ON THE RATE OF EQUIDISTRIBUTION OF EXPANDING
TRANSLATES OF HOROSPHERES IN \( \Gamma \backslash G \)

SAMUEL C. EDWARDS

ABSTRACT. Let \( G \) be a semisimple Lie group and \( \Gamma \) a lattice in \( G \). We generalize a method of Burger to prove precise effective equidistribution results for translates of pieces of horospheres in the homogeneous space \( \Gamma \backslash G \).

1. Introduction

1.1. Background. Let \( G \) be a connected Lie group and \( \Gamma \) a lattice in \( G \). Several outstanding problems, particularly in the field of number theory, have been resolved thanks to breakthroughs in the study of orbits related to various subgroups of \( G \) on the homogeneous space \( \Gamma \backslash G \). Amongst the most well-understood types of orbits on \( \Gamma \backslash G \) are those that involve so-called horospherical subgroups of \( G \). Equidistribution properties of the orbits of such subgroups play a role in the proofs of several more difficult results in the field, as well as in a number of applications in areas such as number theory, mathematical physics, and geometry; for a few examples and further references, cf. Kleinbock and Margulis [19], Marklof [29, 30, 31], Marklof and Strömbergsson [32], and Mohammadi and Oh [35].

Recall that for a one-parameter subgroup \( \{g_t\}_{t \in \mathbb{R}} \) (henceforth denoted \( g_\mathbb{R} \)) of \( G \), the expanding horospherical subgroup \( U^+ \) with respect to \( g_\mathbb{R} \) is defined as

\[
U^+ = \{ h \in G : \lim_{t \to +\infty} g_t^{-1} h g_t = e \},
\]

and the contracting horospherical subgroup \( U^- \) with respect to \( g_\mathbb{R} \) as

\[
U^- = \{ h \in G : \lim_{t \to -\infty} g_t^{-1} h g_t = e \}.
\]

This article is a continuation of work started in [8]: attempting to obtain precise rates of effective equidistribution for expanding translates of pieces of horospherical orbits.

Under certain conditions, translates of horospheres equidistribute in \( \Gamma \backslash G \). Assume that \( g_\mathbb{R} \) is \( \mathbb{R} \)-diagonalizable, i.e. each operator \( \text{Ad}_{g_t} \) (on the Lie algebra \( \mathfrak{g} \) of \( G \)) is diagonalizable over \( \mathbb{R} \). Let \( \mu \) denote the unique \( G \)-invariant Borel probability measure on \( \Gamma \backslash G \). Then for any bounded continuous function \( f \) on \( \Gamma \backslash G \) and any probability measure \( \lambda \) on \( U^+ \) which is absolutely continuous with respect to a Haar measure \( du \) on \( U^+ \), we have (cf., e.g., [19 Proposition 2.2.1])

\[
\lim_{t \to -\infty} \int_{U^+} f(xug_t) \, d\lambda(u) = \int_{\Gamma \backslash G} f \, d\mu \quad \forall x \in \Gamma \backslash G.
\]

One of the most common ways of proving equidistribution statements similar to (1) is by making use of the ubiquitous "Margulis' thickening technique", the behind-lying idea of which originates in the thesis of Margulis, cf. [28]. By imposing certain restrictions on \( g_\mathbb{R} \), \( f \), and \( \lambda \) (1) can be made effective: from [19 Proposition 2.4.8], if the action of \( g_\mathbb{R} \) is exponentially mixing, we have that for \( f \in C^\infty_c(\Gamma \backslash G) \) and \( \chi \in C^\infty_c(U^+) \),

\[
\left| \int_{U^+} \chi(u) f(xug_t) \, du - \int_{U^+} \chi \, du \int_{\Gamma \backslash G} f \, d\mu \right| \ll_{G,f,\chi,x} e^{\delta t} \quad \forall t \leq 0,
\]

Date: November 10, 2018.
for some explicit $\delta > 0$. In this article, we consider the problem of giving a precise bound for this difference, with particular focus on obtaining as great a rate of exponential decay as possible. Our results follow in a line of work by Hejhal [13], Strömbergsson [42], and Södergren [41] studying this type of question on hyperbolic manifolds, i.e., establishing corresponding equidistribution statements for spaces $\Gamma \backslash \SO(n,1)/\SO(n)$. The results mentioned above make extensive use of the spectral theory of automorphic forms on hyperbolic space, and consider only the equidistribution of translates of pieces of closed horospheres. The connection between the spectral theory of automorphic forms, in particular Eisenstein series, and the equidistribution of closed horospheres goes back to Selberg (unpublished work), Zagier [46], and Sarnak [38], who studied the rate of equidistribution for translates of entire closed horocycles in the case $G = \SL(2, \R)$.

In [43], Strömbergsson, while generalizing results of Burger [3] on precise equidistribution statements for horospherical orbits on $\Gamma \backslash G$ in the case $G = \SL(2, \R)$, observed that the method used there can be used to strengthen the results of [12] regarding the equidistribution of translates of pieces of horospheres (cf. [43, Remark 3.4]). We note that Flaminio and Forni [10] also proved precise equidistribution results in the case $G = \SL(2, \C)$, cf. [10] Theorems 1.7 and 5.14. The methods of [10] are somewhat different to those of [3, 43].

The main tool of [3, 43] is a representation-theoretic method first developed by Burger in [3]. At the heart of this method is an identity, [3, Lemma 1], associating the action of $U^+$ with that of $g_\R$ in an arbitrary irreducible unitary representation. In [9], we used a similar identity in the case $G = \SL(2, \C)$ to prove precise effective equidistribution results similar to those of [13, 41, 42], though for translates of pieces of all horospherical orbits; i.e. not just the closed ones. It is this method that we now develop for more general groups $G$.

1.2. Main Results. From now on we let $G$ be a connected semisimple Lie group with finite center, and $\Gamma$ a lattice in $G$ satisfying the assumptions of Langlands (cf. [25, Chapter 2], [37, Chapter 2]). The one-parameter subgroup $g_\R$ is assumed to be $\R$-diagonalizable throughout. We also fix an element $Y \in \g$ such that

$$g_t = \exp(tY) \quad \forall t \in \R.$$ 

These assumptions ensure that there exists a parabolic subgroup $P$ with Langlands decomposition $P = N AM$ (cf. [21, Chapter VII.4]) such that $g_\R \subset A$ and $U^+ = N$; we will review the structure theory of $G$ and its parabolic subgroups in detail in Section 2.1.

As previously mentioned, our method of proof is heavily representation-theoretic. The previous uses of this method [3, 8, 43] are all restricted to the cases $G = \SL(2, \R)$ and $\SL(2, \C)$, and make use of the classification of the unitary dual of $G$, as well as the decomposition of the right-regular representation of $G$ on $L^2(\Gamma \backslash G)$, denoted $\rho$, into irreducible unitary representations. In these cases, this information allows one to relate the rate of equidistribution with the spectrum of the Laplace operator on the hyperbolic orbifold $\Gamma \backslash G/K$, where $K = \SO(2)$ or $K = \SU(2)$. For more general groups $G$, a complete classification of the unitary dual is not available to us. Furthermore, the decomposition of $(\rho, L^2(\Gamma \backslash G))$ is somewhat more complicated, cf. [25]. For these reasons, we instead compare the equidistribution of the relevant translates with the decay of the matrix coefficients of $g_\R$. The main result of this article is that translates of pieces of horospheres equidistribute with the same exponential rate as the matrix coefficients of $g_\R$ decay. As far as we are aware, all previous results on the effective equidistribution of these types of translates either give a worse rate, cf., eg., [19, Proposition 2.4.8], or are restricted to groups $G$ of real rank one (and focus mainly on pieces of closed horospheres), cf. [3, 8, 10, 13, 38, 41, 42, 43, 46].

In order to state our main result, we make a maximal compact subgroup $K$ of $G$, and make the following definition:

**Definition 1.1.** A number $\eta \in \R_{>0}$ is said to be a *rate of decay* for the matrix coefficients of $g_\R$ in a unitary representation $(\pi, \mathcal{H})$ of $G$ if there exist $C$, $q \geq 0$ such that for all $K$-finite
vectors $\mathbf{u}, \mathbf{v} \in \mathcal{H}$ and $t \in \mathbb{R}$,
\begin{equation}
|\langle \pi(g)t \mathbf{u}, \mathbf{v} \rangle| \leq C\|\mathbf{u}\|_{\mathcal{H}}\|\mathbf{v}\|_{\mathcal{H}} (\dim(\pi(K)\mathbf{u}) \dim(\pi(K)\mathbf{v}))^{1/2} (1 + |t|^\eta)e^{-\eta|t|}.
\end{equation}

The decay of matrix coefficients is one of the most studied aspects of the unitary representations of $G$; cf., eg., \cite{4,6,14,36}. We let $L^2(\Gamma \backslash G)$ denote the orthogonal complement of the one-dimensional subspace of $L^2(\Gamma \backslash G)$ consisting of the constant functions. Rates of decay of matrix coefficients for $(\rho, L^2(\Gamma \backslash G))$ have deep connections to many important problems in the study of lattices in Lie groups.

We must also introduce Sobolev norms and spaces; these will be important for quantifying the regularity of the function $f$ and the measure $\lambda$ in \cite{11}. For an open subset $B$ of $U^+$, and a choice of basis for the Lie algebra of $U^+$, we define a Sobolev norm $\| \cdot \|_{W^{k,p}(B)}$ on $C^\infty_c(B)$ by letting $\|\chi\|_{W^{k,p}(B)}$ denote the sum of the $L^p$ norms of all the Lie derivatives of $\chi$ corresponding to monomials in our chosen basis with order not greater than $k$. The closure of $C^\infty_c(B)$ with respect to the norm $\| \cdot \|_{W^{k,p}(B)}$ is denoted $W^{k,p}(B)$.

In an analogous way, we choose a basis for the Lie algebra of $G$, and use this to define Sobolev norms $\| \cdot \|_{S^m(\Gamma \backslash G)}$ (this is done in greater detail in Section 2.3) for a smooth function $f$ on $\Gamma \backslash G$ such that $f$ and all its derivatives are in $L^2(\Gamma \backslash G)$, let $\|f\|_{S^m(\Gamma \backslash G)}^2$ denote the sum of the squares of the $L^2$ norms of $f$ and all its Lie derivatives corresponding to monomials in our chosen basis with order not greater than $m$. The closure of the space of functions $f \in C^\infty(\Gamma \backslash G) \cap L^2(\Gamma \backslash G)$ such that $\|f\|_{S^m(\Gamma \backslash G)} < \infty$ with respect to $\| \cdot \|_{S^m(\Gamma \backslash G)}$ is denoted $S^m(\Gamma \backslash G)$.

Related to Sobolev norms is the invariant height function $\mathcal{Y}_\Gamma$. This function, in conjunction with Sobolev norms, is used to give pointwise bounds for functions in $S^m(\Gamma \backslash G)$. If $\Gamma$ is not cocompact, $\mathcal{Y}_\Gamma(x)$ provides a measure of how far into a cusp the point $x \in \Gamma \backslash G$ is. A stringent definition of $\mathcal{Y}_\Gamma$ is given in Section 1.2.

Our main result can now be stated:

**Theorem 1.** Let $(\rho, V)$ be a subrepresentation of $(\rho, L^2(\Gamma \backslash G)_0)$, and $\eta$ a rate of decay for the matrix coefficients of $g_\mathcal{S}$ in $(\rho, V)$. For any open, relatively compact subset $B$ of $U^+$ there exist $C = C(G, \Gamma, g_\mathcal{S}, B, \eta) > 0$, $W > 0$, $m_1, m_2 \in \mathbb{N}$ such that
\begin{equation}
\left| \int_{U^+} \chi(u)f(xug_\mathcal{S}) \, du \right| \leq C\|\chi\|_{W^{m_2}(B)}\|f\|_{S^m(\Gamma \backslash G)}\mathcal{Y}_\Gamma(x)^{2}(1 + |t|^W)e^{\eta t}
\end{equation}
for all $t \leq 0$, $f \in V \cap S^{m_1}(\Gamma \backslash G)$, $\chi \in W^{m_2}(B)$, and $x \in \Gamma \backslash G$.

**Remark 1.** While it is possible to keep track of, and explicitly state, the dependency of the implied constants on $\eta$ and $g_\mathcal{S}$, it is somewhat tedious to do so. We simply note that the constants are inversely proportional to the quotient of $\eta$ and the “distance” of $Y$ to the walls of the Weyl chamber in which it lies. It is also possible to make the dependency on $B$ completely explicit; it is related to the “covering number” of $B$ with respect to some fixed neighbourhood of the identity in $G$.

**Remark 2.** If we restrict to the case of translates of entire closed horospheres, i.e. when $\Gamma \cap U^+$ is a lattice in $U^+$, $x = \Gamma e$, and $\chi$ is suitably chosen so that the integral in the left-hand side of (3) may be written as $\int_{\Gamma \cap U^+} f(\Gamma g_\mathcal{S}) \, du$, simplifications occur in the proof of Theorem 1 that produce a result similar in nature to \cite{38} Theorem 1. Also, if $G$ is algebraic and $\Gamma$ is arithmetic, the proof of Theorem 1 may once again be adjusted so as to permit $\chi \in W^{m_2}(B)$, with a corresponding change of norm in the right-hand side of (3). Both these points are discussed further in Remarks 3 and 4 below.

As previously stated, the main focus of this article has been the rate of exponential decay in Theorem 1. For this reason, we have assumed that $f$ and $\chi$ are as smooth as needed to obtain this rate. Using Proposition 1 below, one sees that our proof may be slightly adjusted so as to also give an exponential rate of decay for corresponding averages when the test functions
\( \chi \) are restricted to \( W_{0}^{k_{0},\infty}(B) \) for any \( k_{0} \geq 1 \). The rate obtained for such \( \chi \) may, however, be much smaller than the rate of decay of the matrix coefficients of \( g_{R} \). Furthermore, it is unclear whether this rate is greater than that given by using approximations of \( \chi \in W_{0}^{k_{0},\infty}(B) \) by elements of \( W_{0}^{m_{2},\infty}(B) \) together with Theorem 1 \((m_{2} \text{ being as in Theorem 1})\). We also note that the results of [3, 8, 43] give variations of Theorem 1 for \( G = \text{SL}(2, \mathbb{R}) \) or \( \text{SL}(2, \mathbb{C}) \) in the case that \( \chi \) is an indicator function of a subset \( B \subset U^{+} \). It seems possible that our method of proof may also be used to directly (i.e. without having to use approximations by Sobolev functions) give effective equidistribution statements for such \( \chi \) for general groups \( G \) as well, though not without complications; compare Corollary 16 below with [8, Proposition 6]. The rate one obtains in this manner, however, seems to be bounded by \( \min_{a \in \Sigma^{+}(g,a)} \alpha(Y) \), where \( \Sigma^{+}(g,a) \) is the set of positive restricted roots corresponding to the Langlands decomposition \( P = U^{+}AM \). For the cases \( G = \text{SL}(2, \mathbb{R}) \) or \( \text{SL}(2, \mathbb{C}) \), this value is greater than or equal to the rate of decay for the matrix coefficients of \( g_{R} \) in any non-tempered unitary representation of \( G \) (this is due to the fact that in these cases \( G \) has real rank one and the dimension of the restricted root spaces is less than or equal to two). This is not the case though for more general groups \( G \), and at present we do not see how one might prove a result with a rate of exponential decay similar to that of Theorem 1 for indicator functions.

Shifting perspective slightly, we now instead fix a horospherical subgroup and consider the equidistribution of translates of pieces of \( \pi \) in a unitary representation \((\pi, \mathcal{H})\) of \( G \) if there exist \( C, q \geq 0 \) such that for all \( K \)-finite vectors \( u, v \in \mathcal{H} \) and all \( H \in \mathfrak{a}^{\mp} \),

\[
|\langle \pi(\exp(H))u, v \rangle| \leq C\|u\|\|v\| (\dim(\pi(K)u)\dim(\pi(K)v))^{1/2} (1 + \|H\|^{q}) e^{-\|H\|}.
\]

(4)

The following theorem shows that the translates of pieces of \( N \) equidistribute with the same speed as the matrix coefficients decay, essentially uniformly over all \( \mathfrak{a}^{\mp} \):

**Theorem 2.** Let \((\rho, V)\) be a subrepresentation of \((\rho, L^{2}(\Gamma \backslash G)_{0})\), and \( \| \cdot \| \) a norm on \( \mathfrak{a} \) that controls the decay of matrix coefficients of \( \mathfrak{a}^{\mp} \) in \((\rho, V)\). Then there exist constants \( c \) (depending only on \( \mathfrak{a} \) and \( \| \cdot \| \)) and \( W \) (depending only on \( G \) and \( A \)) such that for any relatively compact subset \( B \subset N \) and \( 0 < \epsilon < \min\{1, (2c)^{-1}\} \), there exists \( C = C(G, \Gamma, A, B, \| \cdot \|, \epsilon) \) such that if \( m_{1} = W(\left[\frac{2}{2}\right] + 3) + 1 + \left\lceil \frac{\dim G}{2} \right\rceil \) and \( m_{2} = \left[\frac{2}{2}\right] \), we have

\[
\int_{N} \chi(n)f(xn \exp(-H)) \, dn \leq C\|\chi\|W_{m_{2},\infty}(B)\|f\|_{S^{m_{1}}(\Gamma \backslash G)}\mathcal{V}_{\Gamma}(x)^{2} (1 + \|H\|^{W+1}) e^{-(1-c\epsilon)\|H\|}
\]

for all \( H \in \mathfrak{a}^{\mp} \), \( f \in V \cap S^{m_{1}}(\Gamma \backslash G) \), \( \chi \in W_{0}^{m_{2},\infty}(B) \), and \( x \in \Gamma \backslash G \).

**Remark 3.** As was the case for Theorem 1 in order to give a (relatively) simple statement, we have refrained from stating (and keeping track of) the dependency of the constant \( C \) on the various parameters. From the proofs, one sees that it is possible to make many of these dependencies completely explicit. It can also be seen that the constant \( W \) is bounded over all
choices of $A$ by the order of the Weyl group of $G$. Furthermore, given a set $K \subset \mathfrak{a}^{\perp}$ whose intersection with $\mathfrak{a}^{\perp}$ and the relative interior of any face of $\mathfrak{a}^{\perp}$ is compact, there exists $\epsilon_K > 0$ such that if $\epsilon \leq \epsilon_K$, (5) holds with $e^{-\|H\|} \epsilon$ in place of $e^{-(1-\epsilon)\|H\|}$ for all $H \in \mathbb{R}K$. Finally, it can be shown that for a given norm $\|\cdot\|$ and compact subset $C \subset \mathbb{R}^{+}$, the dependency of the constants $C$ and $c$ with respect to the choice of norm is uniformly bounded over the set \{\|\cdot\| : \eta \in C\}.

**Remark 4.** Recently, the equidistribution of translates of pieces of $N$ by elements of $G$ that are not necessarily in $A^{\perp}$ has been studied by a number of authors, cf. \cite{[7, 20, 21, 22, 54, 39, 10]}. Another line of recent study has been the equidistribution of translates of horospheres in the case where $G$ is of rank one and $\Gamma$ a geometrically finite group, but not necessarily a lattice, cf., eg., \cite{[17, 18, 26, 35]}. These equidistribution results have applications to a number of problems in number theory. It would be interesting to see if the method developed here can be adapted to these settings.

1.3. **Outline of Article.** Since our line of proof is quite similar to that of \cite{[8]}, the general structure of this article is more or less the same. After reviewing the necessary background material in Section 2, in Section 3 we turn our attention to proving the main technical result required for the proof of Theorem 1, a generalization of “Burger’s formula” \cite{[3, Lemma 1]} for general semisimple Lie groups. The first order of business is to establish a Lie algebra identity, Lemma 5, which will enable us to express an ordinary differential equation (with respect to $t$) for averaging operators, in irreducible unitary representations $(\pi, \mathcal{H})$, of the form $v \mapsto \int_{U^+} \chi(u\pi(u)g) \mathcal{H} = \chi(\pi(u)g) \mathcal{H} \, du$ (where $v \in \mathcal{H}$). We then use this differential equation to give an integral representation of the averaging operators (see Proposition 6); it is this integral representation that corresponds to \cite{[3, Lemma 1]} and \cite{[8, Proposition 4]}. Unlike in \cite{[3, 8]}, this integral representation is not sufficient for our purposes, and we must instead make use of the differentiability of $\chi$ and an iterative application of Proposition 8 to obtain an integral identity which suits our needs. This is done in Proposition 10.

We then proceed to prove the required properties of the invariant height function $\mathcal{Y}_\Gamma$. The main tool here is the reduction theory of lattices in semisimple Lie groups; using this, we prove that $\mathcal{Y}_\Gamma \in L^1(\Gamma \backslash G)$ (Proposition 14 and Corollary 15). This fact, together with a “thickening” argument allows us to give a bound $\int_B \mathcal{Y}_\Gamma(xu) \, du \ll_{\Gamma, B} \mathcal{Y}_\Gamma(x)^2$ for all $x \in \Gamma \backslash G$, see Corollary 16 below. Matters here are considerably less complicated than when proving the corresponding results needed when considering “sharp cut-off functions” $\chi$ in the case $G = SL(2, \mathbb{C})$ (cf. \cite{[8, Proposition 6]}).

Section 5 is devoted to the proof of Theorem 1. Having established the necessary pre-requisites, the proof is fairly straightforward: the representation $\rho, \gamma$ is decomposed into irreducible unitary representations $(\pi, \mathcal{H})$ and the identity (37) of Proposition 10 is used in each of these. The bounds provided in Proposition 10 then suffice to exchange the direct integral decomposition with the integrals occurring in (37). These bounds are then used once again, together with Corollary 16 to give the bound stated in Theorem 1.

Finally, in Section 6, we prove Theorem 2. Firstly, for each $\epsilon > 0$ we construct certain compact subsets of $\mathfrak{a}^{\perp}$ on which we will prove that Theorem 1 holds uniformly. We then review the results from Sections 3 and 4, making relatively minor adjustments in order to obtain statements which are uniform in a new parameter $H \in \mathfrak{a}^{\perp}$. The proof of Theorem 2 is based on the observation that every $H \in \mathfrak{a}^{\perp}$ such that $\|H\| > 0$ may be decomposed as $H = H_e + J_e$, with $\frac{H}{\|H\|}$ lying in one of the compact subsets constructed at the beginning of the section. Applying Theorem 1 with respect to $Y = \frac{H}{\|H\|}$ and showing that the contribution from $J_e$ is negligible concludes the proof.

1.4. **Acknowledgements.** The research leading to these results was funded by Swedish Research Council Grant 621-2011-3629. I am grateful to my advisor Andreas Strömbergsson for
2. Preliminaries

2.1. Structure Theory. We start by reviewing the structure theory of (real) semisimple Lie groups and their parabolic subgroups; our source for this is [24] Chapters 6 and 7, and we summarize some of the results there that we will require.

Recall that we assume throughout that $G$ is a connected Lie group with finite center whose Lie algebra $\mathfrak{g}$ is semisimple. We have also fixed a maximal compact subgroup $K$ of $G$, the Lie algebra of which is denoted $\mathfrak{k}$. Let $\theta$ be a Cartan involution of $\mathfrak{g}$ such that $\mathfrak{k}$ is the $-1$-eigenspace for $\theta$. Denoting the $+1$-eigenspace of $\theta$ by $\mathfrak{p}$, we have the corresponding Cartan decomposition $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$; recall that the direct sum is orthogonal with respect to the Killing form $B = B(\cdot, \cdot)$ of $\mathfrak{g}$. We now choose and fix once and for all a maximal abelian subalgebra $\mathfrak{a}_0$ of $\mathfrak{p}$. Note that any other choice of maximal abelian subalgebra of $\mathfrak{p}$ is of the form $Ad_\kappa \mathfrak{a}_0$ for some $k \in K$. Let $\Sigma(\mathfrak{g}, \mathfrak{a}_0)$ denote the restricted roots of $\mathfrak{a}_0$; this is the set of all $\lambda \in \mathfrak{a}_0^* \setminus \{0\}$ such that the corresponding root space $\mathfrak{g}_\lambda = \{ X \in \mathfrak{g} : ad_H X = \lambda(H) \ \forall H \in \mathfrak{a}_0 \}$ is non-zero.

