representations of general linear groups over $\mathfrak{o}$ that are regular at depth $\mathfrak{q}^2$ is closed under parabolic induction. That is to say, writing $n = \mathrm{rank}(\mathbf{V})$, suppose given a partition $n = m_1 + \cdots + m_k$, corresponding to a parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$, together with a collection of representations π_j ($j = 1, \ldots, k$) of $\mathrm{GL}_{m_j}(\mathfrak{o})$ each of which is regular at depth $\mathfrak{q}^2$. Then the same holds for the parabolic induction*

$$\pi := \mathrm{Ind}_{\mathbf{P}(\mathfrak{o})}^{\mathbf{G}(\mathfrak{o})}(\pi_1 \otimes \cdots \otimes \pi_k).$$

Moreover:

(i) *Every polynomial for π at depth $\mathfrak{q}^2$ is a product of polynomials for $\pi_1, \ldots, \pi_k$ at depth $\mathfrak{q}^2$.*
(ii) *For any polynomials $P_1, \ldots, P_k$ for $\pi_1, \ldots, \pi_k$ at depth $\mathfrak{q}^2$, the product $P_1 \cdots P_k$ is a polynomial for π at depth $\mathfrak{q}^2$.*
(iii) *If π_i admits a unique polynomial P_i at depth $\mathfrak{q}^2$, and if P_i occurs with multiplicity one, then $P_1 \cdots P_k$ is likewise the unique polynomial for π at depth $\mathfrak{q}^2$, and occurs with multiplicity one.*

Proof. For $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$, let us temporarily write π^τ for the subspace of π on which $K(\mathfrak{q})$ acts via χ_τ. Similarly, for an $m_i \times m_i$ matrix τ_i over $\mathfrak{o}/\mathfrak{q}$, write $\pi_i^{\tau_i}$ for the associated subspace of π_i. Let us take for $\mathbf{P}$ the standard upper-triangular parabolic subgroup associated to the given partition. Let $\mathbf{U} < \mathbf{P}$ denote the unipotent radical and $\mathbf{L}$ the block-diagonal Levi factor, thus $\mathbf{L} \cong \prod_{i=1}^k \mathbf{G}_i$, where each $\mathbf{G}_i$ is a general linear group of rank m_i. We may regard

$$\Pi := \otimes_i \pi_i$$

as a representation of $\mathbf{L}(\mathfrak{o})$. We extend it to a representation of $\mathbf{P}(\mathfrak{o})$ by letting $\mathbf{U}(\mathfrak{o})$ act trivially, and realize π as the space of functions $v : \mathbf{G}(\mathfrak{o}) \to \Pi$ such that for all $p \in \mathbf{P}(\mathfrak{o})$,

$$v(pg) = \Pi(p)v(g),$$

with the action given by right translation.

By Mackey theory, a basis for π^τ is indexed by representatives g for the double quotient

$$\mathbf{P}(\mathfrak{o}) \backslash \mathbf{G}(\mathfrak{o}) / K(\mathfrak{q}) \cong \mathbf{P}(\mathfrak{o}/\mathfrak{q}) \backslash \mathbf{G}(\mathfrak{o}/\mathfrak{q}) \tag{8.5}$$

together with a basis for the space of vectors $t \in \Pi$ such that the formula

$$v_g(pgh) = \begin{cases} \chi_\tau(h)\Pi(p)t & \text{if } (p,h) \in \mathbf{P}(\mathfrak{o}) \times K(\mathfrak{q}), \\ 0 & \text{otherwise} \end{cases}$$

is well-defined, that is, for which $pgh = g \implies \chi_\tau(h)\Pi(p)t = t$, or equivalently,

$$p \in \mathbf{P}(\mathfrak{o}) \cap gK(\mathfrak{q})g^{-1} \implies \chi_{\mathrm{Ad}(g)\tau}(p)t = \chi_\tau(g^{-1}pg)t = \Pi(p)t. \tag{8.6}$$

Since $K(\mathfrak{q})$ is normal in $\mathbf{G}(\mathfrak{o})$, we have

$$\mathbf{P}(\mathfrak{o}) \cap gK(\mathfrak{q})g^{-1} = K_P(\mathfrak{q}) := \ker(\mathbf{P}(\mathfrak{o}) \to \mathbf{P}(\mathfrak{o}/\mathfrak{q})).$$

Any element p of the latter may be written uniquely as $p = mu$, where (m, u) lies in $K_L(\mathfrak{q}) \times K_U(\mathfrak{q})$ (with each factor defined like $K_P(\mathfrak{q})$). Assuming that the vector t is nonzero, the condition (8.6) is then equivalent to the following pair of assertions.

- $\chi_{\mathrm{Ad}(g)\tau}(u) = 1$ for all $u \in K_U(\mathfrak{q})$. Translating this assertion via the trace pairing, it says that the element $\mathrm{Ad}(g)\tau$ of $\mathbf{M}(\mathfrak{o}/\mathfrak{q})$ preserves the flag $\mathcal{F}$ in $\mathbf{M}(\mathfrak{o}/\mathfrak{q})$ of which $\mathbf{P}(\mathfrak{o}/\mathfrak{q})$ is the stabilizer.
- $\Pi(m)t = \chi_{\mathrm{Ad}(g)\tau}(m)t$ for all $m \in K_L(\mathfrak{q})$. Writing $(\mathrm{Ad}(g)\tau)_{ii}$ for the ith block diagonal $m_i \times m_i$ matrix, this is equivalent to the condition $t \in \otimes_{i=1}^k \pi_i^{(\mathrm{Ad}(g)\tau)_{ii}}$.

