A HAAR COMPONENT FOR QUANTUM LIMITS ON LOCALLY SYMMETRIC SPACES

NALINI ANANTHARAMAN AND LIOR SILBERMAN

Abstract. We prove lower bounds for the entropy of limit measures associated to non-degenerate sequences of eigenfunctions on locally symmetric spaces of non-positive curvature. In the case of certain compact quotients of the space of positive definite $n \times n$ matrices (any quotient for $n = 3$, quotients associated to inner forms in general), measure classification results then show that the limit measures must have a Haar component. This is consistent with the conjecture that the limit measures are absolutely continuous.

Contents

1. Introduction

1.1. **Background and motivations.** The study of high-energy Laplacian eigenfunctions on negatively curved manifolds has progressed considerably in recent years. In the so-called "arithmetic" case, Elon Lindenstrauss has proved the Quantum Unique Ergodicity conjecture for Hecke eigenfunctions on congruence quotients of the hyperbolic plane [16] (for non-compact congruence quotients his methods left open the possibility of "escape of mass", and the matter was finally resolved by Soundararajan in [25]). In greater generality (variable negative curvature, no arithmetic structure), the first author has proved (partly joint with Stéphane Nonnenmacher) that semiclassical limits of eigenfunctions have positive Kolmogorov-Sinai entropy [1, 3, 4].

The two approaches are very different, but have in common the central role of the notion of entropy. In Lindenstrauss' work, an entropy bound is obtained from arithmetic

N.A. was partly supported by the grant ANR-05-JCJC-0107-01. She is also grateful to the Miller Institute for Basic Research in Science at the University of California Berkeley, for supporting her work in the spring of 2009.

L.S. was partly supported by an NSERC Discovery Grant.

considerations [5], and then combined with the measure rigidity phenomenon to prove Quantum Unique Ergodicity.

It is very natural to ask about a possible generalization of these results to locally symmetric spaces of higher rank and non-positive curvature. In this case the Laplacian will be replaced by the entire algebra of translation-invariant differential operators, as proposed by Silberman and Venkatesh in [23]. A generalization of the entropic bound of [5] has been worked out by these authors in the adelic case, and as a result they could prove a form of Arithmetic Quantum Unique Ergodicity in the case of the locally symmetric space $\Gamma\backslash SL_n(\mathbb{R})$, when n is prime and Γ is a lattice derived from a division algebra over $\mathbb{Q}$ [24]. The goal of this paper is to generalize the "non-arithmetic" approach of [3, 4] in this context – that is to say, prove an entropy bound without using the Hecke operators or other arithmetic techniques. Doing so, we will not require all the assumptions used in [24]: we will work with an arbitrary connected semisimple Lie group with finite center G, Γ will be any *cocompact* lattice in G, and we will not use the Hecke operators. Combining the entropy bound with the measure classification results of [8, 9, 17], in the case of $G = SL_3(\mathbb{R})$, Γ arbitrary, or $G = SL_n(\mathbb{R})$, n arbitrary but Γ derived from a division algebra over $\mathbb{Q}$, we will prove a weakened form of Quantum Unique Ergodicity : any semiclassical measure has the Haar measure as an ergodic component[1].

In addition to the intrinsic interest of locally symmetric spaces, there is another motivation to study these models. So far, the entropic bound of [3, 4] is not fully satisfactory for manifolds of *variable* negative sectional curvature ([1] proves that the entropy is positive, but without giving an explicit bound). Gabriel Rivière has been able to treat the case of surfaces [19, 20]; he is even able to work in non-positive curvature, but the case of higher dimensions remains open. The problem comes from the existence of several distinct Lyapunov exponents for the geodesic flow. Understanding the case of negatively curved locally symmetric spaces constitutes a progress in this direction : we will deal with dynamical systems that have distinct Lyapunov exponents, some of which may even vanish. Still, considerable simplifications arise from the fact that the space is homogeneous, and that the stable and unstable foliations are smooth. It would be interesting but extremely challenging to extend the techniques of [3, 4, 19, 20] to systems that are non-uniformly hyperbolic (euclidean billiards would be the ultimate goal).

Let G be a connected semisimple Lie group with finite center, $K < G$ be a maximal compact subgroup, $\Gamma < G$ a uniform lattice. We will work on the symmetric space $\mathbf{S} = G/K$, the compact quotient[2] $\mathbf{Y} = \Gamma\backslash G/K$, and the homogeneous space $\mathbf{X} = \Gamma\backslash G$. We will endow G with its Killing metric, yielding a G-invariant Riemannian metric on G/K, with non-positive curvature.

Call $\mathcal{D}$ the algebra of G-invariant differential operators on $\mathbf{S}$; it follows from the structure of semisimple Lie algebras that this algebra is commutative and finitely generated [11, Ch.

[1]Unfortunately, we are not able to extend the method to the case of $\Gamma = SL_n(\mathbb{Z})$, which is not cocompact – unless we input the extra assumption that there is no escape of mass to infinity, or that the mass escapes very fast.

[2]We do not assume that Γ is torsion free. When speaking of smooth functions on $\mathbf{Y}$, we have in mind smooth functions on $\mathbf{S}$ that are Γ-invariant.

II §4.1, §5.2]. The number of generators, to be denoted by r, coincides with the real rank of $\mathbf{S}$ (that is the dimension of a maximal flat totally geodesic submanifold), and, in a more algebraic fashion, with the dimension of $\mathfrak{a}$, a maximal abelian semisimple subalgebra[3] of $\mathfrak{g}$ orthogonal to $\mathfrak{k}$. More background and notation concerning Lie groups are given in Section 2.

Remark 1.1. The algebra $\mathcal{D}$ always contains the Laplacian. If the symmetric space $\mathbf{S}$ has rank $r = 1$, then $\mathcal{D}$ is generated by the Laplacian.

Example 1.2. The case $G = SO_o(d, 1)$ yields the d-dimensional hyperbolic space $\mathbf{S} = \mathbb{H}^d$ (of rank 1), already dealt with in [3, 4].