A “notion of positivity on $\mathfrak{a}^*$” (cf. [24] Chapter II.5]) is fixed, and the set of positive restricted roots for $\mathfrak{a}_0$ is denoted $\Sigma^+(\mathfrak{g}, \mathfrak{a}_0)$. Let $\Sigma^+_0(\mathfrak{g}, \mathfrak{a}_0) \subset \Sigma^+(\mathfrak{g}, \mathfrak{a}_0)$ denote the set of simple (positive) restricted roots; recall that this is the unique subset such that every element of $\Sigma^+(\mathfrak{g}, \mathfrak{a}_0)$ may be written as a non-negative integral linear combination of elements $\Sigma^+_0(\mathfrak{g}, \mathfrak{a}_0)$. A useful fact is that the elements of $\Sigma^+_0(\mathfrak{g}, \mathfrak{a}_0)$ constitute a basis of $\mathfrak{a}_0^*$. Letting $\mathfrak{n}_0 = \bigoplus_{\lambda \in \Sigma^+_0(\mathfrak{g}, \mathfrak{a}_0)} \mathfrak{g}_\lambda$ gives rise to an Iwasawa decomposition of $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{n}_0 \oplus \mathfrak{a}_0 \oplus \mathfrak{k}.$$

We let $A_0$ be the analytic subgroup of $G$ corresponding to $\mathfrak{a}_0$ and $N_0$ the analytic subgroup corresponding to $\mathfrak{n}_0$. Both $A_0$ and $N_0$ are simply connected, $A_0$ is abelian, and $N_0$ is Ad-unipotent. The corresponding Iwasawa decomposition of $G$ is $G = N_0 A_0 K$. Recall that the map $N_0 \times A_0 \times K \to G$, $(n, a, k) \mapsto nak$ is a diffeomorphism onto.

We now turn our attention to the parabolic subgroups of $G$. Here we follow [23] Chapter V.5]. The subgroup $P_0 = N_G(N_0)$ is fixed as the standard minimal parabolic subgroup. Recall that then any closed subgroup of $G$ containing $P_0$ is called a standard parabolic subgroup, and a parabolic subgroup is any subgroup of $G$ which is conjugate to a standard parabolic subgroup. Letting $\mathfrak{m}_0 = Z_G(\mathfrak{a}_0)$ and $M_0 = Z_K(\mathfrak{a}_0)$, we obtain a Langlands decomposition of $P_0$: $P_0 = N_0 A_0 M_0$. If $q$ is the Lie algebra of a parabolic subgroup $P$, then there is a Langlands decomposition of $q$:

$$q = n \oplus \tilde{a} \oplus \tilde{m},$$

where $n$, $\tilde{a}$, and $\tilde{m}$ are the subalgebras of $\mathfrak{g}$ which are uniquely defined by the following properties:

i) $n$, $\tilde{a}$, and $\tilde{m}$ are orthogonal with respect to the inner product on $\mathfrak{g}$ defined by $(X_1, X_2) := -B(X_1, \theta X_2)$.

ii) $\tilde{a} \oplus \tilde{m} = q \cap \theta q$.

iii) $\tilde{a} = p \cap Z_{\mathfrak{k} @ \mathfrak{m}}$.

The Langlands decomposition described above may be used to give other Langlands decompositions of $q$: let $a = Ad_k \tilde{a}$ for some $k \in N_K(P)$ and $m = \{ J \in Z_\mathfrak{g}(a) : \langle H, J \rangle = 0 \ \forall H \in a \}$. We then have another Langlands decomposition of $q$:

$$q = n \oplus a \oplus m.$$
Moreover, the subset \( F \) of parabolic subgroups, and we may write

\[
\mathfrak{n} = \bigoplus_{\lambda \in \Sigma^+(\mathfrak{g}, \mathfrak{a})} \mathfrak{g}_\lambda,
\]

\( \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) being the set of positive elements in \( \Sigma(\mathfrak{g}, \mathfrak{a}) \). As was the case for \( \mathfrak{a}_0 \), there then exists a unique subset \( \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \subset \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) such that \( \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \) forms a basis for \( \mathfrak{a}^* \), and the coordinates of every element of \( \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) with respect to this basis are non-negative integers.

An important subset of \( \mathfrak{n} \) and the map \( \lambda : \Sigma^+(\mathfrak{g}, \mathfrak{a}) \to \mathbb{R} \) is uniquely determined by \( \lambda(\mathfrak{a}) = 1 \). We conclude this discussion by recalling the explicit construction of standard parabolic subgroups (cf. [24, Chapter VII.7]). Assume now that \( \lambda \) is a positive Weyl chamber \( \mathfrak{a}^+ \):

\[
\mathfrak{a}^+ = \left\{ H \in \mathfrak{a} : \lambda(H) > 0 \quad \forall \lambda \in \Sigma^+(\mathfrak{g}, \mathfrak{a}) \right\} = \left\{ H \in \mathfrak{a} : \lambda(H) > 0 \quad \forall \lambda \in \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \right\}.
\]

The (topological) closure of \( \mathfrak{a}^+ \) is denoted \( \overline{\mathfrak{a}^+} \). Now defining

\[
\mathfrak{n}^- = \theta \mathfrak{n} = \bigoplus_{\lambda \in \Sigma^+(\mathfrak{g}, \mathfrak{a})} \mathfrak{g}_{-\lambda},
\]

we see that

\[
\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m} \oplus \mathfrak{n}^-.
\]

Another, more general, type of Langlands decomposition will be needed when discussing the reduction theory of lattices in \( G \); as this will only be needed in Section II.1 we postpone the introduction of this until then.

A Langlands decomposition of \( \mathfrak{g} \) gives rise to a corresponding Langlands decomposition of \( P \): let \( N, A \) and \( M' \) be the analytic subgroups of \( G \) with Lie algebras \( \mathfrak{n}, \mathfrak{a}, \) and \( \mathfrak{m}, \) respectively. Letting \( M = M'Z_K(\mathfrak{a}) \), we have

\[
P = NAM,
\]

and the map \( N \times A \times M \to P, (n, a, m) \mapsto nam \) is a diffeomorphism onto. Similarly to the groups \( N_0, A_0, N \) and \( A \) are simply connected, \( A \) is abelian, and \( N \) is Ad-unipotent. The exponents of the positive Weyl chamber \( \mathfrak{a}^+ \) and its closure \( \overline{\mathfrak{a}^+} \) are denoted \( \mathfrak{a}^+ \) and \( \overline{\mathfrak{a}^+} \), respectively. We make two final definitions that will prove useful when integrating over \( G \): let \( N^- \) be the analytic subgroup with Lie algebra \( \mathfrak{n}^- \), and

\[
\rho_a = \frac{1}{2} \sum_{\lambda \in \Sigma^+(\mathfrak{g}, \mathfrak{a})} \dim(\mathfrak{g}_\lambda) \lambda \in \mathfrak{a}^*.
\]

Since \( A \) is simply connected, we may also view \( \rho_a \) as a function on \( A \); by abusing notation slightly, we write \( \rho_a(a) = \rho_a(\log_a a) \) for \( a \in A \).

We conclude this discussion by recalling the explicit construction of standard parabolic subgroups (cf. [24, Chapter VII.7]). Assume now that \( P = NAM \) is a standard parabolic subgroup. There then exists a subset \( \mathcal{F} \subset \Sigma_0^+(\mathfrak{g}, \mathfrak{a}_0) \) such that

\[
\mathfrak{a} = \mathfrak{a}_\mathcal{F} = \left\{ H \in \mathfrak{a}_0 : \alpha(H) = 0 \quad \forall \alpha \in \mathcal{F} \right\},
\]

\[
\mathfrak{n} = \mathfrak{n}_\mathcal{F} = \bigoplus_{\lambda \in \Sigma^+(\mathfrak{g}, \mathfrak{a}_0)} \mathfrak{g}_\lambda
\]

\[
\mathfrak{m} = \mathfrak{m}_\mathcal{F} = \left\{ J \in Z_\mathfrak{a}_\mathcal{F} : B(H, J) = 0 \quad \forall H \in \mathfrak{a}_\mathcal{F} \right\}.
\]

Moreover, the subset \( \mathcal{F} \) uniquely determines \( P \); there are therefore exactly \( 2^{\dim \mathfrak{a}_0} \) standard parabolic subgroups, and we may write \( P = P_\mathcal{F} = P_\mathfrak{F} A_\mathcal{F} M_\mathcal{F} \). Given subsets \( \mathcal{F}_1 \subset \mathcal{F}_2 \subset \Sigma_0^+(\mathfrak{g}, \mathfrak{a}_0) \), we have \( P_{\mathcal{F}_1} \subset P_{\mathcal{F}_2}, M_{\mathcal{F}_1} \subset M_{\mathcal{F}_2}, A_{\mathcal{F}_1} \subset A_{\mathcal{F}_2}, N_{\mathcal{F}_2} \subset N_{\mathcal{F}_1} \), and \( \rho_{\alpha_{\mathcal{F}_2}} = \rho_{\alpha_{\mathcal{F}_1}} |_{\mathfrak{a}_{\mathcal{F}_2}} \). Finally, observe that \( P_0 = P_\emptyset \) and \( G = P_{\Sigma_0^+(\mathfrak{g}, \mathfrak{a}_0)} \).
2.2. Measures and Integration. We will denote the Haar measure on $G$ by $d\mu$, or $dg$, and the corresponding measure on $\Gamma \backslash G$ by $\mu_{\Gamma \backslash G}$ or $dx$ (recall that the Haar measure on $G$ has been normalized so that $(\Gamma \backslash G, \mu_{\Gamma \backslash G})$ is a probability space). We will use similar notation for unimodular subgroups $H$ of $G$, i.e., $\mu_H$ or $dh$ will denote a Haar measure on $H$. For subgroups $H$ which are not unimodular, $dh$ and $drh$ will be used to denote left and right Haar measures, respectively. The Haar measure on $G$ can be decomposed with respect to various subgroups, cf. [24, Chapter VIII.3]; the most important decomposition for us is the following (cf. [24, Proposition 8.45]): let $P = NAM$ be a parabolic subgroup. Then $NAM^{-1}$ is an open dense subset of $G$, and the Haar measures on $N, A, M, N^{-1}$ can be normalized so that $dg = e^{-2\rho_a(a)} d\pi dm da dn$, i.e. for $f \in C_c(G)$,

$$\int_G f(g) dg = \int_N \int_A \int_M \int_{N^{-1}} f(nam\pi) e^{-2\rho_a(a)} d\pi dm da dn.$$  

(6)

2.3. Representation Theory. As in [8], the key ingredient in the proofs of our main results is the theory of unitary representations. Recall that a unitary representation of $G$ is a pair $(\pi,\mathcal{H})$, where $\mathcal{H}$ is a separable Hilbert space, and $\pi$ is a homomorphism of topological groups from $G$ to the group of unitary operators on $\mathcal{H}$ (equipped with the strong operator topology). We say that a unitary representation $(\pi,\mathcal{H})$ is irreducible if $\mathcal{H}$ has no non-trivial closed subspaces that are invariant under $\pi(G)$. The irreducible unitary representations may be viewed as the "building blocks" for all other unitary representations in the following way: given a unitary representation $(\pi,\mathcal{H})$ of $G$, there exist a locally compact Hausdorff space $Z$ and a positive Radon measure $\nu$ on $Z$ such that

$$\pi \cong \left( \int_Z \pi_{\zeta} d\nu(\zeta), \int_Z \mathcal{H}_{\zeta} d\nu(\zeta) \right),$$

(7)

where $(\pi_{\zeta},\mathcal{H}_{\zeta})$ is an irreducible unitary representation of $G$ for $\nu$-a.e. $\zeta$ (cf., e.g., [44, Corollary 14.9.5]). This is called the direct integral decomposition of $(\pi,\mathcal{H})$. In Section 3 we will prove certain identities for operators in irreducible unitary representations of $G$. The direct integral decomposition of $(\rho,\mathcal{H})$ will then be used to prove that similar identities hold in $(\rho,\mathcal{H})$.

The complexification of $\mathfrak{g}$ is denoted $\mathfrak{g}_{\mathbb{C}}$; similar notation is used for subalgebras of $\mathfrak{g}$. We let $\mathcal{U}(\mathfrak{g}_{\mathbb{C}})$ denote the universal enveloping algebra of $\mathfrak{g}_{\mathbb{C}}$, and the terms in the canonical filtration of $\mathcal{U}(\mathfrak{g}_{\mathbb{C}})$ are denoted $\mathcal{U}^m(\mathfrak{g}_{\mathbb{C}})$, where $m \in \mathbb{N}$. Recall that a smooth vector of $(\pi,\mathcal{H})$ is a vector $v \in \mathcal{H}$ such that the map $G \to \mathcal{H}, g \mapsto \pi(g)v$ is a $C^\infty$-function. The subspace of smooth vectors is denoted $\mathcal{H}^{\infty}$; by a well-known result of Garding, $\mathcal{H}^{\infty}$ is dense in $\mathcal{H}$. We then have a (Lie algebra) representation $d\pi$ of $\mathfrak{g}$ on $\mathcal{H}^{\infty}$, called the derived representation. For $X \in \mathfrak{g}$, $d\pi(X)$ is defined through

$$d\pi(X)v := \left. \frac{d}{dt} \pi(\exp(tX))v \right|_{t=0}, \quad v \in \mathcal{H}^{\infty}. $$

(8)

This extends (by linearity and composition) to a representation $d\pi$ of $\mathcal{U}(\mathfrak{g}_{\mathbb{C}})$ in the canonical way. Of particular importance to us is the action of the center of $\mathcal{U}(\mathfrak{g}_{\mathbb{C}})$, which we denote by $Z(\mathfrak{g}_{\mathbb{C}})$. By Schur’s lemma, if $(\pi,\mathcal{H})$ is irreducible, then for each $Z \in Z(\mathfrak{g}_{\mathbb{C}})$ there exists a scalar $\lambda_Z \in \mathbb{C}$ such that $d\pi(Z)v = \lambda_Z v$ for all $v \in \mathcal{H}^{\infty}$.

While it is common to consider the $K$-finite vectors of a representation $(\pi,\mathcal{H})$ and use the dimensions of the spaces of $K$-translates of such vectors to help quantify certain aspects of the action of $G$ on $\mathcal{H}$ (cf., e.g., [23, Chapters VII & VIII] as well as Definitions 1.1 and 1.2, for our purposes it is more convenient to use Sobolev norms for $\pi$ on $\mathcal{H}^{\infty}$. We choose a basis $\{X_j\}_{j=1, \ldots, \dim \mathfrak{g}}$ of $\mathfrak{g}$, and define the Sobolev norm of order $m$, $\| \cdot \|_{S^m(\mathcal{H})}$, on $\mathcal{H}^{\infty}$ by

$$\|v\|^2_{S^m(\mathcal{H})} := \sum_U \| d\pi(U)v \|^2, \quad v \in \mathcal{H}^{\infty},$$

(9)

the sum running over all monomials $U$ in the elements $X_j$ of degree not greater than $m$. Note that this sum includes the term "1" of order zero. While different bases of $\mathfrak{g}$ give rise to different Sobolev norms, these norms will be equivalent. The closure of $\mathcal{H}^{\infty}$ with respect to the norm
\[ \| \cdot \|_{S^m(H)} \] is denoted \( S^m(H) \). For notational convenience, in the case \( (\pi, H) = (\rho, L^2(\Gamma \backslash G)) \) we denote \( S^m(L^2(\Gamma \backslash G)) \) by \( S^m(G) \). Note that for our subrepresentation \((\rho, V)\) we have \( \| f \|_{S^m(V)} = \| f \|_{S^m(G)} \) for all \( f \in V^\infty \) and hence \( S^m(V) = V \cap S^m(G) \).

For any \( m \geq 0 \), there exists a continuous function \( C : G \to \mathbb{R}_{>0} \) (depending only on \( G, m \) and the choice of basis \( \{ X_\nu \}_{\nu=1,...,\dim g} \)) such that for any unitary representation \((\pi, H)\),

\[
\| \pi(g)v \|_{S^m(H)} \leq C(g) \| v \|_{S^m(H)} \quad \forall g \in G, \ v \in H^\infty.
\]

The direct integral decomposition of a unitary representation \((\pi, H)\) into irreducible representations given in (7) may be used to decompose norms on \( H \) and \( H^\infty \): for \( v \in H \),

\[
\| v \|_H^2 = \int_Z \| v_\zeta \|_{H_\zeta}^2 \, d\nu(\zeta),
\]

and

\[
\pi(g)v = \int_Z \pi_\zeta(g)v_\zeta \, d\nu(\zeta).
\]

Moreover,

\[
\| v \|_{S^m(H)}^2 = \int_Z \| v_\zeta \|_{S^m(H_\zeta)}^2 \, d\nu(\zeta) \quad \forall v \in H^\infty, \ m \in \mathbb{N}.
\]

Direct integral decompositions allow us to give explicit constructions of intertwining operators; recall that for a representation \((\pi, H)\), these are the operators that commute with \( \pi(g) \) for all \( g \in G \). Assume that a representation \((\pi, H)\) has a direct integral decomposition as in (7), and let \( f : Z \to \mathbb{C} \) be a bounded, measurable function. We then define an operator \( T_f \) on \( H \) through

\[
T_f v := \int_Z f(\zeta)v_\zeta \, d\nu(\zeta) \quad \forall v \in H.
\]

This then gives

\[
T_f \pi(g)v = \int_Z f(\zeta)\pi_\zeta(g)v_\zeta \, d\nu(\zeta) = \int_Z \pi_\zeta(g)f(\zeta)v_\zeta \, d\nu(\zeta) = \pi(g)T_f v,
\]

as well as \( \| T_f v \| \leq (\sup_{\zeta \in Z} |f(\zeta)|) \| v \| \) for all \( v \in H \). We will have use of intertwining operators \( T_f \) for measurable functions \( f \) which are unbounded on \( Z \). Operators of this type will, in general, be unbounded on \( H \). The operators of this kind that will be of use to us will, however, be bounded operators on \( H^\infty \) with respect to suitably chosen Sobolev norms.

2.4. Decay of Matrix Coefficients. We review some material regarding the quantitative decay of matrix coefficients of unitary representations. Two results regarding rates of decay of matrix coefficients will be needed; let \((\pi, H)\) be an arbitrary unitary representation of \( G \).

We then have the following:

**Lemma 3.** Suppose that \((\pi, H) \cong (f_\zeta \pi_\zeta \, d\nu(\zeta), f_\zeta \mathcal{H}_\zeta \, d\nu(\zeta))\), where each \((\pi_\zeta, \mathcal{H}_\zeta)\) is an irreducible unitary representation of \( G \).

i) If \( \eta \) is a rate of decay for the matrix coefficients of \( g_\mathbb{R} \) in \((\pi, H)\), then \( \eta \) is a rate of decay for the matrix coefficients of \( g_\mathbb{R} \) in \((\pi_\zeta, \mathcal{H}_\zeta)\) for \( \nu\text{-a.e.} \zeta \).

ii) Let \( P = NAM \) be a parabolic subgroup of \( G \). If \( \| \cdot \| \) controls the rate of decay for the matrix coefficients of \( A^\infty \) in \((\pi, H)\), then \( \| \cdot \| \) controls the rate of decay for the matrix coefficients of \( A^\infty \) in \((\pi_\zeta, \mathcal{H}_\zeta)\) for \( \nu\text{-a.e.} \zeta \).

**Proof.** By e.g. [9], Theorem 1.7, for \( \nu\text{-a.e.} \zeta \), \((\pi_\zeta, \mathcal{H}_\zeta)\) is weakly contained in \((\pi, H)\). The unitary dual of \( K \) is denoted by \( \hat{K} \); we then have the orthogonal decomposition \( H = \bigoplus_{\tau \in \hat{K}} H_\tau \), where \((\pi|_K, \mathcal{H}_\tau)\) is isomorphic to a number (possibly countably infinite) of copies of \( \tau \). Let \( \Psi \)
denote the function on \( \hat{K} \) defined by \( \Psi(\tau) = \dim \tau \). By \[23\] p. 206 (1), \( \dim(\pi(K)v) \leq \Psi(\tau)^2 \) for all \( v \in \mathcal{H}_\tau \). In order to prove \( i \), we let \( C \) and \( q \) be as in \( \[2\] \), and define
\[
\Phi(g) = \begin{cases} 
C(1 + |t|^q)e^{-\eta|t|} & \text{if } g \in KgK \\
1 & \text{otherwise}.
\end{cases}
\]

In the case of \( ii \), without loss of generality we assume that \( a = a_F \), for some \( F \subset \Sigma^+_0(g, a_0) \), and hence \( a \subset a_0 \). We now let \( C \) and \( q \) be as in \( \[4\] \), and define
\[
\Phi(g) = C \min_{w \in N_K(a_0)/Z_K(a_0)} (1 + \|\text{Ad}_w H\|^q)e^{-\|\text{Ad}_w H\|}
\]
if \( g \in K \exp(\pm H)K \) for some \( H \in a^\tau \), and \( \Phi(g) = 1 \) otherwise. In either case, by \[24\] Theorem 7.39 and the fact that \( N_K(a_0)/Z_K(a_0) \) is a finite group, \( \Phi \) is well-defined. Furthermore, \( \Phi \) is bi \( K \)-invariant and \( \Phi(g) = \Phi(g^{-1}) \) for all \( g \in G \). Using the language of \[14\] Chapter 6, \[2\] or \[4\] thus implies that \( (\pi, \mathcal{H}) \) is \( (\Phi, \Psi) \)-bounded (see \[14\] (6.8))). By \[14\] Lemma 6.2(b) and \[14\] p. 287 j) and we then have that for \( v \)-a.e. \( \zeta, (\pi_\zeta, \mathcal{H}_\zeta) \) is \( (\Phi, \sqrt{2}\Psi) \)-bounded. Since in case \( i \) \( \Phi(g_t) \leq C(1 + |t|^q)e^{-\eta|t|} \), and in case \( ii \) \( \Phi(\exp(H)) \leq C(1 + \|H\|^q)e^{-\|H\|} \) for all \( H \in a^\tau \), we are done.