In summary,

$$\pi^\tau = \oplus_{\substack{g \in \mathbf{P}(\mathfrak{o}) \backslash \mathbf{G}(\mathfrak{o}) / K(\mathfrak{q}): \\ \mathrm{Ad}(g)\tau \text{ preserves } \mathcal{F}}} \otimes_{i=1}^k \pi_i^{(\mathrm{Ad}(g)\tau)_{ii}}. \tag{8.7}$$

The matrices $(\mathrm{Ad}(g)\tau)_{ii}$ are cyclic, as their images in $\mathbf{M}_{m_i}(\mathfrak{o}/\mathfrak{q})$ are cyclic by Lemma 8.15. If the spaces $\pi_i^{(\mathrm{Ad}(g)\tau)_{ii}}$ are nonzero, then the characteristic polynomial of $(\mathrm{Ad}(g)\tau)_{ii}$ must be a polynomial for π_j at depth $\mathfrak{q}^2$. Thus every polynomial for π is a product of polynomials for the π_j inside the ring $(\mathfrak{o}/\mathfrak{q})[X]$.

Conversely, suppose given polynomials $P_1, \ldots, P_k$ for $\pi_1, \ldots, \pi_k$ at depth $\mathfrak{q}^2$. By Lemma 8.17, there is a cyclic element τ which preserves the flag and for which each τ_{ii} has characteristic polynomial P_i. Then the characteristic polynomial of τ is $P_1 \cdots P_k$. On the other hand, the contribution to (8.7) from $g = 1$ is positive-dimensional, so π^τ is nonzero. Thus $P_1 \cdots P_k$ is a polynomial for π.

It remains to verify the final assertion concerning uniqueness and multiplicity one. Suppose that π^τ is nonzero for some cyclic $\tau \in M$. By (8.7), we may assume (after conjugating τ if necessary) that τ preserves $\mathcal{F}$ and that each $\pi_i^{\tau_{ii}}$ is nonzero. By hypothesis, the characteristic polynomial of τ_{ii} is P_i and $\pi_i^{\tau_{ii}}$ is one-dimensional. Moreover, by Lemma 8.18 (applied over $R = \mathfrak{o}/\mathfrak{q}$), the only $g \in \mathbf{G}(\mathfrak{o})$ for which $\mathrm{Ad}(g)\tau$ preserves $\mathcal{F}$ and $(\mathrm{Ad}(g)\tau)_{ii}$ has characteristic polynomial P_i lie in the trivial double coset. Thus (8.7) reduces to the one-dimensional summand $\otimes_{i=1}^k \pi_i^{\tau_{ii}}$. This proves that $P_1 \cdots P_k$ is the unique polynomial for π and occurs with multiplicity one. $\qquad\square$

8.7. Stable pairs of representations.
Recall that in Definition 8.6, we defined when a pair of representations (π, σ) is stable at depth $\mathfrak{q}^2$.

Example 8.20. Let χ and η be characters of $\mathfrak{o}^\times = \mathrm{GL}_1(\mathfrak{o})$. Then the pair (χ, η) is stable at depth $\mathfrak{q}^2$ precisely when χ and η have conductor dividing $\mathfrak{q}^2$ and the ratio χ/η has conductor exactly $\mathfrak{q}^2$. In verifying this, we observe that $\chi(x)/\eta(x) = \psi(x(\xi_\chi - \xi_\eta))$, which is nontrivial for some $x \in \mathfrak{p}^{-1}\mathfrak{q}^2$ precisely when $\xi_\chi - \xi_\eta$ is a unit, or equivalently, when the polynomials $X - \xi_\chi$ and $X - \xi_\eta$ generate the unit ideal in $(\mathfrak{o}/\mathfrak{q})[X]$.

Proposition 8.21. *The class of pairs of representations (π, σ) that are stable at depth $\mathfrak{q}^2$ is closed under parabolic induction, in the following sense. Suppose given a finite collection of representations π_i and σ_j, where each pair (π_i, σ_j) is stable at depth $\mathfrak{q}^2$. Let π (resp. σ) denote the parabolic induction of the π_i (resp. σ_j), as in Proposition 8.19. Then the pair (π, σ) is likewise stable at depth $\mathfrak{q}^2$.*

Proof. This reduces to Proposition 8.19 and the following observation: given finite collections of polynomials $\{P_i\}$ and $\{Q_j\}$ in $(\mathfrak{o}/\mathfrak{q})[X]$ such that each pair (P_i, Q_j) generates the unit ideal, the same holds for the pair (P, Q) given by $P := \prod_i P_i$ and $Q := \prod_j Q_j$. (To see this, write $1 = a_i P_i + b_j Q_j$ and observe that $\prod_{i,j}(a_i P_i + b_j Q_j)$ lies in the ideal generated by P and Q.) $\qquad\square$

Example 8.22. Let π (resp. σ) be a principal series representation induced by characters χ_i (resp. η_j), each of conductor dividing $\mathfrak{q}^2$. Suppose that each ratio χ_i/η_j has conductor $\mathfrak{q}^2$. Then (π, σ) is stable at depth $\mathfrak{q}^2$. In verifying this, we may reduce via Proposition 8.21 to the case of a pair of characters (χ, η) discussed in Example 8.20.

8.8. Supercuspidals.
We verify that the above terminology applies to certain supercuspidal representations of $\mathbf{G}(F)$. We do not aim here for an exhaustive treatment, as representation-theoretic issues are not the focus of this paper.

Write $n := \dim(V)$. Let $\tau \in \mathbf{M}(\mathfrak{o})$ be such that the F-subalgebra

$$E := F[\tau] \subseteq \mathbf{M}(F)$$

is a field extension of F having the largest possibly degree, namely n. Let

$$\eta : E^\times \to \mathbb{C}^\times$$

be a character. We consider the following special case:

Assumption 8.23. The element τ satisfies the following conditions:

(1) The image $\bar{\tau}$ of τ in $\mathbf{M}(\mathfrak{o}/\mathfrak{p})$ also generates a degree n extension of the residue field. In particular, $\det(\tau) \in \mathfrak{o}^\times$.