We will pay special attention to the example of $G = SL_n(\mathbb{R})$, $K = SO(n, \mathbb{R})$. In that case, $\mathfrak{g}$ is the set of matrices with trace 0, $\mathfrak{k}$ the antisymmetric matrices, and one can take $\mathfrak{a}$ to be the set of diagonal matrices with trace 0. The connected group generated by $\mathfrak{a}$ is denoted by A, in this example it is the set of diagonal matrices of determinant 1 and with positive entries. The rank is $r = n - 1$.

We will be interested in Γ-invariant joint eigenfunctions of $\mathcal{D}$; in other words, eigenfunctions of $\mathcal{D}$ that go to the quotient $\mathbf{Y} = \Gamma\backslash G/K$. If we choose a set of generators of $\mathcal{D}$, the collection of eigenvalues can be represented as an element of $\mathbb{C}^r$. We will recall in Sections 2.2 and 2.3 that it is more natural to parametrize the eigenvalue by an element $\nu \in \mathfrak{a}^*_{\mathbb{C}}$, the complexified dual of $\mathfrak{a}$. More precisely, $\nu \in \mathfrak{a}^*_{\mathbb{C}}/W$ where W is the Weyl group of G, a finite group given by M'/M where M' is the normalizer of A, and M the centralizer of A, in K.

1.2. **Semiclassical limit.** Silberman and Venkatesh suggested to study the L^2-normalized eigenfunctions (ψ) in the limit $\|\nu\| \longrightarrow +\infty$, as a variant of the very popular question of understanding high-energy eigenfunctions of the Laplacian. The question of "quantum ergodicity" is to understand the asymptotic behaviour of the family of probability measures $d\bar{\mu}_\psi(y) = |\psi(y)|^2 dy$ on $\mathbf{Y} = \Gamma\backslash G/K$. They considered the case where $\frac{\nu}{\|\nu\|}$ has a limit $\nu_\infty \in \mathfrak{a}^*_{\mathbb{C}}/W$, with the sequence ν satisfying a certain number of additional properties that will be stated in the next paragraphs. For the moment, we just note that the real parts $\Re e(\nu)$ are uniformly bounded, so that $\Re e(\nu_\infty) = 0$ [15, §16.5(7) & Thm. 16.6]. We will denote $\Lambda_\infty = \Im m(\nu_\infty) = -i\nu_\infty$.

1.3. **Symplectic lift vs. representation-theoretic lift.** The locally symmetric space $\mathbf{Y}$ should be thought of as the *configuration space* of our dynamical system. To introduce tools from dynamical systems it is necessary to move to an appropriate *phase space*. Once we lift the eigenfunctions there, the measures we study become approximately invariant under some dynamics and we can apply the tools of ergodic theory. Two different kinds of lifts have been considered thus far: the *microlocal lift* (we also call it the *symplectic lift*) lifts the measure $\bar{\mu}_\psi$ to a distribution $\tilde{\mu}_\psi$ on the cotangent bundle $T^*\mathbf{Y} = \Gamma\backslash T^*(G/K)$, taking advantage of its symplectic structure. This construction applies in great generality, for example when $\mathbf{Y}$ is any compact Riemannian manifold. The *representation-theoretic*

[3] We shall denote by $\mathfrak{g}$ the Lie algebra of G, by $\mathfrak{k}$ the Lie algebra of K, and so on.

lift used in [27, 16, 23, 24, 6], specific to locally symmetric spaces, lifts the measure $\bar{\mu}_\psi$ to a measure μ_ψ defined on $\mathbf{X} = \Gamma\backslash G$, taking advantage of the homogeneous space structure of G/K.

The two lifts are very natural, and closely related. The symplectic lift will be defined in §3.4; we shall not give the construction of the representation-theoretic lift since we do not use it. Nevertheless, the representation theoretic vision will enter the picture when it comes to applying some measure classification results from [8, 9].

In the symplectic point of view, the dynamics is defined as follows. On $T^*(G/K)$, consider the algebra $\mathcal{H}$ of smooth G-invariant Hamiltonians (*i.e.* functions), that are polynomial in the fibers of the projection $T^*(G/K) \longrightarrow G/K$. This algebra is isomorphic to the algebra of W-invariant polynomials on $\mathfrak{a}^*$ (consider the restriction on $\mathfrak{a}^* \subset T^*_o(G/K)$, where $o = eK$ serves as the "origin" in G/K). The structure theory of semisimple Lie algebras shows that $\mathcal{H}$ is isomorphic to a polynomial ring in r generators. Moreover, the elements of $\mathcal{H}$ commute under the canonical Poisson bracket on $T^*(G/K)$. Thus, we have a family of r independent commuting Hamiltonians $H_1, ..., H_r$, which generate r commuting hamiltonian flows. The Killing metric, seen as a function on $T^*(G/K)$, is one of them, and its symplectic gradient generates the geodesic flow. Since all these flows are G-equivariant, they descend to the quotient $T^*\mathbf{Y}$.

Joint energy layers of $\mathcal{H}$ are naturally parametrized by elements $\Lambda \in \mathfrak{a}^*/W$. Here is the geometric explanation : fix a point in G/K, say the origin $o = eK$. Consider the flat totally geodesic submanifold $A.o \subset G/K$ going through o. It is isometric to $\mathbb{R}^r$, and the cotangent space $T^*_o(A.o)$ is naturally isomorphic to $\mathfrak{a}^*$. If $\mathcal{E} \subset T^*(G/K)$ is a joint energy layer of $\mathcal{H}$ (or equivalently a G-orbit in $T^*(G/K)$), then there exists $\Lambda \in \mathfrak{a}^*$ such that $\mathcal{E} \cap T^*_o(A.o) = W \cdot \Lambda$, the orbit of Λ under the Weyl group W. See [13] for details. We will denote $\mathcal{E}_\Lambda$ the energy layer of parameter Λ.