**Lemma 4.**

i) If \( \eta \) is a rate of decay for the matrix coefficients of \( g_\tau \) in \( (\pi, \mathcal{H}) \), then there exist \( m \in \mathbb{N} \), and \( C, q \geq 0 \) such that for all \( t \in \mathbb{R}, u, v \in \mathcal{H}_\tau^\infty \),
\[
|\langle \pi(g_t)u, v \rangle| \leq C\|u\|_{S^m(\mathcal{H})}\|v\|_{S^m(\mathcal{H})}(1 + |t|^q)e^{-\eta|t|}.
\]

ii) Let \( P = \text{NAM} \) be a parabolic subgroup of \( G \). If \( \|\cdot\| \) controls the decay of matrix coefficients of \( \mathcal{H}_\tau^\infty \) in \( (\pi, \mathcal{H}) \), then there exist \( m \in \mathbb{N} \), and \( C, q \geq 0 \) such that for all \( H \in a^\tau \) and \( u, v \in \mathcal{H}_\tau^\infty \),
\[
|\langle \pi(\exp(-H))u, v \rangle| \leq C\|u\|_{S^m(\mathcal{H})}\|v\|_{S^m(\mathcal{H})}(1 + \|H\|^q)e^{-\|H\|}.
\]

**Proof.** This is fairly standard, c.f., e.g., the proofs of \[16\] Theorem 3.1 and \[33\] Theorem 6): let \( X_1, \ldots, X_d \) be an orthonormal basis for \( t \) with respect to an invariant inner product, and define \( C_K = \sum_{i=1}^{d} X_i^2 \). \( \Omega = 1 - C_K \in \mathcal{Z}(\mathbb{C}) \). As in the proof of Lemma 3 we decompose \( \mathcal{H} \) as \( \mathcal{H} = \bigoplus_{\tau \in R} \mathcal{H}_\tau \). By Schur’s lemma, \( d\pi(\Omega) \) acts as a scalar \( c(\tau) \) on each \( \mathcal{H}_\tau^\infty = \mathcal{H}_\tau \cap \mathcal{H}_\tau^\infty \). The skew-symmetry of \( d\pi(X) \) for all \( X \in \mathfrak{g} \) gives \( c(\tau) \in \mathbb{R}_{\geq 1} \). We decompose \( u \) and \( v \) as \( u = \sum_\tau u_\tau \) and \( v = \sum_\tau v_\tau \) with \( u_\tau, v_\tau \in \mathcal{H}_\tau^\infty \). Mimicking the proof of \[45\] 4.4.2.3, for \( m \in \mathbb{N} \) large enough, we have
\[
\sum_{\tau \in R} c(\tau)^{-2m(\dim \tau)^2} < \infty,
\]
where \( \dim \tau \) is the dimension of any vector space in a realization of \( \tau \). For any \( m \in \mathbb{N} \), we have
\[
\|u_\tau\| = c(\tau)^{-m}\|d\pi(\Omega^m)u_\tau\|, \quad \|v_\tau\| = c(\tau)^{-m}\|d\pi(\Omega^m)v_\tau\|.
\]

We now let \( g \) be either \( g_t \) or \( \exp(-H) \) for some \( H \in a^\tau \). Let \( D = (1 + |t|^q)e^{-\eta|t|} \) if \( g = g_t \) and \( D = (1 + \|H\|^q)e^{-\|H\|} \) if \( g = \exp(-H) \). Using either \[2\] or \[4\], we have
\[
|\langle \pi(g)u, v \rangle| \leq \sum_{\tau \in R} \sum_{\sigma \in R} |\langle \pi(g)u_\tau, v_\sigma \rangle| \leq CD \sum_{\tau \in R} \sum_{\sigma \in R} \|u_\tau\|\|v_\sigma\| (\dim(\pi(K)u_\tau) \dim(\pi(K)v_\sigma))^{1/2}.
\]
As before, by \cite[p. 206 (1)]{23}, \( \dim(\pi(K)u_\tau) \leq (\dim \tau)^2 \) and \( \dim(\pi(K)v_\sigma) \leq (\dim \sigma)^2 \). Using this, together with the Cauchy-Schwarz inequality and \cite{15}, gives
\[
\leq CD \sum_{\tau \in \hat{K}} \sum_{\sigma \in \hat{K}} \|u_\tau\| \|v_\sigma\|(\dim \tau)(\dim \sigma)
\]
\[
= CD \left( \sum_{\tau \in \hat{K}} c(\tau)^{-m} \|d\pi(\Omega^m)u_\tau\|(\dim \tau) \right) \left( \sum_{\sigma \in \hat{K}} c(\sigma)^{-m} \|d\pi(\Omega^m)v_\sigma\|(\dim \sigma) \right)
\]
\[
\leq CD \left( \sum_{\tau \in \hat{K}} \|d\pi(\Omega^m)u_\tau\|^2 \right)^{1/2} \left( \sum_{\sigma \in \hat{K}} \|d\pi(\Omega^m)v_\sigma\|^2 \right)^{1/2} \left( \sum_{\varrho \in \hat{K}} c(\varrho)^{-2m} \dim(\varrho)^2 \right)
\]
\[
= CD \|d\pi(\Omega^m)u\| \|d\pi(\Omega^m)v\| \left( \sum_{\tau \in \hat{K}} c(\tau)^{-2m}(\dim \tau)^2 \right)
\]
\[
\ll D \|u\|_{S^{2m}(\mathcal{H})} \|v\|_{S^{2m}(\mathcal{H})},
\]
as desired. \( \square \)

3. Integral Formulas

The aim of this section is to establish formulas similar to \cite[Lemma 1]{3} and \cite{8} Proposition 4] for our case of a general group \( G \).

3.1. Harish-Chandra Isomorphisms. As a starting point, we have the following lemma:

**Lemma 5.** Let \( P = NAM \) be a parabolic subgroup of \( G \), with corresponding Lie algebra \( n \oplus a \oplus m \). Given \( H \in a \), we can find elements \( Z_0, Z_1, \ldots, Z_{W-1} \) of \( \mathcal{Z}(\mathfrak{g}_C) \) such that
\[
H^W + H^{W-1}Z_{W-1} + \ldots + HZ_1 + Z_0 \in n \mathcal{U}(\mathfrak{g}_C).
\]

**Remark 5.** Although the number \( W \) depends on the choice of \( P \) and \( H \), it is uniformly bounded over all such choices, as seen in the proof.

**Proof.** Much of the proof follows by mimicking the proof of \cite[Proposition 8.22]{24}. For the sake of completeness, we write out some of the details: without loss of generality, we may assume that \( P = NAM = NFMFM \supset P_0 \) for some subset \( F \subset \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \). By the construction of standard parabolic subgroups, we then have \( a_0 \oplus m_0 \subset a \oplus m \). Choose a maximal abelian subalgebra \( \mathfrak{a}_0 \) of \( \mathfrak{m}_0 \); by \cite[Proposition 6.47]{24} \( \mathfrak{h}_C = (\mathfrak{a}_0 \oplus \mathfrak{h}_0)_C \) is a Cartan subalgebra of \( \mathfrak{g}_C \). Let \( t = t_0 \oplus a^\perp \), where \( a^\perp \) is the orthogonal complement of \( a \) in \( a_0 \) with respect to the Killing form. Note that \( t \) is a maximal abelian subalgebra of \( \mathfrak{m}_0 \), as seen from the construction of \( \mathfrak{m}_0 \) in \cite[(7.77a)]{21}. Since \( \mathfrak{h}_C \) is a Cartan subalgebra of \( \mathfrak{g}_C \), we have a root space decomposition \( \mathfrak{g}_C = \mathfrak{h}_C \oplus \bigoplus_{\lambda \in \Delta^+(\mathfrak{g}_C, \mathfrak{h}_C)} \mathfrak{g}_\lambda \). We order the roots so that \( \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) consists of the elements of \( \Delta^+(\mathfrak{g}_C, \mathfrak{h}_C) \) that are non-zero when restricted to \( a \). This ensures that \( a_C \subset \mathfrak{g}_C^\perp \), where \( \mathfrak{g}_C^\perp = \bigoplus_{\lambda \in \Delta^+(\mathfrak{g}_C, \mathfrak{h}_C)} \mathfrak{g}_\lambda \). The subalgebra \( \mathfrak{g}_C^\perp \) is defined in a similar fashion, and we get a triangular decomposition of \( \mathfrak{g}_C \): \( \mathfrak{g}_C = \mathfrak{g}_C^+ \oplus \mathfrak{h}_C \oplus \mathfrak{g}_C^\perp \). Defining the action of the Weyl group \( W = W(\mathfrak{g}_C, \mathfrak{h}_C) \) on \( \mathfrak{h}_C \) by \( \lambda(wX) = (w^{-1}\lambda)(X) \) for all \( \lambda \in \mathfrak{h}_C^\perp \), \( X \in \mathfrak{h}_C \), we consider the polynomial (with coefficients in \( \mathcal{U}(\mathfrak{h}_C) \)) given by the product
\[
(16) \quad p_H(x) = \prod_{w \in W/Stab_W(H)} (x - (\mathfrak{h}_C(H)1)) = x^W + J_{W-1}x^{W-1} + \ldots + J_1x + J_0,
\]
where \( \mathfrak{h}_C(H) = \frac{1}{2} \sum_{\lambda \in \Delta^+(\mathfrak{g}_C, \mathfrak{h}_C)} \lambda(H) = P_0(H) \). We denote the set of \( W \)-invariant elements of \( \mathcal{U}(\mathfrak{h}_C) \) by \( \mathcal{U}(\mathfrak{h}_C)^W \); since \( \mathcal{U}(\mathfrak{h}_C) \cong \text{Sym}(\mathfrak{h}_C) \), we may view \( \mathcal{U}(\mathfrak{h}_C) \) as the space of polynomial functions on \( \mathfrak{h}_C^\perp \), and the action of \( W \) is then defined by \( (wJ)(\lambda) = J(w^{-1}\lambda) \), where \( J \in \mathcal{U}(\mathfrak{h}_C) \), \( w \in W \), and \( \lambda \in \mathfrak{h}_C^\perp \). From their definition, we see that the coefficients \( J_{W-1}, \ldots, J_0 \) of \( p(x) \) are in \( \mathcal{U}(\mathfrak{h}_C)^W \), and \( p_H(H + \delta_{\mathfrak{h}_C}(H)) = 0 \). Let \( \sigma \) denote the "\( \delta_{\mathfrak{h}_C} \)-shift" on \( \mathcal{U}(\mathfrak{h}_C) \), i.e. the
unique $\mathbb{C}$-algebra automorphism given by by $\sigma(X) = X + \delta_{h_c}(X)1$ for all $X \in h_c$, and then extending to all of $\mathcal{U}(h_C)$ in the canonical way. The triangular decomposition of $g_C$ and the Poincaré-Birkhoff-Witt theorem give $\mathcal{U}(g_C) = \mathcal{U}(h_C) \oplus (g_C^+ \mathcal{U}(g_C) + \mathcal{U}(g_C) g_C^-)$. Let $\text{Proj}_{h_C} : \mathcal{U}(g_C) \to \mathcal{U}(h_C)$ be the projection onto the first summand in this decomposition, and $\gamma = \sigma \circ \text{Proj}_{h_C}$. Then $\gamma$ is the Harish-Chandra isomorphism, an isomorphism from $Z(g_C)$ to $\mathcal{U}(h_C)^W$. (Indeed, this is seen by following [24, Chapter V.5], but using $\Delta^-$ as the set of positive roots; then "$\chi_n$", "$\delta_n$", and "$\gamma_n" in [24, Chapter V.5] correspond to $\text{Proj}_{h_C}$, $-\delta_{h_c}$, and $\gamma|_{Z(g_C)}$ in our set-up, and the desired statement is given by [24, Theorem 5.44].) We now let $Z_j = \gamma|_{Z(g_C)} (J_j)$, $j = 0, 1, \ldots, W - 1$. Recall that $p_H(\sigma(H)) = 0$, i.e.

$$
\sigma(H)W + \gamma(Z_{W-1})\sigma(H)^{W-1} + \ldots + \gamma(Z_1)\sigma(H) + \gamma(Z_0) = 0.
$$

Applying $\sigma^{-1}$ to this identity gives

$$
H^W + \text{Proj}_{h_C}(Z_{W-1})H^{W-1} + \ldots + \text{Proj}_{h_C}(Z_1)H + \text{Proj}_{h_C}(Z_0) = 0.
$$

Since $Z_j - \text{Proj}_{h_C}(Z_j) \in g_C^+ \mathcal{U}(g_C)$, we now have

$$
P(H) := H^W + H^{W-1}Z_{W-1} + \ldots + HZ_1 + Z_0 \in g_C^+ \mathcal{U}(g_C).
$$

It remains to prove that this polynomial is in fact in $n \mathcal{U}(g_C)$ (which is a subspace of $g_C^+ \mathcal{U}(g_C)$). Note that $\text{Proj}_{h_C}(P(H)) = 0$. We decompose $g$ as $g = n \oplus a \oplus m \oplus n^-$, and, as noted in [23, Proof of Proposition 8.22], the same type of argument as in [24, Proof of Proposition 5.34 (b)] gives

$$
Z(g_C) \subset nC \mathcal{U}(g_C) \oplus \mathcal{U}((a \oplus m)_C),
$$

hence

$$
(17)
Z(g_C) \mathcal{U}(a_C) \subset nC \mathcal{U}(g_C) \oplus \mathcal{U}((a \oplus m)_C).
$$

Note that $a \oplus m$ is a reductive Lie algebra with Cartan subalgebra $a \oplus t$. We thus have a triangular decomposition $(a \oplus m)_C = m_C^- \oplus (a \oplus t)_C \oplus m_C^+$, where the order on $\Delta(m_C, t_C)$ that defines $m_C^+$ is chosen so that $\Delta^+(m_C, t_C) = \Delta^+(g_C, h_C)|_{C}\setminus \{0\}$. This choice of positivity gives $g_C^- = n_C \oplus m_C^+$. Letting $\text{Proj}_{(a\oplus m)_C} : Z(g_C) \mathcal{U}(a_C) \to \mathcal{U}((a \oplus m)_C)$ be the projection onto the second summand of (17), one checks that for all $Z \in Z(g_C)$, $\text{Proj}_{(a\oplus m)_C}(Z) \in Z((a \oplus m)_C)$, hence $\text{Proj}_{(a\oplus m)_C}(Z(g_C) \mathcal{U}(a_C)) \subset Z((a \oplus m)_C)$. Since $a \oplus t$ is a Cartan subalgebra of the reductive Lie algebra, $a \oplus m$, by the same arguments as for $g_C$, $Z((a \oplus m)_C) \subset \mathcal{U}((a \oplus t)_C) \oplus m_C^+ \mathcal{U}((a \oplus t)_C)$. Letting $\text{Proj}((a\oplus t)_C)$ denote the projection from $Z((a \oplus m)_C)$ onto the first summand, we get

$$
\text{Proj}_{h_C}|_{Z(g_C) \mathcal{U}(a_C)} = \text{Proj}_{(a\oplus t)_C} \circ \text{Proj}_{(a\oplus m)_C}
$$

(cf. [23, p. 225 (8.34)]). Since composition with an appropriate half-sum of positive roots turns $\text{Proj}_{(a\oplus t)_C}$ into a Harish-Chandra isomorphism for $Z((a \oplus m)_C)$, $\text{Proj}_{(a\oplus t)_C}$ is invertible. In particular, since $\text{Proj}_{(a\oplus m)_C}( \text{Proj}_{(a\oplus m)_C}(P(H))) = \text{Proj}_{h_C}(P(H)) = 0$, we conclude that $\text{Proj}_{(a\oplus m)_C}(P(H)) = 0$, i.e. $P(H) \in nC \mathcal{U}(g_C) = n \mathcal{U}(g_C)$. \hfill $\square$

### 3.2. Differential Equations

We shall now use Lemma 4 to express the "averaging operator" in terms of differential operators. Let $(\pi, \mathcal{H})$ be an irreducible unitary representation of $G$, and assume that $v \in \mathcal{H}^\infty$. For $\chi \in C_c^\infty(U^+)$ define $I^\chi_v : G \to \mathcal{H}$ by

$$
I^\chi_v(g) = \int_{U^+} \chi(u)\pi(ug)v\, du.
$$

As we shall see later, the restriction that $\chi \in C_c^\infty(U^+)$ may be replaced by the condition that $\chi \in C_c^m(U^+)$ for some fixed number $m$, but for convenience it is assumed throughout
the remainder of Section 3 that \( \chi \) is smooth. Note that \( \mathcal{I}^\chi_{v} \) is linear in both \( v \) and \( \chi \), and for \( X \in \mathfrak{g} \):

\[
(\mathcal{I}^\chi_{v})(g) = \frac{d}{ds}_{s=0} \int_{U^+} \chi(u) \pi(u \exp(sX)) v \, du
\]

\[
= \int_{U^+} \chi(u) \pi(u) \frac{d}{ds}_{s=0} \pi(\exp(sX)) v \, du = \mathcal{I}^\chi_{\pi(U^+)v}(g),
\]

the exchange of integration and differentiation being permitted since \( \chi \) has compact support. We use linearity and composition to extend this to all of \( \mathcal{U}(\mathfrak{g}_C) \), i.e.

\[
(\mathcal{U} \mathcal{I}^\chi_{v})(g) = \mathcal{I}^\chi_{\pi(U)\pi(U^+)v}(g) \quad \forall U \in \mathcal{U}(\mathfrak{g}_C).
\]

Similarly, \( u^+ \) acts on \( C^\infty_c(U^+) \) by

\[
(X \psi)(u) := \frac{d}{ds}_{s=0} \psi(u \exp(sX)) \quad \forall X \in u^+, \, u \in U^+, \, \psi \in C^\infty_c(U^+).\]

By linearity and composition, this defines \( X \psi \) for all \( X \in \mathcal{U}(u^+_C) \). The next lemma relates the action of \( \mathcal{U}(u^+_C) \) on \( \mathcal{I}^\chi_{v}(g) \) and \( \chi \) for certain special \( g \in G \):

**Lemma 6.** Let \( X \in u^+ \) and \( g \in G \) be such that \( \text{Ad}_g X \in u^+ \). Then for all \( v \in \mathcal{H}^\infty \) and \( \chi \in C^\infty_c(U^+) \),

\[
(X \mathcal{I}^\chi_{v})(g) = -\mathcal{I}^v_{\text{Ad}_g X \chi}(g).
\]

**Proof.**

\[
(X \mathcal{I}^\chi_{v})(g) = \frac{d}{ds}_{s=0} \int_{U^+} \chi(u) \pi(u \exp(sX)) v \, du
\]

\[
= \frac{d}{ds}_{s=0} \int_{U^+} \chi(u) \pi(u \exp(s\text{Ad}_g X)g) v \, du
\]

\[
= \int_{U^+} \chi(u \exp(-s\text{Ad}_g X)) \pi(ug)v \, du = \mathcal{I}^v_{\text{Ad}_g X \chi}(g).
\]

We now let \( P = NAM \) be the *unique* parabolic subgroup such that \( Y \in \mathfrak{a}^+ \) and \( U^+ = N \). This allows us to fix a root basis \( X_1, \ldots, X_d \) of \( u^+ = \mathfrak{n} \); for each \( j = 1, \ldots, d \) there is some \( \alpha_j > 0 \) such that \( \text{Ad}_{g_j} X_j = e^{\alpha_j t} X_j \) for all \( t \in \mathbb{R} \). Since \( \log(g_1) = Y \in \mathfrak{a} \), we apply Lemma 5 to find \( W \in N_{>0}, \, Z_0, \ldots, Z_{W-1} \in \mathcal{Z}(\mathfrak{g}_C) \), and \( U_1, \ldots, U_d \in \mathcal{U}(\mathfrak{g}_C) \) such that

\[
Y^W + Y^{W-1}Z_{W-1} + \ldots + YZ_1 + Z_0 = \sum_{j=1}^{d} X_j U_j
\]

(as is clear from the proof of Lemma 5 if \( k_0 \in K \) is such that \( \text{Ad}_{k_0} a = a_F \), for some \( F \subset \Sigma^+_0(\mathfrak{g}, a_0) \), then \( W = |W/\text{Stab}_W(\text{Ad}_{k_0} Y)| \), \( W \) being the Weyl group defined with respect to \( (a_0 \oplus k_0)_C \)). Combining (13), (20) and Lemma 6 gives

\[
(Y^W \mathcal{I}^\chi_{v})(g_t) + \sum_{i=0}^{W-1} (Y^i \mathcal{I}^\chi_{\pi(U^+)_i v})(g_t) = -\sum_{j=1}^{d} e^{\alpha_j t} \mathcal{I}^X_{\mathcal{I}^\chi_{\pi(U^+_j)_v}}(g_t).
\]

By Schur’s lemma, each \( d\pi(Z_i) \) acts as a scalar \( a_i \in \mathbb{C} \). This fact and calculations analogous to those in (18) give

\[
(Y^i \mathcal{I}^\chi_{\pi(U^+_i)_v})(g_t) = a_i \frac{d^i}{dt^i} \mathcal{I}^\chi_{v}(g_t), \quad a_i \in \mathbb{C}.
\]
We let $\lambda_1, \ldots, \lambda_W$ be the roots of the polynomial $z^W + a_{W-1}z^{W-1} + \ldots + a_1z + a_0$. Abusing notation slightly by letting $T^\chi_\psi(t) = T^\chi_\psi(g_t)$, (24) may be rewritten as

$$
\left( \prod_{i=1}^W \left( \frac{d}{dt} - \lambda_i \right) \right) T^\chi_\psi(t) = - \sum_{j=1}^d e^{\alpha_j t} T^{X_j^\chi}_{ds(\eta_j)}(t).
$$

Now, each $T^{X_j^\chi}_{ds(\eta_j)}(t)$ will solve the same type of equation, i.e.

$$
\left( \prod_{i=1}^W \left( \frac{d}{dt} - \lambda_i \right) \right) T^{X_j^\chi}_{ds(\eta_j)}(t) = - \sum_{k=1}^d e^{\alpha_k t} T^{X_k^\chi X_k^\chi}_{ds(\eta_k)}(t).
$$

This property will be important in obtaining a better rate than $\min_j \alpha_j$ in Theorem I.

3.3. Burger’s Formula. Here we present “integral formulas” for the solution of (22). From now on assume that $(\pi, H)$ is an irreducible unitary representation with $\eta$ as a rate of decay for the matrix coefficients of $g_\mathbb{R}$ in $(\pi, H)$; in particular (cf. Lemma 2), we assume that there exists $m \in \mathbb{N}$ such that

$$
|\langle \pi(g_t)u, v \rangle| \leq C_{\psi, \gamma} e^{\gamma t}
$$

for all $\epsilon > 0$, $t \leq 0$, and $u, v \in H^\infty$. Throughout this section we assume that we have the differential equation

$$
\left( \prod_{i=1}^W \left( \frac{d}{dt} - \lambda_i \right) \right) T^\chi_\psi(t) = \psi(t).
$$

We let $\Lambda$ be the multi-set of order $W$ consisting of the complex numbers $\lambda_i$, $i = 1, \ldots, W$. Since we will require several different orderings of the $\lambda_i$s, we consider the following construction: let $\text{Sym}_W$ denote the group of permutations of the set $\{1, 2, \ldots, W\}$. We may identify $\mathbb{C}^W/\text{Sym}_W$ with the family of multi-sets of complex numbers having order $W$. The element $\Lambda$ may thus be viewed as an element of $\mathbb{C}^W/\text{Sym}_W$.

**Lemma 7.** Assume that $\|\psi(t)\| \leq C_{\psi, \gamma} e^{\gamma t}$ for some positive number $\gamma$ and all $t \leq 0$, and choose a number $\beta_\gamma$ so that $i) 0 < \beta_\gamma < \gamma$ ii) $\beta_\gamma \leq \eta$. Label and order the $\lambda_i$s in the following way:

$$
\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_W) = (\lambda_1^-, \lambda_2^-, \ldots, \lambda_{m_1}^-, \lambda_{m_1}^+, \lambda_{m_1}^+, \ldots, \lambda_{m_2}^+, \ldots, \lambda_{m_2}^+),
$$

where $\beta_\gamma > \text{Re}(\lambda_1^-) \geq \text{Re}(\lambda_2^-) \geq \ldots \geq \text{Re}(\lambda_{m_1}^-)$, and $\text{Re}(\lambda_1^+) \geq \text{Re}(\lambda_2^+) \geq \ldots \geq \text{Re}(\lambda_{m_2}^+) \geq \beta_\gamma$. If $m_1 > 0$, then for all $t \in \mathbb{R}$ we have

$$
\left( \prod_{i=1}^{m_2} \left( \frac{d}{dt} - \lambda_j^+ \right) \right) T^\chi_\psi(t) = e^{\lambda_{m_1}^- t} \int_{-\infty}^t e^{(\lambda_{m_1}^- - \lambda_{m_1}^-) t_{m_1}} \int_{-\infty}^{t_{m_1}} e^{(\lambda_{m_1}^- - \lambda_{m_1}^-) t_{m_1-1}} \ldots \int_{-\infty}^{t_{m_2}} e^{(\lambda_{m_2}^- - \lambda_{m_2}^-) t_{m_2}} e^{-\lambda^- t} \psi(t) dt_1 dt_2 \ldots dt_{m_1-1} dt_{m_1}
$$

(where the product in the left-hand side is viewed as empty if $m_2 = 0$).