(2) For each $e \in E^\times \cap K(\mathfrak{q})$, we have the compatibility condition

$$\eta(e) = \psi(\mathrm{trace}(\tau(e - 1))).$$

Recall that $\psi : \mathfrak{q}/\mathfrak{q}^2 \to \mathbb{C}^\times$ is a character nontrivial on $\mathfrak{p}^{-1}\mathfrak{q}^2/\mathfrak{q}^2$. As τ normalizes $K(\mathfrak{q})$, so does $E^\times$. The subgroup

$$J_E := E^\times K(\mathfrak{q})$$

is well-defined. It admits the character $\chi : J_E \to \mathbb{C}^\times$ defined by the formula

$$\chi(eg) := \eta(e)\psi(\text{trace}(\tau(g-1))) \quad \text{for } (e,g) \in E^\times \times K(\mathfrak{q}). \tag{8.8}$$

Remark 8.24. The assumption (1) implies that E is an unramified extension of F. It also implies that any nonzero element in $\mathbf{V}(\mathfrak{o}/\mathfrak{p})$ is cyclic with respect to the image $\overline{\tau}$ of τ in $\mathbf{M}(\mathfrak{o}/\mathfrak{p})$, as $\mathbf{V}(\mathfrak{o}/\mathfrak{p})$ is then a one-dimensional vector space over the field $(\mathfrak{o}/\mathfrak{p})[\overline{\tau}] \subseteq \mathbf{M}(\mathfrak{o}/\mathfrak{p})$.

Lemma 8.25. The compact induction

$$\pi := \text{ind}_{J_E}^{\mathbf{G}(F)} \chi$$

is an irreducible supercuspidal representation of $\mathbf{G}(F)$.

Proof. For the notations as above, the pair (J_E, χ) is a maximal type in the sense of [**?**, (6.2)]. Thus by [**?**, (6.2.2)], the representation π constructed above is a supercuspidal representation. $\qquad\square$

Lemma 8.26. The representation π constructed in Lemma 8.25 is regular at depth $\mathfrak{q}^2$, with regular parameter given by τ modulo $\mathfrak{q}$ (see §8.1 for definitions).

Proof. By definition of compact induction, there exists an element $v \in \pi$ which identifies with χ on J_E and vanishes elsewhere. In particular J_E acts on v by χ, and so $K(\mathfrak{q}) \subset J_E$ acts on v by χ_τ as in Definition 8.1.

The element τ is also cyclic in the sense of Definition 3.1, by Remark 8.24 above. Thus τ is a regular parameter for π at depth $\mathfrak{q}^2$. $\qquad\square$

Remark 8.27. We expect that one could verify that every supercuspidal representation with "odd integral depth" and "generic induction datum" arises via Lemma 8.25. Here "depth" should be understood according to the conventions of [**?**, §5.1], while by "generic induction datum", we mean that the parameter β associated to a simple type contained in π (see [**?**, (3.2.1)]) generates a field extension of F of maximal degree, and that β is minimal in the sense of [**?**, (1.4.14)]. For the above construction, if $\mathfrak{q} = \mathfrak{p}^k$, then the "depth" is $2k-1$, while we may take β to be $\varpi^{-2k+1}\tau$.

Lemma 8.28. Let m, n be positive integers, and let π, σ be regular supercuspidal representations of GL_m, GL_n at depth $\mathfrak{q}^2$, constructed above in terms of parameters τ_π and τ_σ. We consider the following two cases:

- *Case $m \neq n$.*
- *Case $m = n$. Then the degree n inert field extensions generated by τ_π and τ_σ can be identified, uniquely up to the action of $\text{Gal}(E/F) \cong \text{Gal}(k_E/k_F)$. We suppose that the images of τ_π and τ_σ in the residue field k_E of E satisfy*

$$\overline{\tau}_\sigma - \overline{\tau}_\pi^\iota \neq 0$$

for all $\iota \in \text{Gal}(k_E/k_F)$.

Then, in either case, the pair (π, σ) is stable at depth $\mathfrak{q}^2$. Furthermore, the local analytic conductor of the Rankin–Selberg pair $\pi \times \sigma$ is

$$C(\pi \times \sigma) = [\mathfrak{o} : \mathfrak{q}]^{2mn}. \tag{8.9}$$

Note that (8.9) generalizes the conductor formula given in Example 8.20, which concerns the special case of characters (i.e., $m = n = 1$).

Proof. Let P_π, P_σ be the polynomials for π and σ at depth $\mathfrak{q}^2$. By Assumption 8.23, part (1), they are irreducible polynomials of degrees m and n. When $m \neq n$, they necessarily generate the unit ideal.

When $m = n$, the assumption on $\overline{\tau}_\pi, \overline{\tau}_\sigma$ implies that the images $\overline{P}_\pi, \overline{P}_\sigma$ of P_π, P_σ mod $\mathfrak{p}$ satisfy $(\overline{P}_\pi, \overline{P}_\sigma) = 1$. By Nakayama's lemma, we then have $(P_\pi, P_\sigma) = 1$.

By Definition 8.6 and Lemma 8.26, we deduce that the pair (π, σ) is stable at depth $\mathfrak{q}^2$.

The conductor formula follows readily from [4, Theorem 6.5(ii)]. Indeed, observe first that π, σ are "completely distinct" in the sense of [4, §6.2], so the cited result is applicable. The ramification indices e_i there are all 1 in our case. The claimed conductor formula follows by writing $\mathfrak{q} = \mathfrak{p}^k$ and taking $m = 2k - 1$ in [4, Theorem 6.5(ii)].   □

Remark 8.29. Note that the additional condition in the $m = n$ case is necessary, as the conductor would otherwise be smaller. It is the analogue of the condition we imposed for a pair of characters in Example 8.20.

9. MAIN LOCAL RESULT

In this section, we retain the general notation and conventions of §2, and denote by $n+1$ the rank of $\mathbf{V}$. We use the notation $A \ll B$ to denote that $|A| \leq C|B|$, where C depends at most upon n. We write $A \asymp B$ to denote that $A \ll B \ll A$.