In Section 3 we will use a quantization procedure to associate to every Γ-invariant eigenfunction ψ a distribution $\tilde{\mu}_\psi$ on $T^*\mathbf{Y}$, called its microlocal lift. This distribution projects down to $\bar{\mu}_\psi$ on $\mathbf{Y}$. This is a standard construction, and the following theorem is an avatar of propagation of singularities for solutions of partial differential equations :

Theorem 1.3. *Assume that $\|\nu\| \longrightarrow +\infty$, and that $\frac{\nu}{\|\nu\|}$ has a limit ν_∞. Let $\Lambda_\infty = -i\nu_\infty \in \mathfrak{a}^*/W$. Any limit (in the distribution sense) of the sequence $\tilde{\mu}_\psi$ is a probability measure on $T^*\mathbf{Y}$, carried by the energy layer $\mathcal{E}_{\Lambda_\infty}$, and invariant under the family of Hamiltonian flows generated by $\mathcal{H}$.*

In order to transport this statement to get an A-invariant measure on $\Gamma\backslash G$, we must now make some assumptions on Λ_∞. We follow Silberman and Venkatesh, who assume in [23, 24] that ν_∞ is a regular element of $\mathfrak{a}^*_{\mathbb{C}}$, in the sense that it is not fixed by any non-trivial element of the Weyl group W. They show that it implies $\Re e(\nu_n) = 0$ for all but a finite number of ν_ns in the sequence [23, Thm. 2.7 (3)]. The element ν_∞ being regular is, of course, equivalent to Λ_∞ being regular; and this is also equivalent to the energy layer Λ_∞ being regular, in the sense that the differentials $dH_1, ..., dH_r$ are independent there [13].

As above, we denote by M the centralizer of A in K. There is a surjective map

$$\pi : G/M \times \mathfrak{a}^* \longrightarrow T^*(G/K) \tag{1.1}$$

$$(gM, \lambda) \mapsto (gK, g.\lambda) \tag{1.2}$$

The image of $G/M \times \{\lambda\}$ under π is the energy layer $\mathcal{E}_\lambda$. The map $\pi_\lambda : G/M \times \{\lambda\} \longrightarrow \mathcal{E}_\lambda$ is a diffeomorphism if and only if λ is regular (otherwise π_λ is not even injective). Under π_λ^{-1}, the action of the Hamiltonian flow Φ_H^t generated by $H \in \mathcal{H}$ on $\mathcal{E}_\lambda$ is conjugate to

$$gM \mapsto g \exp(t\, dH(\lambda))M.$$

The same statements hold after quotienting on the left by Γ. Since H is a function on $\mathfrak{a}^*$, the differential $dH(\lambda)$ is an element of $\mathfrak{a}$. Denoting by $R(e^{tX})$ the one–parameter flow on G/M generated by $X \in \mathfrak{a}$ (acting by multiplication on the right), we can rephrase this by writing

$$\pi \circ R(e^{t dH(\lambda)}) = \Phi_H^t \circ \pi \qquad \text{on } \mathcal{E}_\lambda.$$

If λ is regular, the elements $dH(\lambda)$ can be shown to span $\mathfrak{a}$ as H varies over $\mathcal{H}$. Otherwise, we have [13]

$$\{dH(\lambda), H \in \mathcal{H}\} = \{X \in \mathfrak{a}, \forall \alpha \in \Delta, (\langle \alpha, \lambda \rangle = 0 \Longrightarrow \alpha(X) = 0)\}, \tag{1.3}$$

where $\Delta \subset \mathfrak{a}^*$ is the set of roots.

Thus, Theorem 1.3 may be rephrased as follows:

Theorem 1.4. *Assume Λ_∞ is regular. Then any limit (in the distribution sense) of the sequence $\tilde{\mu}_\psi$ yields a probability measure on $\Gamma\backslash G/M$, invariant under the right action of A by multiplication.*

This theorem was proved in [23, Thm. 1.6 (3)] using the representation-theoretic lift; the equivariance of that lift shows that the construction is compatible with the Hecke operators on $\Gamma\backslash G$. It is also shown there that the symplectic lift $\tilde{\mu}_\psi$ and the representation theoretic lift μ_ψ have the same asymptotic behaviour as ν tends to infinity, if we identify $\mathcal{E}_{\Lambda_\infty} \subset \Gamma\backslash T^*(G/K)$ with $\Gamma\backslash G/M$.

Definition 1.5. We will call any limit point of the sequence $\tilde{\mu}_\psi$ (or μ_ψ) a *semiclassical measure* in the direction Λ_∞.

Thus, we can equivalently see the semiclassical measures in a regular direction as positive measures on $T^*(\Gamma\backslash G/K)$ carried by a regular energy layer, positive measures on $\Gamma\backslash G/M$, or M-invariant positive measures on $\Gamma\backslash G$.

1.4. **Entropy bounds.** Our main result is a non-trivial lower bound on the entropy of semiclassical measures. We fix $H \in \mathcal{H}$, and we consider the corresponding Hamiltonian flow Φ_H^t, which has Lyapunov exponents

$$-\chi_J(H) \leq \cdots \leq -\chi_1(H) \leq 0 \leq \chi_1(H) \leq \cdots \leq \chi_J(H).$$

In addition, the Lyapunov exponent 0 appears trivially with multiplicity r, as a consequence of the existence of r integrals of motion. The dimension of a regular energy layer is $r + 2J$. The integer J, the rank r and the dimension d of G/K are related by $d = J + r$.