**Proof.** The proof follows that of [8, Lemma 3]: we set $\lambda_0^- = 0$, and for $i = 0, 1, \ldots, m_1$, let $\psi_i(t)$ be such that

$$
e^{\lambda_i^- t} \psi_i(t) = \left( \prod_{j=i+1}^{m_1} \left( \frac{d}{dt} - \lambda_j^- \right) \right) \left( \prod_{j=1}^{m_2} \left( \frac{d}{dt} - \lambda_j^+ \right) \right) T^\chi_\psi(t).
$$

In particular,

$$
e^{\lambda_{m_1}^- t} \psi_{m_1}(t) = \left( \prod_{j=1}^{m_2} \left( \frac{d}{dt} - \lambda_j^+ \right) \right) T^\chi_\psi(t).
$$
Note that for $i \geq 1$,
\[ \psi_i'(t) = e^{(\lambda_i^- - \lambda_i^+)_t} \psi_{i-1}(t). \]

By the fundamental theorem of calculus,
\[ \psi_i(t) - \psi_i(r) = \int_r^t e^{(\lambda_i^- - \lambda_i^+)_s} \psi_{i-1}(s) \, ds. \]

Now let $i \in \{1, \ldots, m_1\}$, and assume we have the bound
\[ \|\psi_{i-1}(s)\| \leq D_{i-1} e^{(\gamma - \Re(\lambda_i^-))s} \quad \text{for all } s \leq 0. \]

Then for all $r \leq t \leq 0$ we have
\[ \|\psi_i(t) - \psi_i(r)\| \leq \int_r^t e^{(\Re(\lambda_i^-) - \Re(\lambda_i^+))}_s D_{i-1} e^{(\gamma - \Re(\lambda_i^-))s} \, ds \]
\[ \ll e^{(\gamma - \Re(\lambda_i^-))t - (\gamma - \Re(\lambda_i^-))r}. \]

It follows that $\psi_i(r)$ converges to some vector $\mathbf{v}_{i, \infty}$ as $r \to -\infty$, and
\[ \psi_i(t) = \mathbf{v}_{i, \infty} + \int_{-\infty}^t e^{(\lambda_i^- - \lambda_i^+)_s} \psi_{i-1}(s) \, ds. \]

We now wish to prove that $\mathbf{v}_{i, \infty} = 0$. Let $\mathbf{w}_i = \left( \prod_{j=i+1}^W (d\pi_j - \lambda_j) \right) \mathbf{v}$, and note that from the definition of $\psi_i(r)$,
\[ \psi_i(r) = e^{-\lambda_i^- r} T_{\mathbf{w}_i}(r), \]

hence
\[ \mathbf{v}_{i, \infty} = \lim_{r \to -\infty} e^{-\lambda_i^- r} T_{\mathbf{w}_i}(r). \]

We now use the rate of decay of matrix coefficients to prove that $\langle \mathbf{v}_{i, \infty}, \mathbf{u} \rangle = 0$ for all $\mathbf{u} \in \mathcal{H}^\infty$. Since $\mathcal{H}^\infty$ is dense in $\mathcal{H}$, we may then conclude that $\mathbf{v}_{i, \infty} = 0$. We have
\[ |\langle \mathbf{v}_{i, \infty}, \mathbf{u} \rangle| = \lim_{r \to -\infty} |\langle e^{-\lambda_i^- r} T_{\mathbf{w}_i}(r), \mathbf{u} \rangle| \leq \liminf_{r \to -\infty} e^{-\Re(\lambda_i^-)r} \int_{U^+} |\chi(u)||\pi(g_i)\mathbf{w}_i, \mathbf{u}|| \, du \]
\[ = \liminf_{r \to -\infty} e^{-\Re(\lambda_i^-)r} \int_{U^+} |\chi(u)||\pi(g_i)\mathbf{w}_i, \pi(u^{-1})\mathbf{u}|| \, du \]
\[ \ll e \liminf_{r \to -\infty} \|\chi\|_{L^\infty(N)} \|\mathbf{w}_i\|_{S^m(\mathcal{H})} e^{(\eta - \Re(\lambda_i^-))r} \int_{\text{supp}(\chi)} \|\pi(u^{-1})\mathbf{u}\|_{S^m(\mathcal{H})} \, du \]

where (23) was used in the last inequality. Choosing $\epsilon$ so that $\eta - \epsilon - \Re(\lambda_i^-) > 0$, and noting that by (11), the integral over $\text{supp}(\chi)$ is finite, we conclude that $\langle \mathbf{v}_{i, \infty}, \mathbf{u} \rangle = 0$. We have thus proved that under the assumption (23), we have
\[ \psi_i(t) = \int_{-\infty}^t e^{(\lambda_i^- - \lambda_i^+)_s} \psi_{i-1}(s) \, ds, \]

and
\[ \|\psi_i(t)\| \leq \int_{-\infty}^t e^{(\Re(\lambda_i^-) - \Re(\lambda_i^+))}_s D_{i-1} e^{(\gamma - \Re(\lambda_i^-))s} \, ds \]
\[ = \frac{D_{i-1} e^{(\gamma - \Re(\lambda_i^-))s}}{\gamma - \Re(\lambda_i^-)}. \]

Since $\|\psi_0(t)\| = \|\psi(t)\| \leq C_{\gamma, \gamma} e^{\gamma t}$, induction establishes that (27) and (26) hold for all $i$, $1 \leq i \leq m_1$. Using this fact together with $\psi_0(t) = \psi(t)$ and (25), we conclude that (24) holds. \qed

Our first version of “Burger’s formula” (cf. [3] Lemma 1 and [8] Proposition 4) can now be stated. Let $|\lambda|_\infty = \max_{1 \leq i \leq W} |\lambda_i|$. We then have the following:
Proposition 8. There exist (completely explicit) functions

\[ F : (\mathbb{C}^W/\text{Sym}_W) \times \mathbb{R}_{>0} \times \mathbb{R}_{\leq 0} \times \mathbb{R}_{\leq 0} \to \mathbb{C} \]

and

\[ F_0, F_1, \ldots, F_W : (\mathbb{C}^W/\text{Sym}_W) \times \mathbb{R}_{>0} \times \mathbb{R}_{\leq 0} \to \mathbb{C}, \]

such that the following hold:

i) Whenever \( I_0(\ell,t) \), \( \psi(t) \) and \( \beta_\gamma \) satisfy the assumptions of Lemma 7, we have for all \( t \leq 0 \):

\[ I_0(\ell,t) = \int_{-\infty}^{0} F(\Delta, \beta_\gamma, t, s) \psi(s) \, ds + \sum_{i=0}^{W} F_i(\Delta, \beta_\gamma, t) I_0^\chi(\gamma_i t, 0). \]

ii) \[ |F(\Delta, \beta_\gamma, t, s)| \ll |t|^{W-1-m_0} |s|^{m_0} e^{\beta_\gamma (t-s)}. \]

iii) \[ |F_i(\Delta, \beta_\gamma, t)| \ll (1 + |\Delta|_\infty) (1 + |t|^W) e^{\beta_\gamma t}. \]

where \( m_0 = \max\{0, m_1 - 1\} \) (\( m_1 \) being the same as in Lemma 7). The implied constants depend only on \( W \).

Proof. We use the same notation as in Lemma 7, as well as letting \( \Psi(t) \) denote the left-hand side of (24). We also let \( \lambda_-^+ = 0 \), and define \( \Psi_i(t) \) successively for \( i = m_2, m_2 - 1, \ldots, 0 \) through

\[ \tag{29} I_0^\chi(t) = e^{\lambda_-^+ t} \Psi_{m_2}(t), \]

\[ \frac{d}{dt} \Psi_{i+1}(t) = e^{(\lambda_-^+ - \lambda_-^+ i)} \Psi_i(t). \]

From these definitions, we obtain

\[ e^{\lambda_-^+ t} \Psi_i(t) = \left( \prod_{j=i+1}^{m_2} \left( \frac{d}{dt} - \lambda_-^j \right) \right) I_0^\chi(t), \]

with (as usual) the product being viewed as empty if \( i = m_2 \). In particular, \( \Psi_0(t) = \psi(t) \) if \( m_1 = 0 \), and \( \Psi_0(t) = \Psi(t) \) otherwise. Making the definition

\[ \psi_i = \left( \prod_{j=i+1}^{m_2} (d\pi(Y) - \lambda_-^j) \right) \psi \]

(so \( \psi_{m_2} = \psi \)) gives

\[ \Psi_i(t) = I_0^\chi(t). \]

There are thus polynomials \( P_{i,j}(\Delta) \) (in fact, elementary symmetric polynomials in the \( \lambda_-^j s \)) such that

\[ \Psi_i(0) = \sum_{j=0}^{m_2} P_{i,j}(\Delta) I_0^\chi(Y_j W(0)). \]

Assuming \( m_1 > 0 \), Lemma 7 gives

\[ \Psi_0(t) = \int_{-\infty}^{t} \ldots \int_{-\infty}^{r_1} \int_{-\infty}^{r_2} \psi(r_1) \exp \left( \lambda_-^1 t - \lambda_- r_1 + \sum_{i=2}^{m_1} (\lambda_-^i - \lambda_-^1) r_i \right) \, dr_1 \, dr_2 \ldots \, dr_{m_1}. \]

Reversing the order of the integration gives

\[ \Psi_0(t) = \int_{-\infty}^{0} \psi(s) \Phi_0(t, s) \, ds, \]

where

\[ \Phi_0(t, s) = \begin{cases} 0 & \text{if } s > t \\ s \int_{s}^{t} \ldots \int_{s}^{r_1} \exp(\lambda_-^1 t - \lambda_- s + \sum_{i=2}^{m_1} (\lambda_-^i - \lambda_-^1) r_i) \, dr_{m_1} \ldots \, dr_2 & \text{if } s \leq t. \end{cases} \]
Repeatedly using $\text{Re}(\lambda_i^{-1}) \geq \text{Re}(\lambda_i^{-})$, and finally $\beta_i > \text{Re}(\lambda_i^{-})$, gives

$$|\Phi_0(t, s)| \leq \begin{cases} 0 & \text{if } s > t \\ (t-s)^{m_1-1} \exp(\beta_i(t-s)) & \text{if } s \leq t. \end{cases}$$

We now write

$$\Psi_i(t) = \int_{-\infty}^0 \psi(s)\Phi_i(t, s)\, ds + \sum_{j=0}^{m_2} \Pi_{i,j}(t, \lambda)\mathcal{I}_{d\pi(Y_j)\nu}(0),$$

where $\Pi_{0,j}(t, \lambda) \equiv 0$, and the other $\Pi_{i,j}$ and $\Phi_i$ are defined iteratively using (29) (if $m_2 > 0$) and the fundamental theorem of calculus, i.e. for $i = 0, 1, \ldots, m_2 - 1$,

$$\Psi_{i+1}(t) = \Psi_{i+1}(0) - \int_t^0 e^{(\lambda_i^+ - \lambda_{i+1}^+)r}\Psi_i(r)\, dr$$

$$= \left(\sum_{j=0}^{m_2} \mathcal{P}_{i+1,j}(\lambda)\mathcal{I}_{d\pi(Y_j)\nu}(0)\right)$$

$$- \int_t^0 e^{(\lambda_i^+ - \lambda_{i+1}^+)r} \left(\int_{-\infty}^0 \psi(s)\Phi_i(r, s)\, ds + \sum_{j=0}^{m_2} \Pi_{i,j}(r, \lambda)\mathcal{I}_{d\pi(Y_j)\nu}(0)\right)\, dr.$$

This yields

$$\Phi_{i+1}(t, s) = -\int_t^0 e^{(\lambda_i^+ - \lambda_{i+1}^+)r}\Phi_i(r, s)\, dr,$$

and

$$\Pi_{i+1,j}(t, \lambda) = \mathcal{P}_{i+1,j}(\lambda) - \int_t^0 e^{(\lambda_i^+ - \lambda_{i+1}^+)r}\Pi_{i,j}(r, \lambda)\, dr.$$

Note that all the $\mathcal{P}_{i,j}$ have degrees less than or equal to $m_2$; there is thus an absolute constant $D$ such that for all $i, j$, $|\mathcal{P}_{i,j}(\lambda)| \leq D(1 + |\lambda|^{m_2})$. By induction, we have

$$|\Pi_{i,j}(t, \lambda)| \ll (1 + |\lambda|^{m_2})(1 + |t|^i).$$

In a similar fashion,

$$|\Phi_i(t, s)| \leq |t|^i|s|^{m_1-1}e^{\beta_i(t-s)}e^{-\text{Re}(\lambda_i^+)t},$$

where the bound for $i = 0$ holds by (29) (recall that $\lambda_0^+ = 0$), and for $i = 1, 2, \ldots, m_2$ by induction. Since $\mathcal{I}_\phi(t) = e^{\lambda_{m_2}^+}t\Psi_{m_2}(t)$, these bounds establish $i), ii), \text{ and } iii)$ when $m_1 > 0$.

It remains to consider the case $m_1 = 0$. We then have $\Psi_0(t) = \psi(t)$, hence

$$\Psi_1(t) = \Psi_1(0) - \int_t^0 e^{-\lambda_1^+s}\psi(s)\, ds$$

$$= \left(\sum_{j=0}^{m_2} \mathcal{P}_{1,j}(\lambda)\mathcal{I}_{d\pi(Y_j)\nu}(0)\right) + \int_{-\infty}^0 \phi_1(t, s)\psi(s)\, ds,$$

where

$$\phi_1(t, s) = \begin{cases} 0 & \text{if } s > t \\ -e^{-\lambda_1^+s} & \text{if } s \leq t. \end{cases}$$

We now proceed as in the case $m_1 > 0$: making repeated use of the fundamental theorem of calculus, we write, for $i = 1, 2, \ldots, m_2$,

$$\Psi_i(t) = \int_{-\infty}^0 \psi(s)\phi_i(t, s)\, ds + \sum_{j=0}^{m_2} \Pi_{i,j}(t, \lambda)\mathcal{I}_{d\pi(Y_j)\nu}(0),$$
where \( \Pi_{i,j}(t, \Delta) = \mathcal{P}_{i,j}(\Delta) \), and \( \Pi_{i,j}(t, \Delta) \) and \( \phi_i(t, s) \) for \( i = 2, 3, \ldots, m_2 \) are given recursively by (31) and
\[
\phi_{i+1}(t, s) = - \int_t^s e^{(\lambda_i^+ - \lambda_{i+1}^+)r} \phi_i(r, s) \, dr,
\]
respectively. Induction now gives
\[
|\phi_i(t, s)| \leq \begin{cases} 
0 & \text{if } s < t \\
(s - t)^{i-1} e^{-\text{Re}(\lambda_i^+)s} & \text{if } s \geq t.
\end{cases}
\]
Observing that \( I_\psi(t) = e^{\lambda_{m_2}^+ t} \psi_{m_2}(t) \), as well as the fact that the bound (32) still holds, concludes the proof.

We now use Proposition 8 to demonstrate how the bound on \( \psi \) can be “lifted” to \( I_\psi \):

**Corollary 9.** Suppose that \( I_\psi(t), \psi(t) \) and \( \beta_\gamma \) satisfy the assumptions of Lemma 7. There then exists \( n_0 \in \mathbb{N} \) such that for all \( t \leq 0, \epsilon > 0, \)
\[
||I_\psi(t)|| \leq (\|\chi\|_{L^1(U^+)} \|\psi\|_{S^{n_0}(\mathcal{H})} + C_{\psi, \gamma}) e^{(\beta_\gamma - \epsilon)t},
\]
where the implied constant depends on \( G, g_\mathcal{R}, \epsilon, \) and \( \gamma - \beta_\gamma \).

**Proof.** By Proposition 8,
\[
||I_\psi(t)|| \leq \int_{-\infty}^0 |F(\lambda, \beta_\gamma, t, s)||\psi(s)||ds + \sum_{i=0}^W |F_i(\lambda, \beta_\gamma, t)||I_{\psi}(Y^{\gamma})v(0)||.
\]
Using \( ||\psi(s)|| \leq C_{\psi, \gamma} e^{\gamma s} \), the bound on \( F_i(\lambda, \beta_\gamma, t) \) from Proposition 8 (iii), and
\[
||I_{\psi}(Y^{\gamma})v(0)|| \leq \int_{U^+} \|\chi(u)||\pi(u)ds\|du \ll \|\chi\|_{L^1(U^+)} \|\psi\|_{S^1(H)}
\]
gives
\[
||I_\psi(t)|| \leq (1 + |\lambda|_\infty W)(1 + |t| W) e^{\beta_\gamma t} \|\chi\|_{L^1(U^+)} \|\psi\|_{S^0(\mathcal{H})} + C_{\psi, \gamma} \int_{-\infty}^0 |F(\lambda, \beta_\gamma, t, s)| e^{\gamma s} ds.
\]
Now, by Proposition 8 (ii),
\[
\int_{-\infty}^0 |F(\lambda, \beta_\gamma, t, s)| e^{\gamma s} ds \ll |t|^{W-1-m_0} \int_{-\infty}^0 |s|^{m_0} e^{(\beta_\gamma - \epsilon)s} e^{\gamma s} ds
\ll (1 + |t| W) e^{\beta_\gamma t}.
\]
We now have
\[
||I_\psi(t)|| \leq (\|\chi\|_{L^1(U^+)} (1 + |\lambda|_\infty W)) \|\psi\|_{S^0(\mathcal{H})} + C_{\psi, \gamma}) e^{(\beta_\gamma - \epsilon)t}.
\]
The definition of the \( \lambda_i \)’s is now used to bound \( |\lambda|_\infty \): recall that the \( \lambda_i \)’s are the solutions to the polynomial equation
\[
z^W + d\pi(Z_{W-1})z^{W-1} + \ldots + d\pi(Z_1)z + d\pi(Z_0) = 0,
\]
so by Cauchy’s bound, \( |\lambda|_\infty \leq 1 + \max_i |d\pi(Z_i)|. \) This gives
\[
|\lambda|_\infty W \|\psi\|_{S^0(\mathcal{H})} \leq \|\psi\|_{S^0(\mathcal{H})} + \max_i |d\pi(Z_i)| \|\psi\|_{S^0(\mathcal{H})}
\]
\[
= \|\psi\|_{S^0(\mathcal{H})} + \max_i |d\pi(Z_i)| \|\psi\|_{S^0(\mathcal{H})} \ll \|\psi\|_{S^{n_0}(\mathcal{H})}
\]
for some \( n_0 \in \mathbb{N} \).
We will now make use of the smoothness of $\chi$ to repeatedly apply Proposition 8 and Corollary 9 to obtain the final version of “Burger’s formula” that will be used in the proof of Theorem 11. In preparation of the proof, we first introduce some more notation.

Recall that we have fixed a basis $X_1, \ldots, X_d$ of $\mathbb{R}^d$, and elements $U_1, \ldots, U_d$ in $\mathcal{U}(G)$. We introduce the following multi-index notation: let $j = (j_1, j_2, \ldots, j_l) \in \bigcup_{l=1}^{\infty} \{1, \ldots, d\}^l$, and define $|j| = l$ (we call $|j|$ the order of $j$). For a multi-index $j$, we define $X_j = X_{j_1}X_{j_2} \cdots X_{j_l}$, and $U_j = U_{j_1}U_{j_2} \cdots U_{j_l}$. We also let $I_k$ and $I_{\leq k}$ denote the sets of multi-indices of order $k$, and order less than or equal to $k$, respectively. We also assume that $I_{\leq k}$ includes $0$, the set of multi-indices of order zero (i.e. if $k \in I_0$, then $dt(U_k)v = v$ and $X_k\chi = \chi$). Note that $I_0$ is in fact a singleton set, consisting only of the “empty string”, which we denote $0$. We also let $I = \bigcup_{k=0}^{\infty} I_k$ be the set of all multi-indices.

This multi-index notation can be used to explicitly express the Sobolev norm $\| \cdot \|_{W^k,p(U^+)}$; for $\chi \in C_c^\infty(U^+)$,

$$\|\chi\|_{W^k,p(U^+)} = \sum_{|\alpha| \leq k} \|X_\alpha \chi\|_{L^p(U^+)}.$$ 

We let $n_0$ be such that for all $j$, $U_j \in \mathcal{U}^{\infty}(G)$ (without loss of generality, we may take $n_0$ to be the same integer as in Corollary 9). From (22), we have that for any multi-index $k$,

$$(35) \quad \left( \prod_{i=1}^d (\frac{d}{\pi} - \lambda_i) \right) T^{X_k\chi}_{d\pi}(U_k)v(t) = -\sum_{\alpha} e^{\alpha t} X_\alpha^{X_k\chi}(U_k)v(t),$$

and, moreover, the following bound holds:

$$(36) \quad \left\| \sum_{\alpha} e^{\alpha t} X_\alpha^{X_k\chi}(U_k)v(t) \right\|_{L^p(U^+)} \leq \|\chi\|_{W^{k+1,1}(U^+)} \|v\|_{S^{k+n_0+1}(H)} e^{(\min_j \alpha_j)t}.$$ 

We may now state the main result of this section:

**Proposition 10.** Let $\alpha \in (0, \min_j \alpha_j)$, and define

$$k_0 = \begin{cases} 1 & \text{if } \alpha \geq 2\eta \\ 2 & \text{if } 2\eta < \alpha < 2\eta \\ \left\lfloor \frac{2\eta}{\alpha} \right\rfloor & \text{if } \alpha \leq \eta. \end{cases}$$

There then exist, for given $(\pi, H)$ and $g_\mathbb{R}$, functions $C_k : \mathbb{R}_{\leq 0} \times \mathbb{R}_{\leq 0} \to \mathbb{C}$ ($k \in I_{k_0}$) and $D_{j,s} : \mathbb{R}_{\leq 0} \to \mathbb{C}$ ($j \in I_{k_0-1}$ and $i \in \{0, \ldots, W\}$) such that for all $v \in H^k$, $\chi \in C_c^\infty(U^+)$, and $t \leq 0$:

$$(37) \quad T^{X}_v(t) = \sum_{k \in I_{k_0}} \int_{-\infty}^0 C_k(t,s) T^{X_k\chi}_{d\pi(U_k)}v(s) \, ds + \sum_{j \in I_{k_0-1}} \sum_{i=0}^W D_{j,i}(t) T^{X_j\chi}_{d\pi(U_j)}v(0).$$

Furthermore, the following hold:

1. $|C_k(t,s)| \ll (1 + |t|^W) e^{\gamma t} s^{\beta_j}$, where the implied constant depends only on $G$, $\eta$, and $\alpha$.
2. There exists $n_0 \in \mathbb{N}$, depending only on $G$, such that for any $m \in \mathbb{N}$,

$$|D_{j,i}(t)| \ll (1 + |t|^W) e^{\gamma t} \inf_{\omega \in H^\infty \setminus \{0\}} \frac{\|\omega\|_{S^{n+m}(H)}}{\|\omega\|_{S^m(H)}},$$

where the implied constant depends only on $G$, $g_\mathbb{R}$, $\eta$, $\alpha$, and $m$.