Let $(F, \mathfrak{o}, \mathfrak{p}, \varpi, q)$ be a non-archimedean local field, with associated data as in §2.4.

9.1. **Statement of result.** The following may be understood as a non-archimedean analogue, uniform with respect to variation of the local field, of the archimedean result [19, Thm 4.2]. It encapsulates the local results needed to prove Theorem 1.1.

Theorem 9.1. *Let $\mathfrak{q} \subseteq \mathfrak{p}$ be a nonzero $\mathfrak{o}$-ideal. Denote by $Q := [\mathfrak{o} : \mathfrak{q}]$ its absolute norm.*

Let π and σ be representations of $\mathbf{G}(F)$ and $\mathbf{H}(F)$ equipped with inner products $\langle,\rangle$ invariant by $K(\mathfrak{q})$ and $K_H(\mathfrak{q})$, respectively. Assume that the pair (π, σ) is stable at depth $\mathfrak{q}^2$ (Definition 8.6). Then there exists

- *a unit vector $v \in \pi$,*
- *a self-adjoint idempotent $\omega \in C_c^\infty(K)$ with $\pi(\omega)v = v$,*
- *a compact open subgroup J_H of K_H, and*
- *a unit vector $u \in \sigma$ that is a J_H-eigenvector,*

with the following properties.

Let $\pi|_Z : \mathbf{Z}(F) \to \mathrm{U}(1)$ denote the central character of π. Define

$$\omega^\sharp(g) := \int_{z \in \mathbf{Z}(F)} \pi|_Z(z)\omega(zg)\,dz.$$

(i) We have

$$\int_{h \in \mathbf{H}(F)} \langle hv, v\rangle\langle u, hu\rangle\,dh \asymp Q^{-n^2}, \tag{9.1}$$

where the integrand is compactly-supported, hence converges absolutely.

(ii) We have

$$\int_{\mathbf{H}(F)} |\omega^\sharp| \ll Q^n. \tag{9.2}$$

(iii) Let $\Psi_1, \Psi_2 : K_H \to \mathbb{C}$ be functions satisfying

$$|\Psi_j(gz)| = |\Psi_j(g)| \text{ for all } (g, z) \in K_H \times J_H.$$

Let

$$\gamma \in \mathbf{G}(F) - \mathbf{H}(F)\mathbf{Z}(F).$$

Then

$$\int_{x,y \in K_H} \left|\Psi_1(x)\Psi_2(y)\omega^\sharp(x^{-1}\gamma y)\right| \ll Q^n \left(\frac{1}{1 + Qd_H(\gamma)} + \frac{d_H(\gamma)^\infty}{Q^*}\right) \|\Psi_1\|_{L^2}\|\Psi_2\|_{L^2}.$$
$$(9.3)$$

Here the notation $d_H(\gamma)$ and $d_H(\gamma)^\infty$ is as in §6.2, while Q^* is as in the statement of Theorem 6.7.

The proof is given in §9.4, following some preliminaries.

9.2. **Stability and matrix coefficients.** Here we adapt some arguments from [20, §19] to the non-archimedean setting, where they simplify considerably.

Lemma 9.2. Let $\tau \in \mathbf{M}_{\mathrm{stab}}(\mathfrak{o}/\mathfrak{q})$.

(i) Let π be a representation of $\mathbf{G}(\mathfrak{o})$, equipped with a $K(\mathfrak{q})$-invariant inner product $\langle,\rangle$, and let $v \in \pi$ be a vector that transforms under $K(\mathfrak{q})$ according to χ_τ. Then for $h \in \mathbf{H}(\mathfrak{o})$, we have

$$\langle hv, v\rangle = 0 \quad \text{unless } h \in K_H(\mathfrak{q}). \qquad (9.4)$$

(ii) Let π be a representation of $\mathbf{G}(F)$ satisfying the same hypotheses as above. Then (9.4) holds for $h \in \mathbf{H}(F)$.

Proof. The group

$$K(\mathfrak{q}) \cap hK(\mathfrak{q})h^{-1} \subseteq K \qquad (9.5)$$

acts on hv by the character $\chi_\tau(h^{-1} \bullet h)$ and on v by χ_τ, so by orthogonality of characters, it suffices to show that whenever $h \notin K_H(\mathfrak{q})$, there exists u in the intersection (9.5) such that

$$\chi_\tau(h^{-1}uh) \neq \chi_\tau(u).$$

We consider two cases.

(i) If $h \in \mathbf{H}(\mathfrak{o})$, then $\chi_\tau(h^{-1}uh) = \chi_{\mathrm{Ad}(h)\tau}(u)$, and $K(\mathfrak{q}) = hK(\mathfrak{q})h^{-1}$. For this reason, it suffices to show that $\mathrm{Ad}(h)\tau \equiv \tau \mod \mathfrak{q}$ if and only if $h \in K_H(\mathfrak{q})$. This is the content of Lemma 4.6.

(ii) The case $h \in \mathbf{H}(F) - \mathbf{H}(\mathfrak{o})$ is addressed by Lemma 9.3, below.

$\square$

Lemma 9.3. Let $\tau \in \mathbf{M}_{\mathrm{stab}}(\mathfrak{o}/\mathfrak{q})$ and $h \in \mathbf{H}(F) - \mathbf{H}(\mathfrak{o})$. Then there exists $u \in K(\mathfrak{q}) \cap hK(\mathfrak{q})h^{-1}$ such that

$$\chi_\tau(h^{-1}uh) \neq \chi_\tau(u). \qquad (9.6)$$

Proof. The proof is an adaptation of [20, Lemma 19.7]. We will show more precisely that there exists

$$u \in K(\mathfrak{q}^2) \cap hK(\mathfrak{q})h^{-1}$$

such that $\chi_\tau(h^{-1}uh) \neq 1$. This suffices in view of the fact that $\chi_\tau(u) = 1$.