In the following theorem we will denote by $\chi_{\max}(H) = \chi_J(H)$, the largest Lyapunov exponent. We denote by $h_{KS}(\mu, H)$ the Kolmogorov-Sinai entropy of a (Φ_H^t)-invariant probability measure μ on $T^*\mathbf{Y}$. We recall the Ruelle-Pesin inequality,

$$h_{KS}(\mu, H) \leq \int \sum_j \chi_j(H) d\mu,$$

which holds for any (Φ_H^t)-invariant probability measure μ. In general, the Lyapunov exponents are measurable functions on the phase space, but here, because of the homogeneous structure, the Lyapunov exponents are constant on each energy layer. Thus, if μ is carried on a single energy layer, this reduced to

$$h_{KS}(\mu, H) \leq \sum_j \chi_j(H).$$

Theorem 1.6. *(Symplectic version) Let μ be a semiclassical measure in the direction Λ_∞. Assume that Λ_∞ is regular.*

For $H \in \mathcal{H}$, we consider the corresponding Hamiltonian flow Φ_H^t on $\mathcal{E}_{\Lambda_\infty}$. Then

$$h_{KS}(\mu, H) \geq \sum_{j:\chi_j(H) \geq \frac{\chi_{\max}(H)}{2}} \left(\chi_j(H) - \frac{\chi_{\max}(H)}{2} \right). \tag{1.4}$$

Continuing with the assumption that Λ_∞ is regular, we can transport the theorem to $\Gamma\backslash G/M$. If we fix a 1-parameter subgroup (e^{tX}) of A (with $X \in \mathfrak{a}$), it is well known that the (non trivial) Lyapunov exponents of the flow (e^{tX}) acting on G/M are the real numbers $(\alpha(X))$, where $\alpha \in \mathfrak{a}^*$ run over the set of roots Δ (see Section 2 for background related to Lie groups). If α is a root then so is $-\alpha$ (one of the two will be called *positive*, the other *negative*, the notion of positivity is explained in detail later). We write $\alpha_{\max}(X)$ for $\max_\alpha \alpha(X)$, this is the largest Lyapunov exponent of the associated Hamiltonian flow. Each root occurs with *multiplicity* m_α, which must be taken into account in the statements below (the corresponding Lyapunov exponent $\alpha(X)$ would be counted repeatedly, m_α times).

Theorem 1.7. *(Group-theoretic version) Let μ be a semiclassical measure in the direction Λ_∞. Assume that Λ_∞ is regular.*

For $X \in \mathfrak{a}$, let $h_{KS}(\mu, X)$ be the entropy of μ with respect to the flow (e^{tX}) acting on G/M. Then

$$h_{KS}(\mu, X) \geq \sum_{\alpha:\alpha(X) \geq \frac{\alpha_{\max}(X)}{2}} m_\alpha \left(\alpha(X) - \frac{\alpha_{\max}(X)}{2} \right). \tag{1.5}$$

Our lower bound is positive for all non-zero X, in fact greater than $\frac{\alpha_{\max}(X)}{2}$. In [1, 3], the first author and S. Nonnenmacher had conjectured the following stronger bound

$$h_{KS}(\mu, H) \geq \frac{1}{2} \sum_j \chi_j(H)$$

or equivalently

$$h_{KS}(\mu, X) \geq \frac{1}{2} \sum_{\alpha, \alpha(X) \geq 0} m_\alpha \cdot \alpha(X). \tag{1.6}$$

We are still unable to prove it, except in one case: when all the *positive* Lyapunov exponents are equal to one another, so that formula (1.5) reduces to (1.6). An example is the case of hyperbolic d-space ($G = SO(d,1)$) alluded to above. Another, the main focus of the present paper, is the case of the "extremely irregular" elements of the torus in $G = SL_n(\mathbb{R})$. These are the elements conjugate under the Weyl group to

$$X = \mathrm{diag}(n-1, -1, ..., -1).$$

1.5. **Application: towards Quantum Unique Ergodicity on locally symmetric spaces.** In Section 6 we combine our entropy bounds with measure classification results. Let $n \geq 3$, $G = SL_n(\mathbb{R})$, $\Gamma < G$ a cocompact lattice. Let μ be a semiclassical measure on $\Gamma \backslash G$ in the regular direction Λ_∞.

The measure μ can be written uniquely as a sum of an absolutely continuous measure and a singular measure (with respect to Haar measure). Since μ is invariant under the action of A, the same holds for both components. Because the Haar measure is known to be ergodic for the action of A, the absolutely continuous part of μ is, in fact, proportional to Haar measure. We call this the *Haar component* of μ. Its total mass is the *weight* of this component.

Theorem 1.8. *Let $n = 3$. Then μ has a Haar component of weight $\geq \frac{1}{4}$.*

Theorem 1.9. *Let $n = 4$. Then either μ has a Haar component, or each ergodic component is the Haar measure on a closed orbit of the group* $\begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}$ *(or one of its 4 images under the Weyl group), and the components invariant by each of these 4 subgroups have total weight $\frac{1}{4}$.*

In fact, the proof of Theorem 1.9 shows the following : if some "extremely irregular" element acts on μ with entropy strictly larger than half of its entropy w.r.t. Haar measure, then there is a Haar component.

It does not seem to be possible to push this technique beyond SL_4. The problem is that there are large subgroups (in the style of those occuring in Theorem 1.9) whose closed orbits support measures of large entropy. For particular lattices, however, these large subgroups do not have closed orbits, so the only possible non-Haar components have small entropy and cannot account for all the entropy. For co-compact lattices this occurs, for example, when Γ is the set of elements of reduced norm 1 of an order in a central division algebra over $\mathbb{Q}$, or more generally for any lattice commensurable with one obtained this way (we say that Γ is *associated* to the division algebra). Such lattices are said to be of "inner type" since they correspond to inner forms of SL_n over $\mathbb{Q}$ (there also exist non-uniform lattices of inner type, corresponding to central simple $\mathbb{Q}$-algebras which are not division algebras). For a brief description of the construction and references see Section 6.