**Proof.** By (22), and the fact that $\left\| - \sum_{j=1}^d e^{\alpha_j t} X_\alpha^{X_j\chi}(U_k)v(t) \right\| \leq C e^{\gamma t}$ for some $C > 0$ and all $t \leq 0$, if $\alpha \geq 2\eta$, we may apply Proposition 8 with $\psi(t) = \sum_{j=1}^d -e^{\alpha_j t} X_\alpha^{X_j\chi}(U_k)v(t)$, $\gamma = \alpha$, and $\beta_j = \eta$. The bounds in Proposition 8 (ii) and (iii), and (34) then suffice to prove the proposition; in particular, (34) provides the bound $\left(1 + \frac{|W|^W}{\lambda^W}\right)\|v\|_{S^{n}(H)} \ll \|v\|_{S^{n+m}(H)}$ (for all $m \in \mathbb{N}$ and $v \in H^\infty$), which, together with Proposition 8 (iii), proves (2).
If \( \eta < \alpha < 2\eta \), we will apply Proposition 8 twice: by (35) and (36), Proposition 8 may be applied to \( I_{\alpha}^{X^jY^k}(v)(t) \) (for each \( j = 1, \ldots, d \)) with \( \psi(t) = -\sum_{k=1}^{d} e^{\alpha_k t} I_{\alpha}^{X^jY^k}(v)(t) \), \( \gamma = \alpha \), and \( \beta_\gamma = \eta - \frac{\alpha}{4} \). This gives

\[
I_{\alpha}^{X^jY^k}(v)(t) = \int_{-\infty}^{0} F(\Delta, \eta - \frac{\alpha}{4}, t, s) \left( -\sum_{k=1}^{d} e^{\alpha_k s} I_{\alpha}^{X^jY^k}(v)(s) \right) ds + \sum_{i=0}^{W} F_i(\Delta, \eta - \frac{\alpha}{4}, t) I_{\alpha}^{X^jY^k}(v)(0).
\]

By letting \( \epsilon = \frac{\alpha}{4} \) in Corollary 9, \( \| I_{\alpha}^{X^jY^k}(v)(t) \| \ll e^{(\eta - \frac{\alpha}{4})t} \) for all \( t \leq 0 \) and \( j = 1, \ldots, d \). This now gives the bound \( \| -\sum_{j=1}^{d} e^{\alpha_j t} I_{\alpha}^{X^jY^k}(v)(t) \| \ll e^{(\eta + \frac{\alpha}{2})t} \), which enables us to apply Proposition 8 to \( I_{\alpha}^X(t) \) with \( \psi(t) = -\sum_{j=1}^{d} e^{\alpha_j t} I_{\alpha}^{X^jY^k}(v)(t) \), \( \gamma = \eta + \frac{\alpha}{4} \), and \( \beta_\gamma = \eta \), hence

\[
I_{\alpha}^X(t) = \int_{-\infty}^{0} F(\Delta, \eta, t, s) \left( -\sum_{j=1}^{d} e^{\alpha_j s} I_{\alpha}^{X^jY^k}(v)(s) \right) ds + \sum_{i=0}^{W} F_i(\Delta, \eta, t) I_{\alpha}^{X^jY^k}(v)(0).
\]

Our previous expression for \( I_{\alpha}^{X^jY^k}(v)(t) \) is now substituted into this, giving

\[
I_{\alpha}^X(t) = \frac{\sum_{j=1}^{d} \sum_{k=1}^{d} \int_{-\infty}^{0} \left( \int_{-\infty}^{0} F(\Delta, \eta, t, r) F(\Delta, \eta - \frac{\alpha}{4}, r, s) e^{\alpha_j r + \alpha_k s} dr \right) I_{\alpha}^{X^jY^k}(v)(s) ds}{\sum_{j=1}^{d} \sum_{i=0}^{W} \left( \int_{-\infty}^{0} F(\Delta, \eta, t, s) F_i(\Delta, \eta - \frac{\alpha}{4}, s) e^{\alpha_j s} ds \right) I_{\alpha}^{X^jY^k}(v)(0)} + \sum_{i=0}^{W} F_i(\Delta, \eta, t) I_{\alpha}^{X^jY^k}(v)(0).
\]

As previously (the case \( \alpha \geq 2\eta \)), the bounds from Proposition 8 and (34) are now used to bound the terms \( \int_{-\infty}^{0} F(\Delta, \eta, t, r) F(\Delta, \eta - \frac{\alpha}{4}, r, s) e^{\alpha_j r + \alpha_k s} dr \), \( \int_{-\infty}^{0} F(\Delta, \eta, t, s) F_i(\Delta, \eta - \frac{\alpha}{4}, s) e^{\alpha_j s} ds \), and \( F_i(\Delta, \eta, t) \) as in the statement of the proposition.

In the case \( \alpha \leq \eta \), a similar argument is used, though now involving higher order multi-indices and backwards induction on the order of these; recall that \( k_0 = \lfloor \frac{2\eta}{\alpha} \rfloor \), and thus \( k_0 \leq \frac{2\eta}{\alpha} < k_0 + 1 \). We now define

\[
\epsilon = \frac{\alpha(k_0 + 1) - 2\eta}{4(k_0 - 1)}.
\]

Note that \( \epsilon \in (0, \frac{\alpha}{4}) \), and \( \eta - \alpha < (k_0 - 1)(\frac{\alpha}{4} - \epsilon) \). For \( 0 \leq \ell \leq k_0 - 1 \), define \( \delta_{\ell} = l(\frac{\alpha}{2} - \epsilon) \). By (35), we have that for any multi-index \( k' \) of order \( k_0 - 1 \),

\[
\left( \prod_{i=1}^{W} \left( \frac{\alpha_i}{\delta_{\ell}} - \lambda_i \right) \right) I_{\alpha}^{X^jY^k}(v)(t) = -\sum_{j=1}^{d} e^{\alpha_j t} I_{\alpha}^{X^jY^k}(v)(t),
\]

and by (36), we have

\[
\left\| -\sum_{j=1}^{d} e^{\alpha_j t} I_{\alpha}^{X^jY^k}(v)(t) \right\| \ll d \| \chi \|_{W^{k_0-1}(U^+)} \| v \|_{S^{(k_0-1)\alpha+1}(U)} e^{\delta_{\ell}t}.
\]

Proposition 8 may thus be applied to \( I_{\alpha}^{X^jY^k}(v)(t) \) with \( \psi(t) = -\sum_{j=1}^{d} e^{\alpha_j t} I_{\alpha}^{X^jY^k}(v)(t) \), \( \gamma = \alpha \), and \( \beta_\gamma = \frac{\eta}{4} \), giving

\[
I_{\alpha}^{X^jY^k}(v)(t) = \sum_{j=1}^{d} \int_{-\infty}^{0} F(\Delta, \frac{\eta}{4}, t, s) e^{\alpha_j s} I_{\alpha}^{X^jY^k}(v)(s) ds + \sum_{i=0}^{W} F_i(\Delta, \frac{\eta}{4}, t) I_{\alpha}^{X^jY^k}(v)(0).
\]
By Proposition 8 ii), \( |F(\Lambda, \frac{\alpha}{2}, t, s)e^{\alpha t}| \leq e^{\delta_1 t + (\frac{\alpha}{2}-\epsilon)s} \), and by Proposition 8 iii) and (34), \( |F_l(\Lambda, \frac{\alpha}{2}, t)| \leq e^{\delta_l t} \inf_{w \in H^\infty} |w|^m \) (for any \( m \in \mathbb{N} \)). We now use induction: assume that for any multi-index \( i \) of order \( k_0 - l \) (for some \( l \in \{1, \ldots, k_0 - 1\} \)), there exist sets of \( \mathbb{C} \)-valued functions \( \{A_{ij}\} \) and \( \{B_{ij}\} \) such that

\[ \mathcal{I}_{d\pi_k}(U_i)u(t) = \sum_{j \in I_k} \int_{-\infty}^{0} A_{ij}(t, s) \mathcal{I}_{d\pi_k}(U_i)u(s) ds + \sum_{l \in I_k} \sum_{i=0}^{W} B_{ij}(t) \mathcal{I}_{d\pi_k}(Y_i U_i)u(0), \]

where

i) \( |A_{ij}(t, s)| \leq C_{G, \eta, \alpha} e^{\delta_1 t + (\frac{\alpha}{2}-\epsilon)s} \),

ii) \( |B_{ij}(t)| \leq C_{G, \eta, \alpha} e^{\delta_l t} \inf_{w \in H^\infty} |w|^m \) \( \forall m \in \mathbb{N} \).

Under these assumptions, using the bound \( \|\mathcal{I}_{d\pi_k}(X_i Y_i)u\| \leq C \) in (38) immediately gives

\[ \|\mathcal{I}_{d\pi_k}(U_i)u\| \leq C e^{\delta_1 t}. \]

We now select an arbitrary multi-index \( \kappa \) of order \( k_0 - l + 1 \). Since \( \mathcal{I}_{d\pi_k}(U_i)u(t) \) satisfies the differential equation (35), the right-hand side of which consists of terms of the form \( e^{\alpha t} \mathcal{I}_{d\pi_k}(X_i Y_i)u(t) \), (39) may be used to bound the right-hand side of (35) by \( e^{(\delta_1 + \alpha)t} \). This enables us to apply Proposition 8 to \( \mathcal{I}_{d\pi_k}(U_i)u(t) \) with \( \psi(t) = -\sum_{j=1}^{d} e^{\alpha_j t} \mathcal{I}_{d\pi_k}(X_k Y_k)u(t), \gamma = \delta_t + \alpha, \) and \( \beta_t = \delta_t + \frac{\alpha}{2} \) (indeed, note that \( \delta_t + \frac{\alpha}{2} \leq (k_0 - 1)(\frac{\alpha}{2} - \epsilon) + 2 < k_0 \frac{\alpha}{2} \leq \epsilon \)), giving

\[ \mathcal{I}_{d\pi_k}(X_k Y_k)u(t) = -\sum_{j=1}^{d} \int_{-\infty}^{0} F(\Lambda, \delta_t + \frac{\alpha}{2}, t, s)e^{\alpha_j t} \mathcal{I}_{d\pi_k}(X_k Y_k)u(s) ds + \sum_{i=0}^{W} F_i(\Lambda, \delta_t + \frac{\alpha}{2}, t) \mathcal{I}_{d\pi_k}(Y_i U_i)u(0). \]

We may now apply Proposition 8 iii) and (31) to get that for any \( m \in \mathbb{N} \) and \( u \in H^\infty \), \( |F_l(\Lambda, \delta_t + \frac{\alpha}{2}, t)| \leq e^{\delta_l t} \inf_{w \in H^\infty} |w|^m \) (since \( \delta_t = \delta_t + \frac{\alpha}{2} \)). Let \( j_i \) be the multi-index of order \( k_0 - l \) corresponding to \( X_k Y_k \) and \( U_i U_k \). The induction hypothesis is then used once again: for each \( j_i \), we have

\[ \mathcal{I}_{d\pi_k}(X_k Y_k)u(t) = \sum_{m \in I_k} \int_{-\infty}^{0} A_{ij,m}(t, s) \mathcal{I}_{d\pi_k}(U_m U_j)u(s) ds + \sum_{l \in I_k} \sum_{i=0}^{W} B_{ij,1,i}(t) \mathcal{I}_{d\pi_k}(Y_i U_i U_j)u(0), \]

where the sets of functions \( \{A_{ij,m}\} \) and \( \{B_{ij,1,i}\} \) satisfy the bounds of the induction hypothesis. This is entered into (40), giving

\[ \mathcal{I}_{d\pi_k}(X_k Y_k)u(t) = \sum_{j=1}^{d} \sum_{m \in I_k} \int_{-\infty}^{0} \left( \int_{-\infty}^{0} -F(\Lambda, \delta_t + \frac{\alpha}{2}, t, r)e^{\alpha_j r} A_{ij,m}(r, s) dr \right) \mathcal{I}_{d\pi_k}(U_m U_j)u(s) ds + \sum_{j=1}^{d} \sum_{l \in I_k} \sum_{i=0}^{W} \left( \int_{-\infty}^{0} -F(\Lambda, \delta_t + \frac{\alpha}{2}, t, s)e^{\alpha_j s} B_{ij,1,i}(s) ds \right) \mathcal{I}_{d\pi_k}(Y_i U_i U_j)u(0) + \sum_{i=0}^{W} F_i(\Lambda, \delta_t + \frac{\alpha}{2}, t) \mathcal{I}_{d\pi_k}(Y_i U_k)u(0). \]

Collecting terms and using the induction hypothesis, the bounds from Proposition 8 and (31) complete the induction: (38), i, and ii) are thus valid for any \( k \) of order \( k_0 - l \), for all \( l, 1 \leq l \leq k_0 \). In particular, for \( l = k_0 \), (38), i and ii) give bounds of the same form as those in Proposition 10 though not as sharp. In order to obtain the desired bounds, we carry out the last step (viz., \( l = k_0 - 1 \)) of the previous induction with the single modification
that we use $\beta_\gamma = \eta$ in place of $\beta_\gamma = \delta_{k_0-1} + \frac{\pi}{2}$; we thus apply Proposition 8 to $I_\nu^X(t)$ with $
abla(t) = -\sum_{j=1}^d e^{\alpha_j r} T_{d_{\nu}(U_j)}^X(t), \gamma = \delta_{k_0-1} + \alpha$ and $\beta_\gamma = \eta$. The assumptions of Proposition 8 are still fulfilled; indeed, the only condition that needs a new verification is $\beta_\gamma < \eta$, i.e. $\eta < \delta_{k_0-1} + \alpha$, and this is an easy consequence of our choice of $\epsilon$. Arguing as before, we obtain

$$I_\nu^X(t) = \sum_{j=1}^d \sum_{m \in B_{k_0-2}} \int_{-\infty}^0 \left( \int_{-\infty}^0 -F(\Delta, \eta, t, r)e^{\alpha_j r} A_{j,m}(r,s) dr \right) T_{d_{\nu}(U_j)}^X(s) ds + \sum_{j=1}^d \sum_{m \in B_{k_0-2}} \sum_{i=0}^W \int_{-\infty}^0 -F(\Delta, \eta, s)e^{\alpha_j s} B_{j,i}(s) ds \right) T_{d_{\nu}(U_j)}^X(s) ds + \sum_{i=0}^W F_i(\Delta, \eta, t) I_\nu^X(0),$$

where $j_j = (j), j = 1, \ldots, d$. Now arguing in the same manner as previously gives the bounds stated in the proposition.

**Remark 6.** The functions $A_{i,j}, B_{i,j}$ in the proof can be made completely explicit; they consist of integrals of products of the functions $F_i, F_i$ from Proposition 8. The uniformity of the bounds in Proposition 10 with respect to the representation $(\pi, \mathcal{H})$ allows us to use the identity (37) in all the representations in the decomposition of $(\rho, V)$ into irreducibles. This will be important in the proof of Theorem 1, where we essentially “integrate” (37) over the direct integral decomposition of $(\rho, V)$. Furthermore, the dependency on $g_{\mathbb{R}}$ in (2) can be quantified, namely it comes from the bound (39), i.e. $1 + |\Delta|^W \ll_{G, g_{\mathbb{R}}, m} \inf_{\nu \in \mathcal{H} \setminus \{0\}} \|\nu\|_{G, g_{\mathbb{R}}, m}$ (the bound in (2) is stated in this manner to ensure that the implied constant does not depend on $(\pi, \mathcal{H})$). This will be of importance when proving Theorem 2 (in Section 6), where we will let $g_{\mathbb{R}}$ vary within the positive Weyl chamber $A^\perp$.

4. **Sobolev Inequalities**

While the results of Section 3 allow us to draw various conclusions regarding convergence in $L^2(\Gamma \backslash G)$ of translates of averages of functions, in order to make pointwise statements for non-uniform $\Gamma$ we use an automorphic Sobolev inequality from [1, Appendix B]. Before introducing this inequality, we recall some facts regarding the reduction theory of $\Gamma \backslash G$.

4.1. **Reduction Theory.** Most of the material of this section is drawn from [37, Chapter 2]. Recall that we have assumed that $\Gamma$ fulfills the assumption of Langlands (cf. [25, Chapter 2, p. 16], also [37, pp. 62-63]). Before stating the main result regarding the reduction theory, we introduce the concept of *split components*. Let $P = NAM$ be a parabolic subgroup with corresponding Lie algebra decomposition $n \oplus a \oplus m$. Given a subalgebra $a' \subset a$, for $\lambda \in a'^*$ we define

$$n_\lambda = \{ X \in n : ad_H(X) = \lambda(H)X \forall H \in a' \}.$$ 

Letting $\Sigma^+(g, a') = \{ \lambda| a' : \lambda \in \Sigma^+(g, a) \}$, we then have that

$$n = \bigoplus_{\lambda \in \Sigma^+(g, a')} n_\lambda.$$ 

Denoting the orthogonal complement of $a'$ in $a$ with respect to the Killing form by $a'^\perp$, $a'$ and its corresponding analytic subgroup $A$ are both called *split components* of $P$ if

$$\text{tr}(ad_H|_{n_\lambda}) = 0 \quad \forall H \in a'^\perp, \lambda \in \Sigma^+(g, a'),$$

cf. [25, pp. 3-5] and [37, pp. 30-32]. One can show that if $a'$ is a split component, then $0 \notin \Sigma^+(g, a')$, as well as

$$\rho_{a'} = \rho_{a'} := \frac{1}{2} \sum_{\lambda \in \Sigma^+(g, a')} \dim n_\lambda.$$
Moreover, there is a unique set \( \Sigma_0^+(g, a') \subset \Sigma^+(g, a') \) of simple roots (cf. \([37\text{ p. } 8]\) and \([37\text{ p. } 34]\)) for \( a' \); of importance to us are the facts that \( \Sigma_0^+(g, a') \) is a basis for \( a'^\perp \), and every element of \( \Sigma^+(g, a') \) may be written as a linear combination \( \sum_{\lambda \in \Sigma_0^+(g, a')} m_\lambda \lambda \), with all \( m_\lambda \in \mathbb{Z}_{\geq 0} \).

For a split component \( a' \), let \( A^+ \) be the analytic subgroup of \( G \) corresponding to the subalgebra \( a'^\perp \). We now define a subgroup \( S \subseteq P \) by \( S = NMA^+ \). The pair \( (P, S) \) is called a \textit{split parabolic pair}, and the triple \((P, S; A)\) a \textit{split parabolic pair with split component} \( A \). We now define a few important subsets: for an arbitrary real number \( \tau \), let

\[
A_\tau = \{ \exp(H) : H \in a', \lambda(H) \geq \tau \quad \forall \lambda \in \Sigma_0^+(g, a') \}
\]

\[
A_\Gamma = \{ \exp(H) : H \in a', |\lambda(H)| \leq \tau \quad \forall \lambda \in \Sigma_0^+(g, a') \}.
\]

For a compact subset \( \Omega \subset S \), define

\[
\Omega_\tau = \bigcup_{a \in A_\tau} a^{-1}\Omega a.
\]

Observe that \( A_\Gamma \) is a compact subset of \( A_\tau \), and \( \Omega_\tau \) is a relatively compact subset of \( S \). We now define a \textit{Siegel set} \( \mathfrak{S}_{\Omega, \tau} \) with respect to \((P, S; A)\) by

\[
\mathfrak{S}_{\Omega, \tau} = \Omega A_\tau K.
\]

A split parabolic pair \((P, S)\) is said to be \textit{\( \Gamma \)-percuspidal} if \( i) \Gamma \cap P \subset S \) \( ii) \Gamma \cap N \) is cocompact in \( N \) \( iii) \Gamma \cap S \) is cocompact in \( S \). We can now state the result from reduction theory that will be needed: combining \([37\text{ Theorem } 2.11]\) and the assumption on \( \Gamma \) (cf. \([37\text{ pp. } 62-63]\)), we have:

**Proposition 11.** There exist \( \tau_0 \in \mathbb{R} \), a standard parabolic split pair with split component \((P, S; A)\), and elements \( g_1, g_2, \ldots, g_\kappa \in G \) such that \( (P_i, S_i; A_i) = (g_iPg_i^{-1}, g_iSg_i^{-1}, g_iA_i^{-1}) \) are \( \Gamma \)-percuspidal, and

\[
G = \bigcup_{i=1}^\kappa \Gamma \mathfrak{S}_i,
\]

where the \( \mathfrak{S}_i = \Omega_i A_{i, \tau_0} K \) are Siegel sets with respect to \((P_i, S_i; A_i)\). Furthermore, the set of all \( \gamma \in \Gamma \) such that \( \gamma \left( \bigcup_{i=1}^\kappa \mathfrak{S}_i \right) \cap \left( \bigcup_{i=1}^\kappa \mathfrak{S}_i \right) \neq \emptyset \) is finite.

**Remark 7.** In the case that \( \Gamma \) is an arithmetic lattice in an algebraic group, the reduction theory of Borel and Harish-Chandra \([24]\) gives more: we may take \( P \) as the real points of a minimal \( \mathbb{Q} \)-parabolic subgroup of \( G_{\mathbb{Q}} \), \( A \) as a maximal \( \mathbb{Q} \)-split torus in \( A \), and the \( g_i \in G_{\mathbb{Q}} \). If \( \text{rank}(G) = 1 \), then the reduction theory is provided by Garland and Raghunathan \([12]\) (cf. \([37\text{ pp. } 14-18]\)).

Throughout the remainder of Section 4, we let \( P_i, S_i, A_i, \) and \( \mathfrak{S}_i \) \((i = 1, \ldots, \kappa)\) be as in Proposition 11. We also use the notation \( N_i = g_iNg_i^{-1}, A_i = g_iA_i^{-1}, M_i = g_iMg_i^{-1}, N_i^- = g_iN^-g_i^{-1} \). Finally, the Lie algebra of \( A_i \) is denoted \( a_i \), and the Lie algebra of \( A_i \) is denoted \( a_i' \).

### 4.2. Sobolev Inequalities

We now follow \([1\text{ Appendix } B]\) and state the previously mentioned automorphic Sobolev inequality. For a compact, symmetric neighbourhood \( B \) of \( e \) in \( G \), we define a function \( w_B \) on \( \Gamma \backslash G \) by letting \( w_B(x) \) \((x \in \Gamma \backslash G)\) be the operator norm of the mapping from \( L^2(\Gamma \backslash G) \) to \( L^2(B) \) given by \( f(\cdot) \mapsto f(x \cdot) \), i.e.

\[
w_B(x) = \inf \left\{ c \in \mathbb{R}_{>0} : \left( \int_B |f(xg)|^2 \, dg \right)^{1/2} \leq c \|f\|_{L^2(\Gamma \backslash G)} \quad \forall f \in L^2(\Gamma \backslash G) \right\},
\]

or, equivalently,

\[
w_B(x) = \sup_{f \in L^2(\Gamma \backslash G), f \neq 0} \frac{\left( \int_B |f(xg)|^2 \, dg \right)^{1/2}}{\|f\|_{L^2(\Gamma \backslash G)}}.
\]
From [11 Proposition B.2] we have the following:

**Proposition 12.** For \( m > \frac{\dim(G)}{2} \), we have

\[
|f(x)| \ll_{\Gamma,B} w_B(x) \|f\|_{S^m(\Gamma \setminus G)} \quad \forall f \in S^m(\Gamma \setminus G), \ x \in \Gamma \setminus G.
\]

The two properties of \( w_B \) established in the next lemma will be used at multiple points throughout the remainder of this section:

**Lemma 13.**

i) For any compact subset \( C \) of \( G \) and any \( x \in \Gamma \setminus G \), we have

\[
w_B(xg) \ll_{B,C} w_B(x) \quad \forall x \in \Gamma \setminus G, \ g \in C.
\]

ii) For all \( g \in G \), \( w_B(\Gamma g)^2 \leq \#(\Gamma \cap gBBg^{-1}) \).

**Proof.** Starting with i), assume that \( g \in C \). Let \( g_1, \ldots, g_r \in G \) be such that \( C \subseteq \bigcup_{i=1}^r B_{g_i} \).
We then have that for any \( f \in L^2(\Gamma \setminus G) \), \( x \in \Gamma \setminus G \), and \( g \in C \),

\[
\int_B |f(xg)|^2 \, dh = \int_{gB_g^{-1}} |f(xh)|^2 \, dh \leq \sum_{i=1}^r \int_{gB_{g_i}} |f(xg)|^2 \, dh = \sum_{i=1}^r \int_B |(\rho(g_i g)f)(xh)|^2 \, dh
\]

\[
\leq w_B^2(x) \sum_{i=1}^r \|\rho(g_i g)f\|_{L^2(\Gamma \setminus G)}^2 = w_B^2(x) r \|f\|_{L^2(\Gamma \setminus G)}^2,
\]

proving i). For ii), choose an exact fundamental domain \( \mathcal{F} \) for \( \Gamma \) in \( G \), and observe that (using 1 to denote characteristic functions)

\[
\int_B |f(\Gamma h)|^2 \, dh = \int_G |f(\Gamma h)|^2 1_B(h) \, dh = \int_G |f(\Gamma h)|^2 1_{gB}(h) \, dh
\]

\[
= \sum_{\gamma \in \Gamma} \int_{\gamma \mathcal{F}} |f(\Gamma h)|^2 1_{gB}(h) \, dh = \int_{\mathcal{F}} |f(\Gamma h)|^2 \left( \sum_{\gamma \in \Gamma} 1_{gB}(\gamma h) \right) \, dh
\]

\[
\leq \sup_{h \in \mathcal{F}} \#\{\gamma \in \Gamma : \gamma h \in gB\} \|f\|_{L^2(\Gamma \setminus G)}^2 = \sup_{h \in G} \#\{\gamma \in \Gamma : \gamma h \in gB\} \|f\|_{L^2(\Gamma \setminus G)}^2.
\]

Noting that if \( \gamma_0 \in \Gamma \), \( \#\{\gamma \in \Gamma : \gamma h \in gB\} = \#\{\gamma \in \Gamma : (\gamma_0 h) \in gB\} \), we have

\[
= \max_{h \in gB} \#\{\gamma \in \Gamma : \gamma h \in gB\} \|f\|_{L^2(\Gamma \setminus G)}^2 \leq \#\{\gamma \in \Gamma : gBg \neq gB \} \|f\|_{L^2(\Gamma \setminus G)}^2
\]

\[
= \#(\Gamma \cap gBBg^{-1}) \|f\|_{L^2(\Gamma \setminus G)}^2,
\]

the last equality holding due to \( B \) being symmetric.