We choose a basis $e_1, \ldots, e_n$ for $\mathbf{V}_H(\mathfrak{o})$. We denote by A_H the subgroup of $\mathbf{H}(F)$ diagonalized by this basis. Let A_H^+ denote the subgroup consisting of diagonal matrices whose entries are integral powers of the uniformizer ϖ of $\mathfrak{p}$.

We first reduce to the special case that $h \in A_H^+ - K_H$. To that end, we apply the Cartan decomposition to write
$$h = k_1 a k_2,$$
where $k_1, k_2 \in K_H$ and $a \in A_H^+ - K_H$. Then
$$\chi_\tau(h^{-1}uh) = \chi_{\mathrm{Ad}(k_2)\tau}(a^{-1}k_1^{-1}uk_1a),$$
$$K(\mathfrak{q}^2) \cap hK(\mathfrak{q})h^{-1} = K(\mathfrak{q}^2) \cap k_1aK(\mathfrak{q})a^{-1}k_1^{-1}.$$
We observe that $\mathrm{Ad}(k_2)\tau$ satisfies the same hypotheses as τ. Suppose we can find some $u \in K(\mathfrak{q}^2) \cap aK(\mathfrak{q})a^{-1}$ so that
$$\chi_{\mathrm{Ad}(k_2)\tau}(a^{-1}ua) \neq 1.$$
Then, setting
$$u' := k_1 u k_1^{-1} \in K(\mathfrak{q}^2) \cap hK(\mathfrak{q})h^{-1},$$
we obtain
$$\chi_\tau(h^{-1}u'h) = \chi_{\mathrm{Ad}(k_2)\tau}(a^{-1}ua) \neq 1,$$
as required.

Thus, let $a \in A_H^+ - K_H$. Consider the adjoint action of a on $\mathbf{M}(F)$. Since $a \notin K_H$, there are nontrivial weights for this action. Suppose for instance that there are positive weights (an identical argument will apply if there are negative weights). By conjugating a by a permutation matrix (as we may, by the preceding argument), we may assume that the largest diagonal coordinate of a, say $\varpi^{-\ell}$ with $\ell \geq 1$, appears in components $1, \ldots, m$, where $m \geq 1$. If we use the partition
$$n + 1 = m + (n - m) + 1$$
to describe $\mathbf{M}(F)$ as a space of 3×3 block matrices, then we may describe the weight spaces for a as follows:
$$\begin{pmatrix} 0 & + & + \\ - & * & * \\ - & * & * \end{pmatrix}$$
Here the symbol 0 indicates where a acts trivially, while $+$ (resp. $-$) describe $\mathfrak{o}$-submodules
$$\mathbf{M}_\pm \subseteq \mathbf{M}$$
such that for $x \in \mathbf{M}_\pm(F)$, we have
$$a^{-1}xa = \varpi^{\pm\ell}x;$$
asterisks denote some unspecified combination of trivial, positive and negative weights. We obtain in particular subspaces
$$\bar{M}_\pm := \mathbf{M}_\pm(\mathfrak{o}/\mathfrak{q}) \subseteq \bar{M} = \mathbf{M}(\mathfrak{o}/\mathfrak{q}).$$

Let $\tau_+ \in \bar{M}_+$ denote the component of $\tau \in \bar{M}$. We claim that τ_+ is nonzero modulo $\mathfrak{p}$, i.e., that τ modulo $\mathfrak{p}$ is not of the form
$$\begin{pmatrix} * & 0 & 0 \\ * & * & * \\ * & * & * \end{pmatrix}.$$

Indeed, if it were, then τ modulo $\mathfrak{p}$ would stabilize an m-dimensional subspace of $\bar{V}_H^*$ modulo $\mathfrak{p}$. This contradicts the characterization (ii) of stability recorded in Lemma 4.5, noting that τ modulo $\mathfrak{p}$ is likewise stable (Example 4.3).

The trace pairing puts M_+ and M_- in duality, so that for each $x \in M_-$, we have

$$\text{trace}(x\tau) = \text{trace}(x\tau_+). \tag{9.7}$$

Since τ_+ is nonzero modulo $\mathfrak{p}$, we may find such an x for which

$$\text{trace}(x\tau_+) \in \mathfrak{o} - \mathfrak{p}. \tag{9.8}$$

Let $t \in \mathfrak{p}^{-1}\mathfrak{q}^2 - \mathfrak{q}^2$, to be determined later, and take

$$u := 1 + \varpi^\ell tx.$$

We note that, since $\ell \geq 1$, we have

$$u \in K(\mathfrak{q}^2). \tag{9.9}$$

On the other hand,

$$a^{-1}ua = 1 + tx \in K(\mathfrak{p}^{-1}\mathfrak{q}^2) \subseteq K(\mathfrak{q}), \tag{9.10}$$

while

$$\chi_\tau(a^{-1}ua) = \chi_\tau(1 + tx) = \psi(\text{trace}(tx\tau)).$$

By (9.7) and (9.8), we have $t\,\text{trace}(x\tau) \in \mathfrak{p}^{-1}\mathfrak{q}^2 - \mathfrak{q}^2$. By choosing t suitably, we may thus arrange that

$$\chi_\tau(a^{-1}ua) = \psi(t\,\text{trace}(x\tau)) \neq 1. \tag{9.11}$$

The required conclusion is then immediate from (9.9), (9.10) and (9.11). $\qquad\square$

Lemma 9.4. Let π and σ be representations of $\mathbf{G}(\mathfrak{o})$ and $\mathbf{H}(\mathfrak{o})$, respectively, such that the pair (π, σ) is stable at depth $\mathfrak{q}^2$.