Theorem 1.10. *For $n \geq 3$ let $\Gamma < SL_n(\mathbb{R})$ be a lattice associated to a division algebra over $\mathbb{Q}$, and let μ be a semiclassical measure on $\Gamma \backslash SL_n(\mathbb{R})$ in a regular direction. Then μ has a Haar component of weight $\geq \frac{\frac{n+1}{2}-t}{n-t} > 0$ where t is the largest proper divisor of n.*

It is not surprising that the strongest implication is for n prime (so that there are few intermediate algebraic measures). Indeed, setting $t = 1$ we find that the weight of the Haar component is greater than $\frac{1}{2}$ in that case. However for n prime Silberman-Venkatesh [24] show that the semiclassical measures associated to *Hecke eigenfunctions* are equal to Haar measure. The main impact of Theorem 1.10 is thus when the n is composite, where previous methods only showed that semiclassical measures are convex combinations of algebraic measures but could not establish that Haar measure occurs in the combination.

Remark 1.11. We compare here our result with that of [24]. That paper studies the case of lattices in $G = PGL_n(\mathbb{R})$ associated to division algebras of prime degree n and joint eigenfunctions of $\mathcal{D}$ and of the Hecke operators. It is then shown that *any ergodic component of a semiclassical measure* μ has positive entropy; it follows that μ must be the Haar measure. While we cannot get that far, in some respects our result is stronger:

- We do not assume that our eigenfunctions are also eigenfunctions of the Hecke operators: this means that multiplicity of eigenvalues is not an issue in this work.
- Our lower bound on the total entropy (1/2 of the maximal entropy) is explicit and quite strong. This allows us to detect the presence of a Haar component in a variety of cases.
- In particular, for $n = 3$ we do not need any assumption on the cocompact lattice Γ; and for Γ associated to a division algebra, our result holds for all n.
- The Hecke-operator method applies more naturally to adelic quotients $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})/K_\infty K_{\mathrm{f}}$. When G is a form of SL_n there is no distinction, but when $G = PGL_n$ the adelic quotients are typically *disjoint unions* of quotients $\Gamma\backslash G$. Even when the quotient is compact, G-invariance of the limit measure does not show that all components have the same proportion of the mass. Our result applies to each connected component separately.

Conversely, there are features of the other approach we cannot reach.

- We cannot control the entropy of individual ergodic components. In particular, we cannot exclude components of zero entropy.
- The methods of Silberman-Venkatesh apply to non-cocompact lattices as well.

1.6. **Hyperbolic dispersive estimate.** The proof of Theorem 1.6 (and 1.7) follows the main ideas of [3], with a major difference which lies in an improvement of the "hyperbolic dispersive estimate" : by this we mean [1, Thm. 1.3.3] and [3, Thm. 2.7]. If we applied directly the result of [3], we would get

$$h_{KS}(\mu, H) \geq \sum_{j=0}^{J} \left(\chi_j(H) - \frac{\chi_{\max}(H)}{2} \right).$$

This inequality is often trivial (when the right-hand term is negative) whereas in (1.4) we managed to get rid of the negative terms $\left(\chi_j(H) - \frac{\chi_{\max}(H)}{2}\right)$.

Since the "hyperbolic dispersive estimate" has an intrinsic interest, and is the core of this paper, we state it here as one of our main results. We fix a quantization procedure, set at scale $\hbar = \|\nu\|^{-1}$, that associates to any reasonable function a on $T^*\mathbf{Y}$ an operator $\mathrm{Op}_\hbar(a)$ on $L^2(\mathbf{Y})$. An explicit construction is given in Section 3. In particular, it is useful to know that $\mathrm{Op}_\hbar$ can be defined so that, if $H \in \mathcal{H}$ is real valued, $\mathrm{Op}_\hbar(H)$ is a self-adjoint operator belonging to $\mathcal{D}$. More explicitly, $\mathrm{Op}_\hbar(H)$ is defined so that $\mathrm{Op}_\hbar(H)\psi = H(-i\hbar\nu)\psi$ for any $\mathcal{D}$-eigenfunction ψ of spectral parameter ν (hence the choice of the normalisation $\hbar = \|\nu\|^{-1}$).

In order to state the hyperbolic dispersive estimate, we introduce $(P_k)_{k=1,\ldots,\kappa}$ a finite family of smooth real functions on $\mathbf{Y}$, such that

$$\forall x \in \mathbf{Y}, \qquad \sum_{m=1}^{\kappa} P_k^2(x) = 1\ . \tag{1.7}$$

We assume that the diameter of the supports of the functions P_k is smaller than the injectivity radius of $\mathbf{Y}$. We will denote by $\widehat{P}_k$ the operator of multiplication by $P_k(x)$ on the Hilbert space $L^2(\mathbf{Y})$.

We denote by $U^t = \exp(i\hbar^{-1}t\,\mathrm{Op}_\hbar(H))$ the propagator of the "Schrödinger equation" generated by the Hamiltonian H. This is a unitary Fourier Integral Operator associated with the classical Hamiltonian flow Φ_H^{-t}. The $\hbar$-dependence of U^t will be implicit in our notation. We fix a small time step η.

Throughout the paper we will use the notation $\widehat{A}(t) = U^{-t\eta}\widehat{A}U^{t\eta}$ for the quantum evolution at time $t\eta$ of an operator $\widehat{A}$. For each integer $T \in \mathbb{N}$ and any sequence of labels $\underline{\omega} = (\omega_{-T}, \cdots, \omega_{-1}, \omega_0, \cdots \omega_{T-1})$, $\omega_i \in [1, \kappa]$ (we say that the sequence $\underline{\omega}$ is of *length* $|\underline{\omega}| = 2T$), we define the operators

$$\widehat{P}_{\underline{\omega}} = \widehat{P}_{\omega_{T-1}}(T-1)\widehat{P}_{\omega_{T-2}}(T-2)\ldots\widehat{P}_{\omega_0}\widehat{P}_{\omega_{-1}}(-1)\ldots\widehat{P}_{\omega_{-T}}(-T)\ . \tag{1.8}$$

We use a smooth, compactly supported function ϕ on $\mathfrak{a}^*$, supported in a tubular neighbourhood of size ϵ of Λ_∞. The function ϕ is assimilated to the function $\phi \circ \pi^{-1}$ on $T^*\mathbf{Y}$ using the map π (1.1). We define

$$\widehat{P}^{\phi}_{\underline{\omega}} = \widehat{P}_{\omega_{T-1}}(T-1)\widehat{P}_{\omega_{T-2}}(T-2)\ldots\widehat{P}^{1/2}_{\omega_0}\,\mathrm{Op}_\hbar(\phi)\widehat{P}^{1/2}_{\omega_0}\widehat{P}_{\omega_{-1}}(-1)\ldots\widehat{P}_{\omega_{-T}}(-T)\ . \tag{1.9}$$

The operator $\widehat{P}^{\phi}_{\underline{\omega}}$ should be thought of as $\widehat{P}_{\underline{\omega}}$ restricted to a spectral window around Λ_∞.