**Remark 8.** More work shows that there is in fact equality between both sides of ii). This is not needed here, however.

**Proposition 14.** For \( g = s a k \in \mathfrak{S}_i \), with \( \mathfrak{S}_i \) as in Proposition [11]

\[
w_B(\Gamma s a k) \ll e^{\rho_n(a)},
\]

where the implied constant depends on \( \Gamma, \mathfrak{S}_i \) and \( B \).

**Proof.** Firstly, we note that it suffices to find some set \( B \) for which \( w_B \) satisfies (42), since for any other choice of \( B \), we may find \( g_1, g_2, \ldots, g_p \in G \) such that \( B \subseteq \bigcup_{i=1}^p g_j B \). We then have that for all \( x \in \Gamma \setminus G \) and \( f \in L^2(\Gamma \setminus G) \),

\[
\int_B |f(xg)|^2 \, dg = \sum_{j=1}^p \int_{g_j B} |f(xg)|^2 \, dg = \sum_{j=1}^p \int_B |f(xg_j^{-1} g)|^2 \, dg
\]

\[
\leq \sum_{j=1}^p w_B(xg_j^{-1}) \|f\|_{L^2(\Gamma \setminus G)}^2 \ll w_B(x) \|f\|_{L^2(\Gamma \setminus G)},
\]

so if (42) holds for \( w_B \), it will hold for \( w_B \) for all other choices of \( B \) as well.
For each $i$, we will find a set $\mathcal{B}_i$ such that (12) holds on $\mathfrak{g}_i$. Choosing $\mathcal{B}$ contained in the intersection of the sets $\mathcal{B}_i$ will then satisfy the requirements of the proposition. For any fixed $B$, we have

$$w_B(\Gamma\text{SAK}) = w_B(\Gamma a^{-1}a) \ll w_B(\Gamma a),$$

so we need only consider the restriction of $w_B$ to $\mathcal{A}_i$. Since $N_iA_iM_iN_i^-$ is open and dense in $G$, we may choose $\mathcal{B}_i$ small enough so that

$$\mathcal{B}_i\mathcal{B}_i \subset B_{N_i}B_{A_i}M_iB_{N_i}^-,$$

where each of the sets in the right-hand side is a small neighborhood of the identity in the corresponding subgroup. Let us now fix $T \geq 0$ large enough so that

$$\bigcup_{a' \in A_i,T} a'B_{N_i}^{-1} a'^{-1} \subset B_{N_i}^-.$$

Note that $A_i, T \subset A_i^{T + |\tau_0|}, A_{i,T} (\tau_0$ being as in (11); hence given any $a \in A_i, \tau_0$, there exist $a_+ \in A_i, T$ and $a_- \in A_i^{T + |\tau_0|}$ such that $a = a_+a_-$. We now have that

$$w_{\mathcal{B}_i}(\Gamma a) \ll w_{\mathcal{B}_i}(\Gamma a_+),$$

and

$$w_{\mathcal{B}_i}(\Gamma a_+)^2 \leq \#(\Gamma \cap a_+\mathcal{B}_i a_+^{-1}) \leq \#(\Gamma \cap a_+B_{N_i}B_{A_i}M_iB_{N_i}^{-1} a_+^{-1})$$

$$= \#(\Gamma \cap (a_+B_{N_i}^{-1}a_+^{-1})B_{A_i}M_i(a_+B_{N_i}^{-1} a_+^{-1}))$$

$$\leq \#(\Gamma \cap (a_+B_{N_i}^{-1}a_+^{-1})B_{A_i}M_iB_{N_i}^-).$$

Since we have assumed that $P_i$ is $\Gamma$-percuspidal, $\Gamma_{N_i} = \Gamma \cap N_i$ is cocompact in $N_i$, and there therefore exists a neighbourhood $B_0$ of $e$ in $G$ such that $\Gamma \cap N_iB_0 = \Gamma_{N_i}$. We now additionally assume that $\mathcal{B}_i$ has been chosen small enough so that $B_{A_i}M_iB_{N_i}^- \subset B_0$. This further assumption allows us to conclude that

$$w_{\mathcal{B}_i}(\Gamma a_+)^2 \leq \#(\Gamma_{N_i} \cap a_+B_{N_i}a_+^{-1}B_0) = \#(\Gamma_{N_i} \cap a_+B_{N_i}a_+^{-1}).$$

Using now the fact that $N_i$ is simply connected and nilpotent (and that $\Gamma_{N_i}$ is a lattice in $N_i$), we turn this into a Euclidean counting problem. Let $\{L_j\}$ be a basis of the Lie algebra of $N_i, n_i$, that is aligned with the restricted root space decomposition of $\mathfrak{g}$ with respect to $\mathfrak{a}_i$. More precisely, for each $j = 1, \ldots, d = \dim n_i$, there exists some $\lambda_j \in \Sigma^+(\mathfrak{g}, \mathfrak{a}_i)$ such that $[H, L_j] = \lambda_j(H)L_j$ for all $H \in \mathfrak{a}_i$. We now define a map $\phi : \mathbb{R}^d \to N_i$ by

$$\phi(x) = \exp \left( \sum_{j=1}^d x_j L_j \right),$$

and observe that by [24] Theorem 1.127, $\phi$ is a diffeomorphism. Without loss of generality, we may assume that the basis $\{L_j\}$ has been chosen so that $B_{N_i} \subset \mathfrak{B} = \phi([-1,1]^d)$. For any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $H \in \mathfrak{a}_i$, we have

$$\exp(H)\phi(x)\exp(-H) = \exp \left( e^{adH} \sum_{j=1}^d x_j L_j \right) = \exp \left( \sum_{j=1}^d e^{\lambda_j(H)x_j} L_j \right).$$

In particular, if $\phi(x) \in B_{N_i}$, then by assumption $x \in [-1,1]^d$, and so $|x_j e^{\lambda_j(H)}| \leq e^{\lambda_j(H)}$. This gives

$$\phi^{-1}(a_+B_{N_i}a_+^{-1}) \subset \phi^{-1}(a_+\mathfrak{B}a_+^{-1}) = \prod_{j=1}^d [-e^{\lambda_j(\log a_+)}, e^{\lambda_j(\log a_+)}].$$

Since $a_+ \in A_i,T$, we have $\lambda_j(\log a_+) \geq T$ for all $j$, hence

$$1 \ll \text{Vol}(\phi^{-1}(a_+B_{N_i}a_+^{-1})) \leq \text{Vol}(\phi^{-1}(a_+\mathfrak{B}a_+^{-1})) = 2^d e^{2\rho_{a_+}(a_+)}.$$
By \cite{1} Proposition 5.4.8 (b)], there exists a lattice $L_i$ in $n_i$ and a finite number of elements $Y_1, \ldots, Y_k \in n_i$ such that $\log(\Gamma_N) = \bigcup_{j=1}^k Y_j + L_i$. This gives that $\phi^{-1}(\Gamma_N)$ is contained in a finite union of affine lattices in $\mathbb{R}^d$, hence

$$\#(a_+B_N, a_-^{-1} \cap \Gamma_N) = \#(\phi^{-1}(a_+B_N, a_-^{-1}) \cap \phi^{-1}(\Gamma_N)) \leq \#(\phi^{-1}(a_+B, a_-^{-1}) \cap \phi^{-1}(\Gamma_N)).$$

Since the side lengths of the rectangular box $\phi^{-1}(a_+B, a_-^{-1})$ are bounded from below by $2e^T$ (uniformly over all $a_+ \in A_i(T)$), we have

$$\#(\phi^{-1}(a_+B, a_-^{-1}) \cap \phi^{-1}(\Gamma_N)) \ll \text{Vol}(\phi^{-1}(a_+B, a_-^{-1})) \ll e^{2\rho_a(a_+)}.$$

Now, since $a = a_+ a_-$,

$$e^{2\rho_a(a_+)} \leq \left( \max_{a_0 \in A_i} e^{-2\rho_a(a_0)} \right) e^{2\rho_a(a)},$$

giving, for $sak \in S_i$:

$$w_{B_i}(\Gamma sak) \ll e^{\rho_a(a)}.$$

We now fix, once and for all, a compact, symmetric neighbourhood $B$ of the identity in $G$, and define the invariant height function $\gamma_i \Gamma$ on $G \setminus \Gamma$ through $\gamma_i \Gamma(x) := w_B(x)$. Proposition \cite{2} gives that for $m > \frac{\dim G}{2}$,

$$|f(x)| \ll \|f\|_{\mathcal{S}^m(\Gamma \setminus G)} \gamma_i \Gamma(x) \quad \forall f \in \mathcal{S}^m(\Gamma \setminus G), x \in \Gamma \setminus G.$$

**Corollary 15.** $\gamma_i \Gamma \in L^1(\Gamma \setminus G)$.

**Proof.** By decomposing the measure on $S_i$ in a manner similar to \cite{3} (cf. \cite{5} p. 25), and then using \cite{6}, we have

$$\int_{\Gamma \setminus G} \gamma_i \Gamma(x) \, d\mu(x) \leq \sum_{i=1}^\infty \int_{\gamma_i \Gamma g} \gamma_i \Gamma(x) \, dg = \sum_{i=1}^\infty \int_{\gamma_i \Gamma g} \int_{A_i, r_0} \int_{K} \gamma_i \Gamma(\Gamma sak) e^{-2\rho_a(a)} \, dk \, da \, ds \ll \int_{A_i, r_0} e^{\rho_a(a)} e^{-2\rho_a(a)} \, da.$$

We now use the fact that the Haar measure $da$ on $A_i$ may be expressed as the pushforward (under the exponential map) of a constant multiple of the Lebesgue measure $dm$ on $a'_i$. Letting $H_1, \ldots, H_r$ be the basis of $a'_i$ defined by $\alpha_j(H_j) = \delta_{jl}$ for all $\alpha_j \in \Sigma_0^+(g, a'_i) = \{\alpha_1, \ldots, \alpha_r\}$, we have

$$\int_{A_i, r_0} e^{-\rho_a(a)} \, da \times \int_{\{H \in a'_i : \lambda(H) > r_0 \forall \lambda \in \Sigma_0^+(g, a'_i)\}} e^{-\rho_a(H)} \, dm(H) \ll \prod_{j=1}^r \int_{r_0}^\infty e^{-y\rho_a(H_j)} \, dy_j < \infty.$$

We conclude this section by giving bounds on integrals of $\gamma_i \Gamma$ over translates of pieces of our horospherical subgroup $U^+$. Such bounds will only be needed for integrals over relatively compact sets $B$ of positive measure in $U^+$. This will allow us to “thicken” $B$ to a set $C$ of positive measure in $G$ such that $B \subset C$, and make use of the wavefront property: $B g_i$ and $C g_i$ will remain “close” to each other as $t \to -\infty$. In \cite{3} Proposition 6], we considered the case $G = SL(2, \mathbb{C})$, and proved a corresponding result for an integral along the boundary of such a piece of a horosphere (subject to certain restrictions on the shape of the subset). Since the boundary has measure zero, the method used here does not work, and the proof becomes considerably more complicated. Here, though, the proof is relatively straightforward:

**Corollary 16.** Let $B \subset U^+$ be relatively compact. Then

$$\int_B \gamma_i \Gamma(x u g_t) \, du \ll_{G, B} \gamma_i^2 \Gamma(x) \quad \forall x \in \Gamma \setminus G, t \leq 0.$$
Proof. Let $P = \text{NAM}$ be the parabolic subgroup such that $U^+ = N$ and $g_\mathbb{R} \subset A$. We start by choosing a compact symmetric neighbourhood $B_0$ of $e$ in $\text{AMN}^- \subset G$. Let $\tilde{B}_0$ denote the closure of $\bigcup_{s \leq 0} g_{s} B_0 g_{s}$; as in (1), $\tilde{B}_0$ is a compact neighbourhood of $e$ in $\text{AMN}^-$, and

$$\mathcal{Y}_T(xg_t) = \mathcal{Y}_T(xbg_t(g_{-tb^1}g_t)) \leq \mathcal{Y}_T(xbg_t) \quad \forall x \in \Gamma \setminus G, t \leq 0, b \in \tilde{B}_0.$$ 

After taking into account the modular function $\Delta(\alpha m) = e^{-2\rho_a(\alpha)}$ of $\text{AM}$ on $\tilde{B}_0$, we have (cf. (6))

$$\int_B \mathcal{Y}_T(xug_t) \, du \ll \int_{\tilde{B}_0} \mathcal{Y}_T(xgg_t) \, dg.$$ 

We now choose $g_1, \ldots, g_R \in G$ such that $\tilde{B}_0 B_0 \subset \bigcup_{j=1}^R B_0 g_j$. By Corollary $\ref{cor:bound}$, $\mathcal{Y}_T \in L^2(\Gamma \setminus G)$, and so from the definition of $\mathcal{Y}_T$:

$$\int_{\tilde{B}_0} \mathcal{Y}_T(xgg_t) \, dg \leq \sum_{j=1}^R \int_B \left| \rho(g_j g_t) \sqrt{\mathcal{Y}_T}(xg) \right|^2 \, dg$$

$$\leq \sum_{j=1}^R \mathcal{Y}_T^2(x) \left| \rho(g_j g_t) \sqrt{\mathcal{Y}_T} \right|_{L^2(\Gamma \setminus G)}^2 \ll \mathcal{Y}_T^2(x).$$

\[\square\]

If $G$ is an algebraic group and $\Gamma$ is an arithmetic lattice in $G$, we may use Siegel’s Conjecture, proved by Ji $\cite{ji}$ and Leuzinger $\cite{leuzinger}$, to give a bound on a corresponding integral of $\mathcal{Y}_T^2$.

Corollary 17. Suppose $G$ is a semisimple algebraic group defined over $\mathbb{Q}$, $G = G(\mathbb{R})$, with rank greater or equal to 2, and $\Gamma \subset G(\mathbb{Q})$ is an arithmetic subgroup. Then there exists $p > 0$ such that for any relatively compact $B \subset U^+$,

$$\int_B \mathcal{Y}_T^2(xug_t) \, du \ll_{\Gamma, B, g_\mathbb{R}} \left(1 + \text{dist}_{\Gamma \setminus G}(\Gamma e, x) + |t| \right)^p \mathcal{Y}_T^2(x) \quad \forall x \in \Gamma \setminus G, t \leq 0.$$ 

Proof. Letting $R = \max_{a \in B} \text{dist}(e, u)$ and $\|X\| = \sqrt{(X, X)}$ for $X \in g$, we have

$$\text{dist}_{\Gamma \setminus G}(xug_t, \Gamma e) \leq \text{dist}_{\Gamma \setminus G}(x, \Gamma e) + R + |t|\|Y\| \quad \forall x \in \Gamma \setminus G, u \in B, t \in \mathbb{R}.$$ 

Making the definitions $\mathcal{B}_r = \{y \in \Gamma \setminus G : \text{dist}_{\Gamma \setminus G}(\Gamma e, y) \leq r\}$ and $r(x, t) = \text{dist}_{\Gamma \setminus G}(x, \Gamma e) + R + |t|\|Y\|$, we then have that

$$\mathcal{Y}_T^2(xug_t) = \mathcal{Y}_T^2(xug_t) \mathcal{1}_{\mathcal{B}_r(x, t)}(xug_t) \quad \forall x \in \Gamma \setminus G, u \in B, t \in \mathbb{R}.$$ 

We denote $F = \mathcal{Y}_T \mathcal{1}_{\mathcal{B}_r(x, t)}$, and note that $F \in L^2(\Gamma \setminus G)$. Now arguing as in the proof of Corollary $\ref{cor:bounds}$ for $t \leq 0$,

$$\int_B \mathcal{Y}_T^2(xug_t) \, du = \int_B |F(xug_t)|^2 \, du \ll \sum_{j=1}^R \mathcal{Y}_T^2(x) \|\rho(g_j g_t)F\|_{L^2(\Gamma \setminus G)}^2 \ll \mathcal{Y}_T^2(x) \|F\|_{L^2(\Gamma \setminus G)}^2.$$ 

All that now remains is to bound $\|F\|_{L^2(\Gamma \setminus G)}^2$. From $\cite{leuzinger}$ Theorem 5.7 or $\cite{ji}$ Theorem 7.6, there exists a constant $D > 0$ such that

$$\text{dist}_{G}(g, e) \leq \text{dist}_{\Gamma \setminus G}(\Gamma g, \Gamma e) + D \quad \forall g \in \bigcup_{i=1}^\kappa \mathcal{G}_i,$$

and hence

$$\mathcal{1}_{\mathcal{B}_r(x, t)}(\Gamma g) = 1 \Rightarrow \text{dist}(g, e) \leq r(x, t) + D \quad \forall g \in \bigcup_{i=1}^\kappa \mathcal{G}_i.$$
This, together with Proposition 14, gives
\[
\int_{\Omega} F^2(\Gamma g) \, dg \ll \int_{A_i} \int_{A_i, r_0} \int_{K} e^{2\rho(a)} 1_{\mathcal{M}}(\Gamma g) e^{-2\rho(a)} \, dk \, da \, ds.
\]
Since \( \Omega_i \) and \( K \) are compact, there exists a constant \( c > 0 \) that enables us to bound the integral by
\[
\ll \int_{\{H \in \log \, A_i, r_0 : \|H\| \leq c + r(x, t) + D\}} dm(H) \ll (1 + r(x, t))^{\dim A_i}.
\]

\[\blacksquare\]

**Remark 9.** Using Garland and Raghunathan’s reduction theory [12], it should be possible to prove a similar result for any lattice in an arbitrary \( G \) of rank one.

\section*{5. Proof of Theorem 1}

**Proof of Theorem 1.** We start by making some initial reductions. The constant \( m_1 \) for which we prove (3) will satisfy \( m_1 > \frac{\dim G}{2} \), and thus by Proposition 12 both sides of (3) depend continuously on \( f \in S^{m_1}(V) \) (with respect to \( \| \cdot \|_{S^{m_1}(V)} \)). Hence, since \( V^\infty \) is dense in \( S^{n_2}(V) \), without loss of generality, from now on we may assume that \( f \in V^\infty \). Similarly, we may assume that \( \chi \in C^\infty_c(B) \).

We let \( \left( \int Z \, \pi, d\nu(\zeta), \int Z \, \mathcal{H}, d\nu(\zeta) \right) \) denote the direct integral decomposition of \((\rho, V)\) into irreducible unitary representations, and \( \int Z f(\zeta) d\nu(\zeta) \) be the corresponding decomposition of \( f \). Using the notation \( T^s_f(t) = T^s_f(g_t) \) from Section 3.2, we then have
\[
T^s_f(t) = \int_{U^+} \chi(u) r(u, t) f(\zeta) d\nu(\zeta) = \int_{Z} \int_{U^+} \chi(u) r(u, t) f(\zeta) d\nu(\zeta) = \int_{Z} T^s_f(t) d\nu(\zeta).
\]

Proposition 10 is now applied with \( \alpha = \min_j \alpha_j \) to \( v \)-a.e. \( f_\zeta \) (by Lemma 3) \( \eta \) is a rate of decay for the matrix coefficients of \( g_R \) \( v \)-a.e. \((\pi_\zeta, \mathcal{H})\), giving
\[
T^s_{f_\zeta}(t) = \sum_{k \in K_{c, k}} \int_{-\infty}^0 C_k^{X, \zeta}(t, s) T^s_{f_\zeta}(U_k)(s) ds + \sum_{j \leq s_{k+1} - 1} \sum_{i=0}^W D^{X, \zeta}_j(t) T^s_{f_\zeta}(U_j)(0) d\mu(\zeta).
\]

By integrating this identity over \( Z \), we get
(44) \[
T^s_f(t) = \sum_{k \in K_{c, k}} \int_{-\infty}^0 C_k^{X, \zeta}(t, s) T^s_{f_\zeta}(U_k)(s) ds d\nu(\zeta)
\]
\[
+ \sum_{j \leq s_{k+1} - 1} \sum_{i=0}^W D^{X, \zeta}_j(t) T^s_{f_\zeta}(U_j)(0) d\nu(\zeta).
\]

We now form intertwining operators \{\( T_k(\tau, s) \), \( S_{j, i}(\tau) \)\} in the manner of Section 2.3 for \( \phi \in V^\infty \), define
\[
T_k(\tau, s) \phi = \int_{Z} C_k^{X, \zeta}(\tau, s) \phi \, d\mu(\zeta)
\]
\[
S_{j, i}(\tau) \phi = \int_{Z} D^{X, \zeta}_j(\tau) \phi \, d\mu(\zeta).
\]

The bounds provided in Proposition 10 suffice to exchange the order of integration in (44). Furthermore, from Proposition 10 (2), each \( S_{j, i} \) is a linear map \( V^\infty \to V^\infty \) which is continuous with respect to two suitably chosen Sobolev norms. We may thus rewrite (44) as
\[
T^s_f(t) = \sum_{k \in K_{c, k}} \int_{-\infty}^0 T_k(t, s) T^s_{f_\zeta}(U_k)(s) ds + \sum_{j \leq s_{k+1} - 1} \sum_{i=0}^W S_{j, i}(t) T^s_{f_\zeta}(U_j)(0).
\]
The definition of $I$, and the fact that all of the operators $T_k(t,s)$ and $S_{j,i}(t)$ commute with $\rho$, gives

$$\int_{U^+} \chi(u)\rho(ug_t) f du = \sum_{k \in I_{\gamma_0}} \int_{-\infty}^0 \int_{U^+} (X_k \chi)(u)\rho(ug_s)T_k(t,s) d\rho(U_k) f du ds$$

$$+ \sum_{j \in I_{\leq \delta_0-1}} \sum_{i=0}^{W} \int_{U^+} (X_j \chi)(u)\rho(u)S_{j,i}(t)d\rho(Y^iU_j) f du.$$

We now apply the functional “evaluation at $x$” to both sides: if $m > \frac{\dim(G)}{2}$, then by \ref{11} (and recalling that $\text{supp}(\chi) \subseteq B$),

$$\left| \int_{U^+} \chi(u)f(xug_t) du \right| \leq \Gamma_m \sum_{k \in I_{\gamma_0}} \|X_k \chi\|_{L^\infty(B)} \int_{-\infty}^0 \|T_k(t,s) d\rho(U_k) f\|_{S^m(\Gamma \setminus G)} \int_B \gamma_1(xug_t) du ds$$

$$+ \sum_{j \in I_{\leq \delta_0-1}} \sum_{i=0}^{W} \|X_j \chi\|_{L^\infty(B)} \|S_{j,i}(t)d\rho(Y^iU_j) f\|_{S^m(\Gamma \setminus G)} \int_B \gamma_1(xu) du.$$

Corollary \ref{16} is now used to bound the integrals over $B$ by $\gamma_1(x)^2$. Decomposing the Sobolev norms $\|T_k(t,s) d\rho(U_k) f\|_{S^m(\Gamma \setminus G)}$ and $\|S_{j,i}(t) d\rho(Y^iU_j) f\|_{S^m(\Gamma \setminus G)}$ as in \ref{11} and applying the bounds from Proposition \ref{10} gives

$$\|T_k(t,s) d\rho(U_k) f\|_{S^m(\Gamma \setminus G)} \ll (1 + |t|^W e^{\frac{\alpha t}{\hat{s}}} t) \|f\|_{S^m(\Gamma \setminus G)}$$

$$\|S_{j,i}(t) d\rho(Y^iU_j) f\|_{S^m(\Gamma \setminus G)} \ll (1 + |t|^W e^{\frac{\alpha t}{\hat{s}}} t) \|f\|_{S^m(\Gamma \setminus G)}$$

for some $m_1 \in \mathbb{N}$. Entering these bounds into \ref{10} and choosing $m_2$ so that all factors involving $\chi$ in \ref{46} are bounded by $\|\chi\|_{W^{m_2,\infty}(B)}$ proves the theorem for any $f \in V^\infty$ and $\chi \in C_c^\infty(B)$. As noted at the beginning of the proof, the continuous dependency on the functions in \ref{9} may now be used to conclude that the same holds for all $f \in S^{m_1}(\Gamma \setminus G)$ and $\chi \in W_0^{m_2,\infty}(B)$. \hfill \square