  (i) *There is a stable element $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ such that τ (resp. τ_H) is a regular parameter for π (resp. σ) at depth $\mathfrak{q}^2$.*
 (ii) *Suppose π (resp. σ) is equipped with an inner product $\langle,\rangle$ invariant by $K(\mathfrak{q})$ (resp. $K_H(\mathfrak{q})$). Let τ be any element as above. Let v (resp. u) be unit vectors that transform under $K(\mathfrak{q})$ (resp. $K_H(\mathfrak{q})$) according to χ_τ (resp. χ_{τ_H}). Then*

$$\int_{h \in \mathbf{H}(\mathfrak{o})} \langle hv, v \rangle \langle u, hu \rangle \, dh = \text{vol}(K_H(\mathfrak{q})). \tag{9.12}$$

(iii) *Retaining the hypotheses of (ii), if π and σ arise as restrictions of representations of $\mathbf{G}(F)$ and $\mathbf{H}(F)$, then the formula (9.12) remains valid after extending the integral to $\mathbf{H}(F)$.*

Proof. Let P_π and P_σ be polynomials for π and σ at depth $\mathfrak{q}^2$ that generate the unit ideal. By Lemma 4.4, there exists $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ such that $P_\tau = P_\pi$ and $P_{\tau_H} = P_\sigma$. By Lemma 4.5, τ is stable, and in particular, cyclic. By Lemma 4.7, $\tau_H \in \mathbf{M}_H(\mathfrak{o}/\mathfrak{q})$ is cyclic. By Lemma 8.10, it follows that τ (resp. τ_H) is a regular parameter for π (resp. σ) at depth $\mathfrak{q}^2$.

For the second assertion (ii) concerning the matrix coefficient integral, we apply part (i) of Lemma 9.2 to truncate that integral to $h \in K_H(\mathfrak{q})$, where the integrand evaluates to

$$\chi_\tau(h)\chi_{\tau_H}(h)^{-1} = 1.$$

For the final assertion (iii), we argue similarly using part (ii) of Lemma 9.2. $\qquad\square$

9.3. **Setup for the proof.** We retain the hypotheses of Theorem 9.1. By Lemma 9.4, we obtain a stable element $\tau \in \mathbf{M}(\mathfrak{o}/\mathfrak{q})$ such that τ (resp. τ_H) is a regular parameter for π (resp. σ) at depth $\mathfrak{q}^2$. By Lemma 8.12, we may find unit vectors $v \in \pi$ and $u \in \sigma$ that transform under J_τ and J_{τ_H} by characters $\tilde{\chi}_\tau$ and $\tilde{\chi}_{\tau_H}$ that extend χ_τ and χ_{τ_H}, respectively. We set $(J_G, J_H, \chi_G, \chi_H) := (J_\tau, J_{\tau_H}, \tilde{\chi}_\tau, \tilde{\chi}_{\tau_H})$ and

$$\omega := \mathrm{vol}(J_G)^{-1}\chi_G^{-1} \in C_c^\infty(K),$$

so that $\pi(\omega)$ is a self-adjoint idempotent with $\pi(\omega)v = v$.

We pause to clarify the shape of the function $\omega^\sharp$.

Lemma 9.5. For $g \in \mathbf{G}(F)$, we have

$$|\omega^\sharp(g)| \ll \mathrm{vol}(J_G)^{-1}1_{g \in \mathbf{Z}(F)J_G}.$$

Proof. By definition,

$$\omega^\sharp(g) = \int_{z \in \mathbf{Z}(F)} \pi|_Z(z)\omega(zg)\,dz.$$

The function ω is supported on J_G and has L^∞-norm $\mathrm{vol}(J_G)^{-1}$. It follows that if $g \notin \mathbf{Z}(F)J_G$, then $\omega^\sharp(g)$ vanishes. We have $Z(\mathbf{F}) \cap J_G = K_Z$, so the set of z for which $\omega(zg) \neq 0$ is a K_Z-coset, and so has volume $\ll 1$. We have assumed that π is unitary, so $|\pi|_Z(z)| = 1$. The stated estimate follows from these observations and the triangle inequality. $\qquad\square$

9.4. **The proof.** We now verify each numbered assertion in turn.

(i) We see from Lemmas 9.2 and 9.4 that the integrand in (9.1) is supported on $h \in K_H(\mathfrak{q})$ and the integral evaluates to $\mathrm{vol}(K_H(\mathfrak{q})) \asymp Q^{-n^2}$, as required.

(ii) By Lemma 9.5, we have

$$\int_{\mathbf{H}(F)} |\omega^\sharp| \ll \frac{1}{\mathrm{vol}(J_G)} \int_{h \in \mathbf{H}(F)} 1_{\mathbf{Z}(F)J_G}(h)\,dh. \tag{9.13}$$

We must verify that the above is $\ll Q^n$. Suppose $h \in \mathbf{H}(F)$ may be written $h = zg$ with $(z, g) \in \mathbf{Z}(F) \times J_G$. Then

$$z^{-1}h \in J_G \cap \mathbf{H}(F)\mathbf{Z}(F) = J_G \cap K_H K_Z = (J_G \cap K_H)K_Z,$$

since $K_Z \subseteq J_G$. By Lemma 4.6, we have

$$J_G \cap K_H = K_H(\mathfrak{q}).$$

We have

$$\mathrm{vol}(J_G)^{-1}\,\mathrm{vol}(K_H(\mathfrak{q})) = \frac{[K : J_G]}{[K_H : K_H(\mathfrak{q})]} \asymp \frac{Q^{(n+1)^2 - (n+1)}}{Q^{n^2}} = Q^n. \tag{9.14}$$

The required estimate follows.