Theorem 1.12. *(Hyperbolic dispersive estimate) Let Λ_∞ be a regular element of $\mathfrak{a}^*$. Fix $H \in \mathcal{H}$, and a time step η, small enough. Let 2^- be an arbitrary number < 2.*

Then, for any $\mathcal{K} > 0$, there exists $\epsilon > 0$, such that : if $\phi \in \mathcal{C}^\infty(T^*\mathbf{Y})$ is supported in a tubular neighbourhood of size $\leq \epsilon$ of the regular energy layer $\mathcal{E}_{\Lambda_\infty}$, the following statement holds.

For $\hbar$ small enough, for $T = \lfloor \frac{\mathcal{K}|\log \hbar|}{\eta} \rfloor$, and for every sequence $\underline{\omega}$ of length $2T$,

$$\|\widehat{P}^\phi_{\underline{\omega}}\|_{L^2(\mathbf{Y}) \longrightarrow L^2(\mathbf{Y})} \leq \prod_{j,\, \chi_j(H) \geq \frac{1}{2\mathcal{K}}} \frac{e^{-T\eta \inf_{\operatorname{supp}\phi} \chi_j(H)}}{\hbar^{1/2^-}} \tag{1.10}$$

where we take the infimum of each Lyapunov exponent $\chi_j(H)$ over the support of ϕ.

The method used in [3] only yields the upper bound:

$$\|\widehat{P}^\phi_{\underline{\omega}}\| \leq \prod_{j=1}^{J} \frac{e^{-T\eta \inf_{\operatorname{supp}\phi} \chi_j(H)}}{\hbar^{1/2^-}} \tag{1.11}$$

in other words it involved all the Lyapunov exponents. This is clearly not optimal when Φ^t_H has some neutral, or slowly expanding directions. For instance, if $H = 0$ then $\Phi^t_H = I$ has only neutral directions. In this case, (1.11) reads

$$\|\widehat{P}^\phi_{\underline{\omega}}\| \leq \hbar^{-d/2^-}, \tag{1.12}$$

where d is the dimension of $\mathbf{Y}$, which is obviously much worse (for any T) than the trivial bound

$$\|\widehat{P}^\phi_{\underline{\omega}}\| \leq 1. \tag{1.13}$$

On the other hand, if some of the $\chi_j(H)$ are (strictly) positive, then (1.11) is much better than the trivial bound (1.13), for very large $T\eta$. The bound given by Theorem 1.12 takes, in some sense, the best of the two bounds in each Lyapunov direction.

The proof of the hyperbolic dispersion estimates occupies Sections 3, 4, 5. The techniques are completely disjoint from the measure rigidity arguments used in §6, which explains that §3, 4, 5 have a completely different flavor from §6. The tools are those of semiclassical analysis. We use a version of the pseudodifferential calculus adapted to the geometry of locally symmetric spaces, based on Helgason's version of the Fourier transform for these spaces, and inspired by the work of Zelditch in the case of $G = SL(2, \mathbb{R})$ [28]. We point out the fact that an alternative proof of Theorem 1.12 is given in [2], based on more conventional Fourier analysis and the exclusive use of the symplectic language. Sections 3, 4, 5 were written long before the paper [2], as it seemed at first to be the most natural idea to use the Helgason calculus. The high technicality of these sections served as a motivation to look for an alternative presentation, and the comparison of the two approaches is *a posteriori* in favor of the symplectic language. Thus, the reader might prefer to read [2] instead of Sections 3, 4, 5. However, we feel that the two approaches have an interest of their own.

We will not repeat here the argument that leads from Theorem 1.12 to the entropy bound Theorem 1.6; it would be an exact repetition of the argument given in [3, §2]. Let us just make one comment : in this argument, we are limited to $\mathcal{K} = \frac{1}{\chi_{\max}(H)}$ (the time $T_E = \frac{|\log \hbar|}{\chi_{\max}(H)}$ is sometimes called the Ehrenfest time for the Hamiltonian H, and corresponds to the time where the approximation of the quantum flow U^t by the classical flow Φ_H^{-t} breaks down). This means that we eventually keep the Lyapunov exponents such that $\chi_j(H) \geq \frac{\chi_{\max}(H)}{2}$, and explains why this restriction appears in (1.4).

2. Background regarding semisimple Lie groups

Our terminology follows Knapp [15].

2.1. **Structure.** Let G denote a non-compact connected simple Lie group with finite center[4]. The neutral element in G will be denoted by e. We choose a Cartan involution Θ for G, and let $K < G$ be the Θ-fixed maximal compact subgroup. Let $\mathfrak{g} = Lie(G)$, and let ϑ denote the differential of Θ, giving the Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with $\mathfrak{k} = Lie(K)$. Let $\mathbf{S} = G/K$ be the symmetric space, with $o = eK \in \mathbf{S}$ the point with stabilizer K. We fix a G-invariant metric on G/K: observe that the tangent space at the point o is naturally identified with $\mathfrak{p}$, and endow it with the Killing form. For a lattice $\Gamma < G$ we write $\mathbf{X} = \Gamma\backslash G$ and $\mathbf{Y} = \Gamma\backslash G/K$, the latter being a locally symmetric space of non-positive curvature. In this paper, we shall always assume that $\mathbf{X}$ and $\mathbf{Y}$ are compact.