\textbf{Remark 10.} The proof of Theorem \ref{1} simplifies somewhat when considering the equidistribution of translates of entire closed horospherical orbits. Suppose that $\Gamma \cap U^+$ is a lattice in $U^+$. Let $\varphi \in C_c^\infty(U^+)$ be such that $\varphi(u) \geq 0$ for all $u \in U^+$ and $\int_{U^+} \varphi(u) du = 1$. We now let $\chi = 1_F * \varphi \in C_c^\infty(U^+)$, where $F$ is a reasonable fundamental domain for $\Gamma \cap U^+$ in $U^+$. For all $u_0 \in U^+$ we have

$$\sum_{\gamma \in \Gamma \cap U^+} \chi(\gamma u_0) = \sum_{\gamma \in \Gamma \cap U^+} (1_F * \varphi)(\gamma u_0) = \int_{U^+} \left( \sum_{\gamma \in \Gamma \cap U^+} 1_F(\gamma u_0 u^{-1}) \right) \varphi(u) du = 1,$$

since $\sum_{\gamma \in \Gamma \cap U^+} 1_F(\gamma u_0 u^{-1}) = 1$ for almost all $u \in U^+$. From this we see that

$$\int_{U^+} \chi(u)\phi(\Gamma u_0) du = \int_{\Gamma \cap U^+} \left( \sum_{\gamma \in \Gamma \cap U^+} \chi(\gamma u_0) \right) \phi(\Gamma u_0) du = \int_{\Gamma \cap U^+} \phi(\Gamma u_0) du$$

for any $\phi \in V^\infty$. For $X \in u^+$, we have $X\chi = 1_F * (X\varphi)$. Since $\varphi$ has compact support, $\int_{U^+} (X\varphi)(u) du = \int_{U^+} \frac{d}{dt} \bigg|_{t=0} \varphi(u \exp(tX)) du = \frac{d}{dt} \bigg|_{t=0} \int_{U^+} \varphi du = 0$. By the same calculation as in \ref{17}, $\sum_{\gamma \in \Gamma \cap U^+} (X\chi)(\gamma u_0) = 0$, and hence

$$\int_{U^+} (X\chi)(u)\phi(\Gamma u_0) du = 0$$

for any $\phi \in V^\infty$ and those $X \in U(u^+_+)$ with vanishing scalar component. The bounds discussed during the proof of Theorem \ref{1} allow us to evaluate \ref{45} at $x = \Gamma e$. Using \ref{15} and \ref{49}, we
then have
\[ \int_{\Gamma \cap U^+} f(\Gamma u g_t) \, du = \sum_{i=0}^W \int_{\Gamma \cap U^+} (S_{0,i}(t)\rho(Y^i)f)(\Gamma u) \, du. \]

Observe that if \( V \) is an irreducible subrepresentation of \( (\rho, L^2(\Gamma \backslash G)_0) \), each operator \( S_{0,i}(t) \), \( i = 0, \ldots, W \), is of the form \( S_{0,i}(t) = \varphi_i(t) \text{Id} \) for some explicit function \( \varphi_i : \mathbb{R} \to \mathbb{C} \) depending only on the representation \( (\rho, V) \), \( \eta \), and \( g_R \). It should thus be possible to use the decomposition of \( (\rho, L^2(\Gamma \backslash G)_0) \) to give an explicit formula for \( \int_{\Gamma \cap U^+} f(\Gamma u g_t) \, du \) similar to that of \([38, \text{Theorem 1}]\).

**Remark 11.** By using the Cauchy-Schwarz inequality, we have
\[ \int_B |(X_k \chi)(u)|^2 \gamma_t(x u g_t) \, du \leq \| X_k \chi \|^2_{L^2(B)} \left( \int_B \gamma_t^2(x u g_t) \, du \right)^{1/2}. \]

In the case that \( G \) is an arithmetic group of rank greater than or equal to two and \( \Gamma \) is an arithmetic lattice in \( G \), we may use Corollary 17 to go from (45) to give an explicit formula for
\[ \int_{\Gamma \cap U^+} f(\Gamma u g_t) \, du, \]

for all \( \Gamma \neq \emptyset \) such that \( \sum_i \gamma_{\eta}(\Gamma \cap U^+ g_i) \gamma_{\eta}(\Gamma \cap U^+ g_i) = 0 \).

6. **Uniform Bounds**

The goal of this section is to prove Theorem 2. Recall that we have fixed a parabolic subgroup \( P = N A M \) (which, without loss of generality, we may assume to be standard) and a norm \( \| \cdot \| \) on \( a \) that controls the rate of decay of matrix coefficients of \( \overline{A} ^\tau\), i.e. there exists \( q \geq 0 \) such that for all \( K \)-finite vectors \( u, v \in V \) and \( H \in \overline{a} ^\tau\),
\[ |\langle \rho(\exp(H))u, v \rangle| \ll \| u \| \| v \| ((\dim(\rho(K)u) \dim(\rho(K)v)))^{1/2} (1 + \| H \|^q)e^{-\| H \|}. \]

By Lemma 3, the same type of bound holds for almost every irreducible representation \( (\pi, \mathcal{H}) \) occurring in the direct integral decomposition of \( V \). Furthermore, by Lemma 4, we may pass to a corresponding bound for the matrix coefficients of smooth vectors: for almost every \( (\pi, \mathcal{H}) \), there exist \( m \in \mathbb{N}, C, q \geq 0 \) such that
\[ |\langle \pi(\exp(H))u, v \rangle| \leq C \| u \|_{S^m(\mathcal{H}_0)} \| v \|_{S^m(\mathcal{H})}(1 + \| H \|^q)e^{-\| H \|} \quad \forall u, v \in \mathcal{H}^\infty, H \in \overline{a} ^\tau. \]

6.1. **The sets \( a_r \).** In order to prove Theorem 2, we first show that for every \( \epsilon > 0 \), we can find an appropriately large set in \( \overline{a} ^\tau \) such that Theorem 1 holds uniformly for all \( g_R = \exp(\mathbb{R} \cdot H) \), with \( H \) in the given set. In order to do this, we must make explicit the dependency on the subgroup \( g_R \) in Section 3.3 in particular Proposition 10. We now form subsets
\[ \overline{a}^+_r = \left\{ H \in \overline{a} ^\tau : \alpha(H) > 0 \Rightarrow \alpha(H) \geq \epsilon \| H \| \right\}, \quad \overline{a}^+ = \left\{ H \in \overline{a} ^\tau : \| H \| = 1 \right\}. \]
Observe that if $H \in \overline{\mathfrak a}^\perp$, then $\delta H \in \overline{\mathfrak a}^\perp$ for all $\delta \geq 0$. Also, if we denote $\Sigma^+_\mathfrak a(g, a) = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$, then $\overline{\mathfrak a}^\perp$ and $\overline{\mathfrak a}^\perp_{e,1}$ have at most $2^r$ connected components; namely

$$\overline{\mathfrak a}^\perp = \bigcup_{F \subseteq \Sigma^+_\mathfrak a(g, a)} \overline{\mathfrak a}^\perp_{F}, \quad \overline{\mathfrak a}^\perp_{e,1} = \bigcup_{F \subseteq \Sigma^+_\mathfrak a(g, a)} \overline{\mathfrak a}^\perp_{e,1,F},$$

where

$$\overline{\mathfrak a}^\perp_{e,F} = \{H \in \mathfrak a : \alpha(H) \geq \epsilon \forall \alpha \in \Sigma^+_\mathfrak a(g, a) \setminus F \text{ and } \alpha(H) = 0 \forall \alpha \in F\},$$

and

$$\overline{\mathfrak a}^\perp_{e,1,F} = \{H \in \overline{\mathfrak a}^\perp_{e,1} : \|H\| = 1\}.$$

It is easily seen that for any $H \in \overline{\mathfrak a}^\perp$, there exists $\epsilon_H > 0$ such that $H \in \overline{\mathfrak a}^\perp$ for all $\epsilon \leq \epsilon_H$. It is the sets $\overline{\mathfrak a}^\perp_{e,1}$ on which Theorem 1 will be shown to hold uniformly. After proving that every element of $\overline{\mathfrak a}^\perp$ may be decomposed in a good way as the sum of an element of $\overline{\mathfrak a}^\perp$ and another element of $\overline{\mathfrak a}^\perp$, we will proceed to prove generalizations of the lemmas and propositions needed in the proof of Theorem 1.

**Lemma 18.** There exists a constant $c > 0$ depending only on $G$, $\mathfrak a$, and $\|\cdot\|$ such that for any $\epsilon > 0$ and $H \in \overline{\mathfrak a}^\perp$, there exists $H_\epsilon \in \overline{\mathfrak a}^\perp$ such that $J_\epsilon := H - H_\epsilon \in \overline{\mathfrak a}^\perp$ and $\|H_\epsilon\| \geq (1-c\epsilon)\|H\|$.

**Proof.** If $H \in \overline{\mathfrak a}^\perp$, then let $H_\epsilon = H$. We now assume otherwise, and let $\{H^j\}$ denote the basis of $\mathfrak a$ defined by $\alpha_i(H^j) = \delta_{ij}$ (recall that $\Sigma^+_\mathfrak a(g, a) = \{\alpha_1, \ldots, \alpha_r\}$); we may thus write (in a unique way) $H = \sum_j x_j H^j$ with all $x_j \geq 0$. Assume that $\epsilon$ is chosen small enough so that $H^j \in \overline{\mathfrak a}^\perp$ for all $j = 1, \ldots, r$. Let $\|\cdot\|_{\infty}$ be the $\ell^\infty$ norm on $\mathfrak a$ with respect to this basis, i.e.

$$\|Y\|_{\infty} := \max_j \{|\alpha_j(Y)|\}, \quad \forall Y \in \mathfrak a.$$

By equivalence of norms, there exist $D \geq D' > 0$ such that $D'\|Y\|_{\infty} \leq \|Y\| \leq D\|Y\|_{\infty}$ for all $Y \in \mathfrak a$. Given $\epsilon > 0$, we set $\delta = D\epsilon\|H\|_{\infty} > 0$, $F = \{j \in \{1, \ldots, r\} : \alpha_j(H) < \delta\}$, and define

$$H_\epsilon = \sum_{j \in F^c} x_j H^j$$

$$J_\epsilon = \sum_{j \in F} x_j H^j.$$

Note that $J_\epsilon \in \overline{\mathfrak a}^\perp$. For each $j \in F^c$, $\alpha_j(H_\epsilon) = 0$, and for each $j \in F$ we have

$$\alpha_j(H_\epsilon) = \alpha_j(H) \geq \delta = D\epsilon\|H\|_{\infty} \geq D\epsilon\|H_\epsilon\|_{\infty} \geq \epsilon\|H_\epsilon\|,$$

so $H_\epsilon \in \overline{\mathfrak a}^\perp$. Also,

$$\|J_\epsilon\| \leq D\|J_\epsilon\|_{\infty} = D\max\{x_j : j \in F\} < D\delta = D^2\epsilon\|H\|_{\infty} \leq \frac{D^2\epsilon}{D'}\|H\|,$$

so

$$\|H_\epsilon\| \geq \|H\| - \|J_\epsilon\| \geq \left(1 - \frac{D^2\epsilon}{D'}\right)\|H\|,$$

proving that $H_\epsilon$ and $J_\epsilon$ have the desired properties (with $c = \frac{D^2}{D'}$).

**Remark 12.** Throughout the remainder of Section 6, we assume that $\epsilon$ is chosen to be small enough so that $\epsilon < \min\{1, (2c)^{-1}\}$, with $c$ being as in Lemma 18; this ensures that $\frac{\|J_\epsilon\|}{\|H_\epsilon\|} \leq \frac{\epsilon\|H_\epsilon\|}{(1-c\epsilon)\|H\|} < 1$. 

\[ \Box \]
6.2. Continuity of the Harish-Chandra Isomorphism. We now fix a restricted root basis \( \{X_j\}_{j=1,\ldots,d} \) of \( \mathfrak{n} \), with corresponding elements \( \beta_j \in \mathfrak{a}^* \), \( j = 1, \ldots, d \) (so \( \text{ad}_H X_j = \beta_j(H) X_j \) for all \( H \in \mathfrak{a} \)). Note that there may be repetitions in the collection \( \beta_1, \ldots, \beta_d \), and each \( \beta_j \) may be written (in a unique way) as a linear combination of elements of \( \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) = \{\alpha_1, \ldots, \alpha_r\} \) with all coefficients being non-negative integers. Recalling the notation of the proof of Lemma \([5]\) let \( \mathfrak{h}_C = \mathfrak{a}^C \oplus \mathfrak{c} \) be a Cartan subalgebra of \( \mathfrak{g}_C \), with Weyl group \( \mathcal{W} = \mathcal{W}(\mathfrak{g}_C, \mathfrak{h}_C) \). Letting \( W = |\mathcal{W}| \) by the Poincaré-Birkhoff-Witt theorem we may view \( \mathcal{U}^W(\mathfrak{g}_C) \) as a finite dimensional topological vector space (over \( \mathbb{C} \)). For notational purposes, given \( H \in \mathfrak{a} \), we let \( W_H \) denote the integer \( |\mathcal{W}/\text{Stab}_W(H)| \). We will now prove a version of Lemma \([5]\) which takes into account the topological structure on \( \mathcal{U}^W(\mathfrak{g}_C) \):

**Lemma 19.** There exist families of maps \( Z_i : \mathfrak{a} \to \mathcal{Z}(\mathfrak{g}_C) \cap \mathcal{U}^W(\mathfrak{g}_C) \), \( i = 0, 1, \ldots, W \), and \( U_j : \mathfrak{a} \to \mathcal{U}^W(\mathfrak{g}_C) \), \( j = 1, \ldots, d \) which satisfy

\[
\sum_{i=0}^{W} Z_i(H) H^i = \sum_{j=1}^{d} X_j U_j(H) \quad \forall H \in \mathfrak{a},
\]

and additionally have the following properties:

i) \( Z_{W_H}(H) = 1 \) and \( Z_i(H) = 0 \) \( \forall H \in \mathfrak{a}, i > W_H \).

ii) \( U_j(H) = 0 \) \( \forall H \in \overline{\mathfrak{a}^+} \cap \ker \beta_j \).

iii) All the maps \( U_1, \ldots, U_d, Z_0, \ldots, Z_W \) are continuous when restricted to any set \( \overline{\mathfrak{a}^+_F} \).

**Proof.** The construction in the proof Lemma \([5]\) gives, for any given \( H \in \mathfrak{a} \), elements \( Z_{H,j} \) in \( \mathcal{Z}(\mathfrak{g}_C) \) such that

\[
H^{W_H} + Z_{H,W-1} H^{W_H-1} + \ldots + Z_{H,1} H + Z_{H,0} = n \mathcal{U}(\mathfrak{g}_C).
\]

We thus define the functions \( Z_i \) on \( \mathfrak{a} \) in accordance with these constructions:

\[
Z_i(H) = \begin{cases} 
Z_{H,i} & \text{if } i < W_H - 1 \\
1 & \text{if } i = W_H \\
0 & \text{if } i > W_H - 1.
\end{cases}
\]

The fact that the Harish-Chandra isomorphism respects the canonical filtrations on \( \mathcal{Z}(\mathfrak{g}_C) \) and \( \mathcal{U}(\mathfrak{h}_C)^W \), as is clear from the proof of [24, Theorem 5.44], gives that \( Z_i(H) \in \mathcal{U}^W(\mathfrak{g}_C) \) for all \( H \in \mathfrak{a}, i = 0, \ldots, W \). Using \([51]\), we have that there exist elements \( U_{j,H} \in \mathcal{U}^W(\mathfrak{g}_C) \) such that

\[
\sum_{i=0}^{W} Z_i(H) H^i = \sum_{j=1}^{d} X_j U_{j,H},
\]

and the functions \( Z_i \) satisfy i). We thus need to use the elements \( U_{j,H} \) to define functions \( U_j \) that satisfy \([50]\), ii) and iii). To any \( H \in \overline{\mathfrak{a}^+} \) we may associate a set \( \mathcal{F} \subset \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \) consisting of all \( \alpha \in \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \) such that \( \alpha(H) = 0 \). Every \( \beta \in \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) for which \( \beta(H) = 0 \) may be thus be written as a linear combination consisting solely of elements of \( \mathcal{F} \) (since each \( \beta \in \Sigma^+(\mathfrak{g}, \mathfrak{a}) \) may be written in a unique way as a linear combination of elements of \( \Sigma_0^+(\mathfrak{g}, \mathfrak{a}) \) with all coefficients being non-negative integers). We may now construct a standard parabolic subgroup \( P_F = N_F A_F M_F \supset P \), with Lie algebra \( \mathfrak{n}_F \oplus \mathfrak{a}_F \oplus \mathfrak{m}_F \) such that \( H \in \mathfrak{a}_F^+ \) and a basis for \( \mathfrak{n}_F \) is given by the elements \( X_j \) for which \( \beta_j(H) \neq 0 \). Since \( \mathfrak{a}_F \subset \mathfrak{a} \subset \mathfrak{h}_C \) and \( \mathfrak{n}_F \subset \mathfrak{n} \), applying the construction in the proof of Lemma \([5]\) to \( H \) with \( P_F \) in place of \( P \) gives rise to the same elements \( Z_{H,i} \) and so

\[
\sum_{i=0}^{W} Z_i(H) H^i = \sum_{j=1}^{d} X_j U_{j,H,F} \in \mathfrak{n}_F \mathcal{U}(\mathfrak{g}_C),
\]

where \( U_{j,H,F} = 0 \) for all \( j \) such that \( X_j \not\in \mathfrak{n}_F \), i.e. those \( X_j \) such that \( \beta_j(H) = 0 \). We thus construct functions \( U_j \) in the following manner: for each subset \( \mathcal{F} \subset \Sigma_0^+(\mathfrak{g}, \mathfrak{a}_0) \), we choose a
Poincaré-Birkhoff-Witt basis of \( g \) with the first elements of this basis being the \( X_j \)'s that form a basis of \( \mathfrak{n}_\mathbb{F} \). Then, for each \( H \in \mathfrak{a} \), we choose the unique maximal subset \( \mathcal{F} \subset \Sigma_{0}^{\mathfrak{a}}(g, a_0) \) such that \( H \in \mathfrak{a}_\mathcal{F} \). We then have that

\[
\sum_{i=0}^{W} Z_i(H)H^i = \sum_{j=1}^{d} X_j U_{j,H,\mathcal{F}} \in \mathfrak{n}_\mathbb{F}\mathcal{U}(g_c),
\]

where the elements \( U_{j,H,\mathcal{F}} \) are uniquely determined by the choice of basis, and \( U_{j,H,\mathcal{F}} = 0 \) if \( X_j \notin \mathfrak{n}_\mathcal{F} \). The functions \( U_j \) are then defined by \( U_j(H) = U_{j,H,\mathcal{F}} \). This choice of functions \( U_j \) will satisfy \( ii \), and (together with the functions \( Z_i \)) \([50]\).

Finally, in order to prove \( iii \), we first note that if \( H_1 \) and \( H_2 \) both belong to some set \( \mathfrak{a}_\mathcal{F} \), then \( \text{Stab}_{\mathcal{F}}(H_1) = \text{Stab}_{\mathcal{F}}(H_2) \), indeed, by Chevalley’s lemma \([24], \text{Proposition 2.72}\), this set is determined by the choice of \( \mathcal{F} \subset \Sigma_{0}^{\mathfrak{a}}(g, a) \). This implies that the coefficients of the polynomial \([16]\) depend continuously on \( H \) as \( H \) is allowed to vary in \( \mathfrak{a}_\mathcal{F} \). More precisely, if

\[
\sum_{i=0}^{W} Z_i(H)H^i = \sum_{j=1}^{d} X_j U_{j,H,\mathcal{F}} \in \mathfrak{n}_\mathbb{F}\mathcal{U}(g_c),
\]

then the maps \( H \mapsto J_{H,i} \) are continuous maps from \( \mathfrak{a}_\mathcal{F} \) to \( \mathcal{U}(g_c) \). The Harish-Chandra isomorphism \( \gamma|_{\mathcal{Z}(g_c)\cap\mathcal{U}(g_c)} \) is an invertible linear map between two subspaces of \( \mathcal{U}(g_c) \), \( \gamma|_{\mathcal{Z}(g_c)\cap\mathcal{U}(g_c)} \) is a continuous linear map from \( \mathcal{U}(g_c) \) to \( \mathcal{U}(g_c) \). Applying this to the right-hand side of \((52)\) gives that for each \( i \), \( H \mapsto \gamma^{-1}(J_{H,i}) = Z_i(H) \) is continuous on \( \mathfrak{a}_\mathcal{F} \). Since the sum \( \sum_{i=0}^{W} Z_i(H)H^i \) depends continuously on \( H \in \mathfrak{a}_\mathcal{F} \), \( \sum_{j=1}^{d} X_j U_j(H) \) must as well. Now using the fact that the coordinates in our Poincaré-Birkhoff-Witt basis are continuous functions on \( \mathcal{U}(g_c) \), we conclude that the maps \( U_j|_{\mathfrak{a}_\mathcal{F}} \) are also continuous. \( \square \)

### 6.3. Differential Equations

Following Section 3.2 for an irreducible unitary representation \((\pi, \mathcal{H})\) for which \( \| \cdot \| \) controls the rate of decay of matrix coefficients of \( \Lambda^+ \), we define

\[
\mathcal{T}_\mathbf{v}(H, t) = \int_N \chi(n)\pi(n \exp(tH)) v \, dn,
\]

where \( v \in \mathbb{H}^\infty \), \( \chi \in C_c^\infty(N) \), \( t \leq 0 \) and \( H \in \mathfrak{a}_\mathcal{F} \). Furthermore (as previously), by Schur’s lemma, there exist functions \( a_i : \mathfrak{a} \rightarrow \mathbb{C} \) such that \( d\pi(Z_i(H))w = a_i(H)w \) for all \( w \in \mathbb{H}^\infty \), \( i = 0, \ldots, W \) \((Z_i(H)\) being as in Lemma \([19]\)\). From Lemma \([19]\) we have

\[
\sum_{i=0}^{W} a_i(H)d\pi(H^i)w = \sum_{j=1}^{d} d\pi(X_j U_j(H))w \quad \forall w \in \mathbb{H}^\infty.
\]

Note that \( a_W(H) = 0 \) and \( a_i(H) = 0 \) for \( i > W_H \). Similarly to Section 3.3 for each \( H \in \mathfrak{a} \) we let \( \Lambda(H) \) denote the multi-set (of order \( W_H \)) of roots to the polynomial

\[
z^{W_H} + a_{W_H-1}(H)z^{W_H-1} + \ldots + a_1(H)z + a_0(H).
\]

Analogously to \((22)\), we have

\[
\prod_{\lambda \in \Lambda(H)} \left( \frac{d}{dt} - \lambda \right)^{\mathcal{T}_\mathbf{v}(H, t)} = \sum_{j=1}^{d} -e^{\hat{\beta}(H) t} \mathcal{T}_{d\pi(U_j(H))w}(H, t).
\]

By Lemma \([19]\) the right-hand side of this equation satisfies a bound \( \ll e^{\alpha(H) t} \) for \( t \leq 0 \), where

\[
\alpha_0(H) = \min\{\alpha(H) : \alpha \in \Sigma_{0}^{\mathfrak{a}}(g, a) \text{ such that } \alpha(H) > 0\}.
\]

This bound will enable us to use Lemma \([7]\) and Proposition \([8]\) uniformly on the sets \( \overline{\mathfrak{a}_\mathcal{F}} \). Letting \( \|\Lambda(H)\|_\infty = \max_{\lambda \in \Lambda(H)} |\lambda| \), we have the following lemma, which gives a “uniform” version of \((43)\):
Lemma 20. There exists a function $\Psi : a \rightarrow \mathbb{R}_{>0}$ such that for all $H \in a$, $w \in \mathcal{H}^\infty$, $m \in \mathbb{N}$,
\[|\Lambda(H)|_{\mathcal{H}^\infty}^{W_H} \|w\|_{S^m(\mathcal{H})} \ll_m \Psi(H) \|w\|_{S^{2W+m}(\mathcal{H})}.
\]
Moreover, $\Psi$ is continuous on each connected component of $\mathfrak{a}^\perp$.