(iii) We are given $\Psi_1, \Psi_2 : K_H \to \mathbb{C}$ whose magnitudes are right-invariant by J_{τ_H}, and $\gamma \in \mathbf{G}(F) - \mathbf{H}(F)\mathbf{Z}(F)$. We must verify that

$$I_1 := \int_{x,y \in K_H} \left|\Psi_1(x)\Psi_2(y)\omega^\sharp(x^{-1}\gamma y)\right| \ll \Delta Q^n\|\Psi_1\|_{L^2}\|\Psi_2\|_{L^2}, \tag{9.15}$$

where Δ is the parenthetical quantity on the right hand side of (9.3). By the bound for $\omega^\sharp$ recorded in Lemma 9.5, we see that the left hand side of (9.15) vanishes unless there exists $z \in \mathbf{Z}(F)$ so that

$$z\gamma \in K_H J_G K_H \subseteq K.$$

Replacing γ by $z^{-1}\gamma$ has no effect on either side of the desired estimate, so we may suppose that $\gamma \in K$. Then, since $x, y \in K_H \subseteq K$ and $\mathbf{Z}(F)J_G \cap K = J_G$, we have

$$|\omega^\sharp(x^{-1}\gamma y)| \ll \mathrm{vol}(J_G)^{-1} 1_{x^{-1}\gamma y \in J_G}.$$

The integral I_1 descends to the quotient

$$\bar{H} := K_H/K_H(\mathfrak{q}).$$

For $j = 1, 2$, let $u_j : \bar{H} \to \mathbb{R}_{\geq 0}$ denote the function induced by $|\Psi_j|$. By definition, the image of J_G in $K/K(\mathfrak{q})$ is the centralizer of τ. The integral I_1 thus satisfies

$$I_1 \ll \mathrm{vol}(J_G)^{-1} I,$$

where I is the integral (7.1) defined in the statement of Theorem 7.1, which gives

$$I \ll \Delta |\bar{H}|^{-1} \|u_1\|_{L^2} \|u_2\|_{L^2}.$$

The required estimate for I_1 follows now from (9.14).

10. COMPLETION OF THE PROOF

We now deduce our main result Theorem 1.1, from our main local result, Theorem 9.1. The deduction is exactly as in [19, §6], so we will be brief. We denote in what follows by $\mathbb{Z}_\mathfrak{l} \leq F_\mathfrak{l}$ the ring of integers in the completion of F at a finite place $\mathfrak{l}$.

Proposition 10.1. *Fix $\alpha > 0$ and $\varepsilon > 0$. Retain the setting of Theorem 1.1; in particular, F is a number field, $(G, H) = (\mathrm{U}(V), \mathrm{U}(W))$ is a pair of unitary groups attached to a nondegenerate codimension one inclusion $W \hookrightarrow V$ of positive-definite hermitian spaces over F, S is a large enough finite set of places, $(\pi, \sigma, \mathfrak{p}, \mathfrak{q}) \in \mathcal{F}$, T is the absolute norm of $\mathfrak{q}^2$, and $\vartheta \in [0, 1/2)$ is such that σ is ϑ-tempered at every finite place $\mathfrak{l} \notin S \cup \{\mathfrak{p}\}$ that splits in E. Set $L := T^\alpha$. Let $\omega \in C_c^\infty(G(\mathbb{Z}_\mathfrak{p}))$ and $J_H \leq H(\mathbb{Z}_\mathfrak{p})$ be as in Theorem 9.1 applied to $(\pi_\mathfrak{p}, \sigma_\mathfrak{p})$. There exists $c > 0$, depending only upon α and (G, H, S), with the following property. For $j = 0, \ldots, n + 1$, let Δ_j denote the infimum of all nonnegative quantities with the following property: for all $u_1, u_2 \in L^2(H(\mathbb{Z}_\mathfrak{p}))$ of unit norm that transform on the right under J_H via some unitary character, and all $\gamma \in G(F_\mathfrak{p})$ with $d_{H_\mathfrak{p}}(\gamma) \geq cL^{-j}$, we have*

$$\int_{x,y \in H(\mathbb{Z}_\mathfrak{p})} \left| u_1(x) u_2(y) \omega^\sharp(x^{-1}\gamma y) \right| dx\, dy \leq T^{n/2} \Delta_j. \tag{10.1}$$

Then

$$\frac{\mathcal{L}(\pi, \sigma)}{T^{n(n+1)/2 + \varepsilon}} \ll \sum_{j=1}^{n+1} \left(L^{-(1-2\vartheta)j} + L^{(2(n+1)^2 - n)j} \Delta_j \right)$$

$$+ L^{-1} \sum_{j=0}^{n+1} \left(L^{-(1-2\vartheta)j} + L^{(2(n+1)^2 - n)j} \Delta_j \right),$$

where the meaning of $\ll$ is "bounded in magnitude up to a factor depending only upon $(\mathcal{F}, \alpha, \varepsilon, \vartheta)$".

Proof. This follows from the arguments of [19, §6], applied with $Q = T^{1/2}$. The positive-definiteness assumption is used in two ways: so that the quotient $H(F)\backslash H(\mathbb{A})$ is compact, and so that each of the groups G and H are compact at infinity. This last property is used in applying [19, Lem 5.1].

In more detail, we repeat [19, §6.1-6.4] verbatim (omitting the adjective "archimedean" in a couple places — this property is never used). In the first line of [19, §6.5.1], we observe

that the integrand in [19, (6.5)] is now invariant merely under $\prod_{\mathfrak{l}\notin S\cup\{\mathfrak{p}\}} H(\mathbb{Z}_\mathfrak{l})$, rather than the larger group $\prod_{\mathfrak{l}\notin S} H(\mathbb{Z}_\mathfrak{l})$. (This is because of a difference in conventions: in [19], S contains the interesting place, while here, it does not.) In the subsequent estimates, we accordingly replace the factorizable neighborhood Θ_S with

$$\Theta_{S\cup\{\mathfrak{p}\}} := \Theta_S H(\mathbb{Z}_\mathfrak{p}).$$

The remainder of [19, §6.5.1-6.5.2] is unchanged. The integral on the left hand side of (10.1) arises naturally following the proof[2] of [19, Lem 6.7]. We feed the resulting estimate into the summary of [19, §6.6] to obtain the stated bound. □