Fix now a maximal abelian subalgebra $\mathfrak{a} \subset \mathfrak{p}$.

We denote by $\mathfrak{a}_{\mathbb{C}}$ the complexification $\mathfrak{a} \otimes \mathbb{C}$ and by $\mathfrak{a}^*$ (resp. $\mathfrak{a}^*_{\mathbb{C}}$) the real dual (resp. the complex dual) of $\mathfrak{a}$. For $\nu \in \mathfrak{a}^*_{\mathbb{C}}$, we define $\Re e(\nu), \Im m(\nu) \in \mathfrak{a}^*$ to be the real and imaginary parts of ν, respectively. For $\alpha \in \mathfrak{a}^*$, set $\mathfrak{g}_\alpha = \{X \in \mathfrak{g}, \forall H \in \mathfrak{a} : ad(H)X = \alpha(H)X\}$, $\Delta = \Delta(\mathfrak{a} : \mathfrak{g}) = \{\alpha \in \mathfrak{a}^* \setminus \{0\}, \mathfrak{g}_\alpha \neq \{0\}\}$ and call the latter the (restricted) roots of $\mathfrak{g}$ with respect to $\mathfrak{a}$. The subalgebra $\mathfrak{g}_0$ is ϑ-invariant, and hence $\mathfrak{g}_0 = (\mathfrak{g}_0 \cap \mathfrak{p}) \oplus (\mathfrak{g}_0 \cap \mathfrak{k})$. By the maximality of $\mathfrak{a}$ in $\mathfrak{p}$, we must then have $\mathfrak{g}_0 = \mathfrak{a} \oplus \mathfrak{m}$ where $\mathfrak{m} = Z_{\mathfrak{k}}(\mathfrak{a})$, the centralizer of $\mathfrak{a}$ in $\mathfrak{k}$.

The Killing form of $\mathfrak{g}$ induces a standard inner product $\langle .,. \rangle$ on $\mathfrak{p}$, and by duality on $\mathfrak{p}^*$. By restriction we get an inner product on $\mathfrak{a}^*$ with respect to which $\Delta(\mathfrak{a} : \mathfrak{g}) \subset \mathfrak{a}^*$ is a root system. The associated Weyl group, generated by the root reflections s_α, will be denoted by $W = W(\mathfrak{a} : \mathfrak{g})$. This group is canonically isomorphic to $N_K(\mathfrak{a})/Z_K(\mathfrak{a})$. In what follows we will represent any element w of the Weyl group by a representative in $N_K(\mathfrak{a}) \subset K$ (taking care to only make statements that do not depend on the choice of a representative), and the action of $w \in W(\mathfrak{a} : \mathfrak{g})$ on $\mathfrak{a}$ or $\mathfrak{a}^*$ will be given by the adjoint representation $\mathrm{Ad}(w)$. The fixed-point set of any s_α is a hyperplane in $\mathfrak{a}^*$, called a wall. The connected components of the complement of the union of the walls are cones, called the (open) Weyl chambers. A subset $\Pi \subset \Delta(\mathfrak{a} : \mathfrak{g})$ will be called a system of simple roots (or a simple system) if every root can be uniquely expressed as an integral combination of elements of Π with either all coefficients non-negative or all coefficients non-positive. For a simple system Π, the open cone $C_\Pi = \{\nu \in \mathfrak{a}^*, \forall \alpha \in \Pi : \langle \nu, \alpha \rangle > 0\}$ is an (open) Weyl

[4]If G is semisimple our discussion remains valid, but one can even do something finer, as remarked in [23, §5.1]. After decomposing $\mathfrak{g}$ into simple factors $\oplus \mathfrak{g}^{(j)}$, and assuming that the Cartan involution, the subalgebra $\mathfrak{a}$, etc. are compatible with this decomposition, one can decompose the spectral parameter ν into its components $\nu^{(j)} \in \mathfrak{a}^{(j)*}$. Instead of assuming that $\|\nu\| \longrightarrow +\infty$ and $\frac{\nu}{\|\nu\|}$ has a regular limit ν_∞, one can assume the same independently for each component $\nu^{(j)}$. This means that we do not have to assume that all the norms $\|\nu^{(j)}\|$ go to infinity at the same speed.

chamber. The closure of an open chamber will be called a closed chamber; we will denote in particular $\overline{C_\Pi} = \{\nu \in \mathfrak{a}^*, \forall \alpha \in \Pi : \langle \nu, \alpha \rangle \geq 0\}$. The Weyl group acts simply transitively on the chambers and simple systems. The action of $W(\mathfrak{a} : \mathfrak{g})$ on $\mathfrak{a}^*$ extends in the complex-linear way to an action on $\mathfrak{a}^*_{\mathbb{C}}$ preserving $i\mathfrak{a}^* \subset \mathfrak{a}^*_{\mathbb{C}}$, and we call an element $\nu \in \mathfrak{a}^*_{\mathbb{C}}$ regular if it is fixed by no non-trivial element of $W(\mathfrak{a} : \mathfrak{g})$. Since $-C_\Pi \subset \mathfrak{a}^*$ is a chamber, there is a unique $w_l \in W(\mathfrak{a} : \mathfrak{g})$, called the "long element", such that $\mathrm{Ad}(w_l).C_\Pi = -C_\Pi$. Note that $w_l^2 C_\Pi = C_\Pi$ and hence $w_l^2 = e$. Also, w_l depends on the choice of Π but we suppress this from the notation.