Proof. Arguing as in the proof of Corollary 9 by Fujiwara’s bound [2], we have
\[|\Lambda(H)|_{\mathcal{H}^\infty}^{W_H} \|w\|_{S^m(\mathcal{H})} \ll \max_{i=0,\ldots,W_H-1} |\alpha_i(H)|_{\mathcal{H}^\infty}^{W_H/(W_H-1)} \|w\|_{S^m(\mathcal{H})}
\]
\[\ll \max_{i=0,\ldots,W_H-1} |w\|_{S^m(\mathcal{H})} + \left\| d\pi \left( Z_i(H)^{W_H/(W_H-1)} \right) w \right\|_{S^m(\mathcal{H})}.
\]
Once again using the fact that the Harish-Chandra isomorphism preserves the canonical filtrations, we have $Z_i(H) \in \mathcal{U}^{W_H-i} (\mathfrak{g}_C)$ (cf. 19), hence $Z_i(H)^{W_H/(W_H-1)} \in \mathcal{U}^{2W_H-i} (\mathfrak{g}_C) \subset \mathcal{U}^{2W} (\mathfrak{g}_C)$. The continuity (on each connected component of $\mathfrak{a}^\perp$) of the maps $H \mapsto Z_i(H)$ completes the proof. \hfill \Box

6.4. Burger’s Formula. The aim of this section is to make Proposition 10 uniform on sets $\mathfrak{a}_{\epsilon,1}^\perp$. We adapt the multi-index notation introduced prior to Proposition 10 to the present situation: for any multi-index $j = (j_1, j_2, \ldots, j_n)$, we define $X_j = X_{j_1} X_{j_2} \cdots X_{j_n}$, and $U_j(H) = U_{j_1}(H) U_{j_2}(H) \cdots U_{j_n}(H)$, where $X_j$ and $U_j(H)$ are as in Lemma 19. We may now state a version of Proposition 10 that is uniform over the sets $\mathfrak{a}_{\epsilon,1}^\perp$.

Proposition 21. Given an irreducible unitary representation $(\pi, \mathcal{H})$, a norm $\| \cdot \|$ on $\mathfrak{a}^\perp$ that controls the matrix coefficients of $\overline{A}^\perp$ in $(\pi, \mathcal{H})$, and $\epsilon > 0$, there exist functions $C$ such that for all $\Psi \in \mathcal{H}^\infty$, $\chi \in C^\infty_c (N)$, $H \in \mathfrak{a}_{\epsilon,1}^\perp$, and $t \leq 0$:
\[I_\Psi^\chi(H, t) = \sum_{k \in \mathfrak{I}_{\mathfrak{a}_{\epsilon,1}^\perp}} \int_{-\infty}^0 C_k(H, t, s) I_{X_k}^{\Psi \chi} (d\pi(k(t))) \omega(H, s) ds + \sum_{j \in \mathfrak{I}_{\mathfrak{a}_{\epsilon,1}^\perp}} \sum_{i=0}^{W} \mathcal{D}_{j,i}(H, t) I_{X_j}^{\Psi \chi} (d\pi(U_j(H))) \omega(H, 0).
\]
Furthermore, the following hold:

(1) $|C_k(H, t, s)| \ll (1 + |t|^W) e^{t + \Psi s}$, where the implied constant depends only on $G$, $\| \cdot \|$, and $\epsilon$.

(2) For any $m \in \mathbb{N}$,
\[|\mathcal{D}_{j,i}(H, t)| \ll (1 + |t|^W) e^t \inf_{\omega \in \mathcal{H}^\infty} \|\omega\|_{S^{2W+m}(\mathcal{H})} \|\omega\|_{S^m(\mathcal{H})},
\]
where the implied constant depends only on $G$, $m$, $\| \cdot \|$, and $\epsilon$.

Proof. For all $H \in \mathfrak{a}_{\epsilon,1}^\perp$, we have $\alpha_0(H) \geq \epsilon$. By letting $\alpha = \epsilon$ and $\eta = 1$, the construction in the proof of Proposition 10 works for all $H \in \mathfrak{a}_{\epsilon,1}^\perp$. Note that we have assumed that $\epsilon < 1$ (cf. Remark 12), so $\alpha < \eta$; it is therefore this case that is used from Proposition 10. Also, even though $N$ is not the expanding horospherical subgroup associated to those $H$ that lie on one or more of the walls of $\mathfrak{a}^\perp$, the fact that $U_j(H) = 0$ for all $j$ such that $\beta_j(H) = 0$ permits the use of the same proof for such $H$, since we have the bound
\[\left\| - \sum_{j=1}^d e^{\beta_j(H)t} I_{X_j}^{\Psi \chi} (d\pi(U_j(H))) \omega(H, t) \right\| \ll \|\chi\|_{W^{d+1}(N)} \|\omega\|_{S^d(\mathcal{H})} e^{\epsilon t} \quad \forall H \in \mathfrak{a}_{\epsilon,1}^\perp, t \leq 0.
\]
Using Lemma 20 instead of (31) throughout the proof replaces the bound in Proposition 10 (2) with the uniform bound stated here (i.e. property (2)); in particular, the dependency of the implied constant on $g_R$ in Proposition 10 (2) is removed. \hfill \Box
6.5. The Invariant Height Function. The final result needed before proceeding to the proof of Theorem 2 is the following generalization of Corollary 16.

Lemma 22. Let $B \subset N$ be a relatively compact subset of positive measure. Then

$$\int_B Y_T (xn \exp(-H)) \, dn \ll_{\Gamma, B} \mathcal{Y}^2_T (x) \quad \forall x \in \Gamma \setminus G, \ H \in \bar{a}^\tau.$$

Proof. The proof is essentially the same as that of Corollary 16; given a compact neighbourhood $B_0$ of $e$ in $AMN^*$, we let $\tilde{B}_0$ be the closure of $\bigcup_{H \in \bar{a}^\tau} \exp(H)B_0 \exp(-H)$. The set $\tilde{B}_0$ is compact, allowing us to use the same argument as previously to bound the integral. \( \square \)

6.6. Proof of Theorem 2. Given $H \in \bar{a}^\tau$, we use Lemma 18 to find $H'_e$ and $J'_e$ such that $H = H'_e + J'_e$, with $H'_e \in \bar{a}^\tau$, $\|H'_e\| \geq (1 - \epsilon\|H\|$, and $J'_e \in \bar{a}^\tau$. Observe that if $\frac{\alpha H}{\|H\|} \geq \epsilon$, then $H'_e = H$. We now let $H_e = \frac{H'_e}{\|H'_e\|} \in \bar{a}^\tau_1$ and $J_e = \frac{J'_e}{\|J'_e\|} \in \bar{a}^\tau$. As in the proof of Theorem 1 we first make the reduction to the case when $f \in V^\infty$ and $\chi \in C_c^\infty(B)$. We then have that for $t \leq 0$,

$$\mathcal{I}(H_e + J_e, t) = \mathcal{I}(H_e, t) + \int_0^1 \frac{d}{dt} \mathcal{I}(H_e + \tau J_e, t) \, d\tau$$

$$= \mathcal{I}(H_e, t) - \int_0^1 \int_N \chi(n) \frac{d}{d\tau} \rho(n \exp(tH_e + \tau J_e)) \, f \, dn \, d\tau$$

$$= \mathcal{I}(H_e, t) - \int_0^1 \int_N \chi(n) \rho(n \exp(tH_e + \tau J_e)) \, d\rho(J_e) \, f \, dn \, d\tau.$$

The $\tau$-integral is rewritten in the following manner:

$$\int_0^1 \int_N \chi(n) \rho(n \exp(tH_e + \tau J_e)) \, d\rho(J_e) \, f \, dn \, d\tau$$

$$= \int_0^1 \rho(\exp(\tau J_e)) \int_N \chi(n) \rho(\exp(-\tau J_e)) n \exp(\tau J_e) \rho(\exp(tH_e)) \, d\rho(J_e) \, f \, dn \, d\tau$$

$$= \int_0^1 e^{2\tau \rho(J_e)} \rho(\exp(\tau J_e)) \int_N \chi(\exp(\tau J_e) n \exp(-\tau J_e)) \rho(n \exp(tH_e)) \, d\rho(J_e) \, f \, dn \, d\tau$$

$$= \int_0^1 e^{2\tau \rho(J_e)} \rho(\exp(\tau J_e)) \mathcal{I}^{\chi}_{\rho(J_e)}(H_e, t) \, d\tau,$$

where, for any $\psi \in C_c^\infty(N)$,

$$\psi_\tau(n) := \psi(\exp(\tau J_e) n \exp(-\tau J_e)).$$

This is now entered into the original equation, giving

$$(54) \quad \mathcal{I}(H_e + J_e, t) = \mathcal{I}(H_e, t) - \int_0^1 e^{2\tau \rho(J_e)} \rho(\exp(\tau J_e)) \mathcal{I}^{\chi}_{\rho(J_e)}(H_e, t) \, d\tau.$$

We will now use Proposition 21 in the same way that Proposition 10 was used in the proof of Theorem 1; let $\{T_k(H_e, t, s)\}$, $\{S_{j, k}(H_e, t)\}$ be the intertwining operators defined on $V^\infty$ by (for $\phi \in V^\infty$)

$$T_k(H_e, t, s) \phi = \int_Z C_{\zeta, k}(H_e, t, s) \phi_\zeta \, dv(\zeta)$$

$$S_{j, k}(H_e, t) \phi = \int_Z D_{\zeta, j, k}(H_e, t) \phi_\zeta \, dv(\zeta).$$

Since $H_e$ will be fixed for the remainder of the proof, for notational convenience we temporarily suppress the argument in the functions $U_k$; that is $U_k = U_k(H_e)$. Analogously to the proof of
Theorem [1] by Proposition [21] these intertwining operators may be used to write [54] as

\[ T_f^X(H_e + J_e, t) = \sum_{k \in \mathfrak{m}_2} \int_{-\infty}^{0} T_k(H_e, t, s) T_{X_k\chi_r}^{X_k\chi_r}(H_e, s) \, ds + \sum_{j \in \mathfrak{m}_{2-1}} \sum_{i=0}^{W} S_{j,i}(H_e, t) T_{\text{d}_d(H^1U_j)f}(H_e, 0) \]

\[ - \sum_{k \in \mathfrak{m}_2} \int_{t}^{0} \int_{-\infty}^{0} e^{2x_{\rho_k}(J_e)} \rho(\exp(\tau J_e)) T_k(H_e, t, s) T_{X_k\chi_r}^{X_k\chi_r}(H_e, s) \, ds \, d\tau \]

\[ - \sum_{j \in \mathfrak{m}_{2-1}} \sum_{i=0}^{W} \int_{t}^{0} e^{2x_{\rho_k}(J_e)} \rho(\exp(\tau J_e)) S_{j,i}(H_e, t) T_{\text{d}_d(H^1U_j)f}(H_e, 0) \, d\tau, \]

where \( m_2 = \frac{1}{2} e \). For any element \( X_j \) of our chosen root basis of \( n \), we have

\[ (X_j\chi_r)(n) = \frac{d}{dv} \bigg|_{v=0} \chi_r(n \exp(vX_{\mu}^j)) = \frac{d}{dv} \bigg|_{v=0} \chi(\exp(\tau J_e)n \exp(vX_j) \exp(-\tau J_e)) \]

\[ = e^{\tau_\beta_j(J_e)}(X_j\chi_r)(n) \]

Hence, for any multi-index \( k = (k_1, \ldots, k_l) \),

\[ (X_k\chi_r)(n) = e^{\tau(\beta_{k_1} + \cdots + \beta_{k_l})(J_e)}(X_k\chi_r)(n). \]

Since \( J_e \in \mathfrak{a}^r \), \( \beta_{J_e} \geq 0 \) for all \( j = 1, \ldots, d \), giving \( e^{\tau_\beta_j(J_e)} \leq 1 \) (\( \tau \leq 0 \)), and so

\[ \|X_j\chi_r\| = \|e^{\tau_\beta_j(J_e)}(X_j\chi_r)\| = \|e^{\tau_\beta_j(J_e)}(X_j\chi_r)(\exp(\tau J_e)n \exp(-\tau J_e))\| \leq \|X_j\chi_r\|_{L^\infty(N)}. \]

Hence, for any multi-index \( k \) and \( \tau \leq 0 \),

\[ \|X_k\chi_r\|_{L^\infty(N)} = \|e^{\tau(\beta_{k_1} + \cdots + \beta_{k_l})(J_e)}(X_k\chi_r)\|_{L^\infty(N)} \leq \|(X_k\chi_r)\|_{L^\infty(N)} = \|X_k\chi_r\|_{L^\infty(N)}. \]

We now consider the terms occurring in the last two lines of [55]. To start with, for \( k = (k_1, \ldots, k_{m_2}) \in \mathfrak{m}_{m_2} \), we have

\[ \int_{t}^{0} \int_{-\infty}^{0} e^{2x_{\rho_k}(J_e)} \rho(\exp(\tau J_e)) T_k(H_e, t, s) T_{X_k\chi_r}^{X_k\chi_r}(H_e, s) \, ds \, d\tau \]

\[ = \int_{t}^{0} \int_{-\infty}^{0} \int_{N} e^{2x_{\rho_k}(J_e)} e^{\tau(\beta_{k_1} + \cdots + \beta_{k_{m_2}})(J_e)}(X_k\chi_r)(n) \]

\[ \times \rho(\exp(\tau J_e)n \exp(sH_e)) T_k(H_e, t, s) d\rho(U_k J_e) f \, dn \, ds \, d\tau \]

\[ = \int_{t}^{0} \int_{-\infty}^{0} \int_{N} e^{\tau(\beta_{k_1} + \cdots + \beta_{k_{m_2}})(J_e)}(X_k\chi_r)(n) \rho(n \exp(\tau J_e + sH_e)) T_k(H_e, t, s) d\rho(U_k J_e) f \, dn \, ds \, d\tau. \]

Taking into account the fact that supp(\( \chi \)) \( \subset B \), the bound given in Proposition [21] (1), Proposition [12] and Lemma [22] (note that \( -\tau J_e - sH_e \in \mathfrak{a}^r \) for all \( \tau, s \leq 0 \)), we have the following bound for \( m > \frac{\dim G}{2} \) and all \( x \in \Gamma \setminus G \):

\[ \int_{N} \left| e^{\tau(\beta_{k_1} + \cdots + \beta_{k_{m_2}})(J_e)}(X_k\chi_r)(n) T_k(H_e, t, s) d\rho(U_k(H_e) J_e) f(xn \exp(\tau J_e + sH_e)) \right| \, dn \]

\[ \ll e^{\tau(\beta_{k_1} + \cdots + \beta_{k_{m_2}})(J_e)} \|X_k\chi_r\|_{L^\infty(B)} \|d\rho(U_k(H_e) J_e) f\|_{S^m(\Gamma \setminus G)} \|T(x)^2 + |t|^W\| e^{t + |t|^W}. \]

Since \( \|J_e\| < 1 \) (cf. Remark [12]), \( \|d\rho(J_e)\|_{L^\infty(\Gamma \setminus G)} \ll \|\phi\|_{S^m(\Gamma \setminus G)} \) for all \( \phi \in V^\infty \). This, together with [50] and Lemma [19] (in particular, the continuity of \( Y \mapsto U_J(Y) \) on each connected
component of $\alpha_{+1}$ gives

$$
\int_{0}^{1} \int_{-\infty}^{0} \int_{N} e^{(\tau_{+1} + \ldots + \tau_{+m})(J)} (X_{\chi}(t)(n)T_{k}(H_{t}, t, s) \rho(U_{k}(H_{t})J)) f(x \exp(\tau_{J_{t}} + sH_{t})) \, dn \, ds \, d\tau
$$

$$
\ll \| \chi \|_{W_{m}^{2, \infty}(B)} \| f \|_{S_{W_{m}^{2, \infty+1+m}(\Gamma \backslash G)}} \bar{Y}_{\Gamma}(x)^{2}(1 + |t|)^{W+1})e^{t},
$$

Similar arguments are used for the other terms of (55) that have an integral term $\int_{0}^{1} d\tau$, i.e. for $j = (j_{1}, \ldots, j_{m}) \in \mathcal{I}_{\leq m-1}$ and $i \in \{0, \ldots, W\}$,

$$
\int_{0}^{1} \int_{N} e^{(\tau_{+1} + \ldots + \tau_{+m})(J)} \left( \rho^{(\chi)}(\tau_{J_{t}})S_{j_{1}}(H_{t}, t) \bar{X}_{\chi}(\bar{X}_{j}) \rho(U_{j_{1}}(H_{t})J_{t}) f(x \exp(\tau_{J_{t}})) \right) \, dn \, d\tau
$$

$$
\ll \| \chi \|_{W_{m}^{2, \infty}(B)} \| f \|_{S_{W_{m}^{2, \infty+1+m}(\Gamma \backslash G)}} \bar{Y}_{\Gamma}(x)^{2}(1 + |t|)^{W+1})e^{t},
$$

where Proposition $\text{(21)} (2)$ was now used. The remaining terms in (55) are dealt with in same way as in the proof of Theorem $\text{[1]}$. Putting these bounds together, we have

$$
\left| \int_{B} \chi(n)(x \exp(\tau_{H_{t}} + J_{t})) \, dn \right| \ll \| \chi \|_{W_{m}^{2, \infty}(B)} \| f \|_{S_{W_{m}^{2, \infty+1+m}(\Gamma \backslash G)}} \bar{Y}_{\Gamma}(x)^{2}(1 + |t|)^{W+1})e^{t}.
$$

Letting $t = -\| H_{t} \|$ (recall that $H = \| H_{t} \|(H_{t} + J_{t})$ and $(1 - \epsilon)\| H \| \leq \| H_{t} \| \leq (1 + \epsilon)\| H \|$) now completes the proof.

\[
\square
\]

References

[1] J. Bernstein and A. Reznikov, Analytic Continuation of Representations and Estimates of Automorphic Forms, Ann. of Math., Vol. 150 (1999), pp. 329-352.
[2] A. Borel and Harish-Chandra, Arithmetic subgroups of algebraic groups, Ann. of Math., Vol. 75 (1962), pp. 485-535.
[3] M. Burger, Horocycle flow on geometrically finite surfaces, Duke Math. J., Vol. 61, (1990), pp. 779-803.
[4] V. Casselman and D. Milicic, Asymptotic behavior of matrix coefficients of admissible representations, Duke Math. J., Vol. 49 (1982), pp. 869-930.
[5] L. Corwin and F. Greenleaf, Representations of Nilpotent Lie Groups and their applications, Part 1: Basic Theory and Examples, Cambridge University Press, 1989.
[6] M. Cowling, Sur les coefficients des representations unitaires des groupes de Lie simple, Springer-Verlag, Lecture Notes in Mathematics, Vol. 739 (1979), pp. 132-178.
[7] K. Dabbs, M. Kelly, and H. Li, Effective equidistribution of maximal horospherical measures in the space of lattices, J. Mod. Dyn., Vol. 10 (2016), pp. 229-254.
[8] S. C. Edwards, On the rate of equidistribution of expanding horospheres in finite-volume quotients of SL(2,C), preprint (2015), arXiv: 1512.02373.
[9] J. Fell, The Dual Spaces of $C^\ast$-algebras, T.A.M.S., Vol. 94 (1960), pp. 364-403.
[10] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows, Duke Math. J., Vol. 119 (2003), pp. 465-526.
[11] M. Fujikawa, Uber die obere Schranke des absoluten Betrages der Wurzeln einer algebraischen Gleichung, Tôhoku Math J. Vol. 10 (1916), pp. 167-171.
[12] H. Garland and M. S. Raghunathan, Fundamental Domains for Lattices in (R-)rank 1 Semisimple Lie Groups, Ann. of Math., Vol. 92 (1970), pp. 279-326.
[13] D. Hejhal, On the uniform equidistribution of long closed horocycles in Loo-Keng Hua: A great mathematician of the twentieth century, Asian J. Math., Vol. 4 (2000), pp. 839-853.
[14] R. Howe, On A Notion of Rank for Unitary Representations of the Classical Groups, Harmonic Analysis and Group Representation, C.I.M.E. Summer Schools, Vol. 82 (1982), pp. 224-331.
[15] L. Ji, Metric compactifications of locally symmetric spaces, Internat. J. Math. 9, 1998, pp. 465-491.
[16] A. Katok and R. Spatzier, First Cohomology of Anosov Actions of Higher Rank Abelian Groups And Applications to Rigidity, Publications Mathématiques de L’Institut des Hautes Scientifiques, Vol. 79 (1994), pp. 131-156.
[17] I. Kim, Counting, mixing and equidistribution of horospheres in geometrically finite rank one locally symmetric manifolds, J. reine angew. Mathematik, Vol. 704 (2015), pp. 85-133.
[18] A. Kontorovich and H. Oh, Apollonian circle packings and closed horospheres on hyperbolic 3-manifolds, With an appendix by Oh and Nimish Shah, J. Amer. Math. Soc., Vol. 24 (2011), pp. 603-648.
[19] D. Kleinbock and G. A. Margulis, Bounded orbits of nonquasiunipotent flows on homogeneous spaces, Sinai’s Moscow Seminar on Dynamical Systems, Amer. Math. Soc. Transl. Ser. 2, 171, Amer. Math. Soc., Providence, RI (1996), pp. 141-172.

[20] D. Kleinbock and G. A. Margulis, On effective equidistribution of expanding translates of certain orbits in the space of lattices, Number Theory, Analysis and Geometry, Springer, New York, 2012, pp. 385-396.

[21] D. Kleinbock, R. Shi, and B. Weiss, Pointwise equidistribution with an error rate and with respect to unbounded functions, Math. Ann. (2016), doi:10.1007/s00208-016-1404-3.

[22] D. Kleinbock and B. Weiss, Dirichlet’s theorem on Diophantine approximation and homogeneous flows, J. Mod. Dyn., Vol. 2 (2008), pp. 43-62.

[23] A. Knapp, Representation Theory of Semisimple Lie Groups, Princeton University Press, 1986.

[24] A. Knapp, Lie Groups Beyond an Introduction Second Edition, Birkhäuser, 2002.

[25] D. Kleinbock and G. A. Margulis, On some aspects of the theory of Anosov systems, with a survey, “Periodic orbits of hyperbolic flows” by Richard Sharp, Springer Monographs in Mathematics, Springer, Berlin, 2004.

[26] J. Marklof, Spectral form factors of rectangle billiards, Comm. Math. Phys., Vol. 190 (1998), pp. 169-202.

[27] J. Marklof, The n-point correlations between values of a linear form, with an appendix by Z. Rudnick, Ergod. Th. & Dynam. Sys., Vol. 20 (2000), pp. 1127-1172.

[28] A. Mohammadi and H. Oh, Matrix coefficients, counting and primes for orbits of geometrically finite groups, J. Eur. Math. Soc., Vol. 17 (2015), pp. 837-897.

[29] H. Oh, Uniform Pointwise Bounds for Matrix Coefficients of Unitary Representations and Applications to Kazhdan Constants, Duke Math. J., Vol. 113 (2002), pp. 133-192.

[30] M. Shah and B. Weiss, On the Deviation of Ergodic Averages for Horocycle Flows, J. Mod. Dyn., Vol. 7 (2013), pp. 291-328.