We now apply Theorem 9.1, part (iii) to see that the quantities Δ_j as in Proposition 10.1 satisfy

$$\Delta_j \ll \frac{L^j}{T^{1/2}} + \frac{1}{R}, \tag{10.2}$$

where R is given by

$$T = q^{2m} \implies R = q^{\lceil m/2 \rceil}.$$

In particular, R satisfies the slightly wasteful estimate $R \geq T^{1/4}$. Substituting this estimate into the right hand side of the bound stated in Proposition 10.1, we see that two terms dominate, giving

$$\frac{\mathcal{L}(\pi,\sigma)}{T^{n(n+1)/2+\varepsilon}} \ll L^{-(1-2\vartheta)} + L^{(2(n+1)^2-n)(n+1)}\left(\frac{L^{n+1}}{T^{1/2}} + \frac{1}{T^{1/4}}\right). \tag{10.3}$$

Expanding the right hand side as a sum of three terms, the first and third terms are equal if

$$L^{-(1-2\vartheta)} = L^{(2(n+1)^2-n)(n+1)}\frac{1}{T^{1/4}},$$

or equivalently,

$$\alpha = \frac{1}{4(A+1-2\vartheta)}, \quad A := (2(n+1)^2-n)(n+1).$$

We then have $A \geq n+1$, hence $(n+1)\alpha \leq 1/4$, so the second term on the right hand side of (10.3) is dominated by the third term. The estimate (10.3) thus simplifies to the required bound

$$\frac{\mathcal{L}(\pi,\sigma)}{T^{n(n+1)/2+\varepsilon}} \ll T^{-\delta}, \quad \delta := \frac{1-2\vartheta}{4(A+1-2\vartheta)}.$$

References

[1] Ehud Moshe Baruch. Bessel functions for GL(3) over a p-adic field. *Pacific J. Math.*, 211(1):1–33, 2003.

[2] Joseph Bernstein and Andre Reznikov. Subconvexity bounds for triple L-functions and representation theory. *Ann. of Math. (2)*, 172(3):1679–1718, 2010.

[3] Raphaël Beuzart-Plessis, Pierre-Henri Chaudouard, and MichałZydor. The global Gan-Gross-Prasad conjecture for unitary groups: the endoscopic case. *Publ. Math. Inst. Hautes Études Sci.*, 135:183–336, 2022.

[4] Colin J. Bushnell, Guy M. Henniart, and Philip C. Kutzko. Local Rankin-Selberg convolutions for GL_n: explicit conductor formula. *J. Amer. Math. Soc.*, 11(3):703–730, 1998.

[5] W. Duke, J. B. Friedlander, and H. Iwaniec. Bounds for automorphic L-functions. II. *Invent. Math.*, 115(2):219–239, 1994.

[6] W. Duke, J. B. Friedlander, and H. Iwaniec. The subconvexity problem for Artin L-functions. *Invent. Math.*, 149(3):489–577, 2002.

[7] David Eisenbud. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.

---

[2] We note that the statement of [19, Lem 6.7] contains a regrettable typo, introduced in the final revision: $d_{H_\mathfrak{q}}(\gamma)^{-1/2}$ should be $d_{H_\mathfrak{q}}(\gamma)^{-1}$. This does not affect the present argument.

[8] R. Neal Harris. The refined Gross-Prasad conjecture for unitary groups. *Int. Math. Res. Not. IMRN*, (2):303–389, 2014.

[9] Roger E. Howe. On the principal series of Gl_n over p-adic fields. *Trans. Amer. Math. Soc.*, 177:275–286, 1973.

[10] Roger E. Howe. Kirillov theory for compact p-adic groups. *Pacific J. Math.*, 73(2):365–381, 1977.

[11] Atsushi Ichino and Tamutsu Ikeda. On the periods of automorphic forms on special orthogonal groups and the Gross-Prasad conjecture. *Geom. Funct. Anal.*, 19(5):1378–1425, 2010.

[12] Henryk Iwaniec. *Spectral methods of automorphic forms*, volume 53 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2002.

[13] Henryk Iwaniec and Peter Sarnak. L^∞ norms of eigenfunctions of arithmetic surfaces. *Ann. of Math. (2)*, 141(2):301–320, 1995.

[14] Erez Lapid and Zhengyu Mao. A conjecture on Whittaker-Fourier coefficients of cusp forms. *J. Number Theory*, 146:448–505, 2015.

[15] Erez Lapid and Zhengyu Mao. On an analogue of the Ichino-Ikeda conjecture for Whittaker coefficients on the metaplectic group. *Algebra Number Theory*, 11(3):713–765, 2017.

[16] Erez Lapid and Zhengyu Mao. Whittaker-Fourier coefficients of cusp forms on $\widetilde{\mathrm{Sp}}_n$: reduction to a local statement. *Amer. J. Math.*, 139(1):1–55, 2017.

[17] Philippe Michel and Akshay Venkatesh. The subconvexity problem for GL_2. *Publ. Math. Inst. Hautes Études Sci.*, (111):171–271, 2010.

[18] Paul D. Nelson. Bounds for standard L-functions. *arXiv e-prints*, page arXiv:2109.15230, September 2021.

[19] Paul D. Nelson. Spectral aspect subconvex bounds for $U_{n+1} \times U_n$. *Invent. Math.*, 232(3):1273–1438, 2023.

[20] Paul D. Nelson and A. Venkatesh. The orbit method and analysis of automorphic forms. *Acta Math.*, 226(1):1–209, 2021.

[21] S. Rallis and G. Schiffmann. Multiplicity one Conjectures. *ArXiv e-prints*, May 2007.

[22] Peter Sarnak. Fourth moments of Grössencharakteren zeta functions. *Comm. Pure Appl. Math.*, 38(2):167–178, 1985.

[23] Akshay Venkatesh. Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math. (2)*, 172(2):989–1094, 2010.

[24] Wei Zhang. Automorphic period and the central value of Rankin-Selberg L-function. *J. Amer. Math. Soc.*, 27(2):541–612, 2014.

YMSC, TSINGHUA UNIVERSITY
Email address: yhumath@tsinghua.edu.cn

AARHUS UNIVERSITY
Email address: paul.nelson@math.au.dk