Fixing a simple system Π we get a notion of positivity. We will denote by Δ^+ the set of positive roots, by $\Delta^- = -\Delta^+$ the set of negative roots. We use $\rho = \frac{1}{2}\sum_{\alpha>0}(\dim \mathfrak{g}_\alpha)\alpha \in \mathfrak{a}^*$ to denote half the sum of the positive roots. For $\mathfrak{n} = \oplus_{\alpha>0}\mathfrak{g}_\alpha$ and $\bar{\mathfrak{n}} = \Theta\mathfrak{n} = \oplus_{\alpha<0}\mathfrak{g}_\alpha$ we have $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m} \oplus \bar{\mathfrak{n}}$. Note that $\bar{\mathfrak{n}} = \mathrm{Ad}(w_l).\mathfrak{n}$. We also have ("Iwasawa decomposition") $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$. We can therefore uniquely write every $X \in \mathfrak{g}$ in the form $X = X_{\mathfrak{n}} + X_{\mathfrak{a}} + X_{\mathfrak{k}}$. We also write $H_0(X)$ for $X_{\mathfrak{a}}$.

Let $N, A, \overline{N} < G$ be the connected subgroups corresponding to the subalgebras $\mathfrak{n}, \mathfrak{a}, \bar{\mathfrak{n}} \subset \mathfrak{g}$ respectively, and let $M = Z_K(\mathfrak{a})$. Then $\mathfrak{m} = Lie(M)$, though M is not necessarily connected. Moreover $P_0 = NAM$ is a minimal parabolic subgroup of G, with the map $N \times A \times M \longrightarrow P_0$ being a diffeomorphism. The map $N \times A \times K \longrightarrow G$ is a diffeomorphism (Iwasawa decomposition), so for $g \in G$ there exists a unique $H_0(g) \in \mathfrak{a}$ such that $g = n\exp(H_0(g))k$ for some $n \in N$, $k \in K$. The map $H_0 : G \longrightarrow \mathfrak{a}$ is continuous; restricted to A, it is the inverse of the exponential map.

We will use the G-equivariant identification between G/M and $G/K \times G/P_0$, given by $gM \mapsto (gK, gP_0)$ (we denote by $gM \in G/M$ the class of g in G/M, and so on). The quotient G/P_0 can also be identified with K/M.

Starting from H_0 we define a "Busemann function" $\mathcal{B}$ on $G/K \times G/P_0 \sim G/M$:

$$\mathcal{B}(gK, g_1P_0) = H_0(k^{-1}g), \tag{2.1}$$

where k is the K-part in the KAN decomposition of g_1 (if g_1 is defined modulo P_0, then k is defined modulo M). Equivalently, if $gM \in G/M$, we have

$$\mathcal{B}(gM) = a, \tag{2.2}$$

where $g = kna$ is the KNA decomposition of g (if g is defined modulo M, then a is uniquely defined and k is defined modulo M).

In G/K, a "flat" is a maximal flat totally geodesic submanifold. Every flat is of the form $\{gaK, a \in A\}$ for some $g \in G$. The space of flats can be naturally identified with G/MA, or with an open dense subset of $G/P_0 \times G/\overline{P}_0$, via the G-equivariant map

$$gMA \mapsto (gP_0, g\overline{P}_0)$$

where $\overline{P}_0 = MA\overline{N} = w_lP_0w_l^{-1}$. We will also use the following injective map from G/MA into $G/P_0 \times G/P_0$,

$$gMA \mapsto (gP_0, gw_lP_0). \tag{2.3}$$

Its image is an open dense subset of $G/P_0 \times G/P_0$, namely $\{(g_1P_0, g_2P_0), g_2^{-1}g_1 \in P_0w_lP_0\}$. We recall the Bruhat decomposition $G = \sqcup_{w\in W(\mathfrak{a}:\mathfrak{g})} P_0wP_0$, with $P_0w_lP_0$ being an open dense subset (the "big cell").

2.2. **The universal enveloping algebra; Harish-Chandra isomorphisms.** Recall that $\mathcal{D}$ is the algebra of G-invariant differential operators on $\mathbf{S}$. We analyze its structure by comparing it with other algebras of differential operators.

For a Lie algebra $\mathfrak{s}$ we write $\mathfrak{s}_{\mathbb{C}}$ for its complexification $\mathfrak{s}\otimes_{\mathbb{R}}\mathbb{C}$. In particular, $\mathfrak{g}_{\mathbb{C}}$ is a complex semisimple Lie algebra. We fix a maximal abelian subalgebra $\mathfrak{b} \subset \mathfrak{m}$ and let $\mathfrak{h} = \mathfrak{a}\oplus\mathfrak{b}$. Then $\mathfrak{h}_{\mathbb{C}}$ is a Cartan subalgebra of $\mathfrak{g}_{\mathbb{C}}$, with an associated root system $\Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ satisfying $\Delta(\mathfrak{a} : \mathfrak{g}) = \{\alpha_{|\mathfrak{a}}\}_{\alpha\in\Delta(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})} \setminus \{0\}$.

If $\mathfrak{s}_{\mathbb{C}}$ is a complex Lie algebra, we denote by $U(\mathfrak{s}_{\mathbb{C}})$ its universal enveloping algebra; $U(\mathfrak{g}_{\mathbb{C}})$ is isomorphic to the algebra of left-G-invariant differential operators on G with complex coefficients [10].

There is an algebra isomorphism, called the Harish-Chandra isomorphism and described below, between $\mathcal{D}$ and the algebra $\mathcal{D}_W(A)$ of A- and W-invariant differential operators on A. The latter is canonically isomorphic to $U(\mathfrak{a}_{\mathbb{C}})^W$, the subalgebra of $U(\mathfrak{a}_{\mathbb{C}})$ formed of W-invariant elements. Since $\mathfrak{a}_{\mathbb{C}}$ is abelian, $U(\mathfrak{a}_{\mathbb{C}})$ can be identified with the space of polynomial functions on $\mathfrak{a}^*$ with complex coefficients.

The Harish-Chandra isomorphism $\Gamma_{HC} : \mathcal{D} \longrightarrow \mathcal{D}_W(A)$ can be realized in a geometric way as follows [11, Cor. II.5.19]. Consider the flat subspace $A.o \subset G/K$, naturally identified with A. Fixing $D \in \mathcal{D}$, let $\Delta_N(D)$ be the translation-invariant differential operator on A (that is, an element of $U(\mathfrak{a})$) given by

$$\left(\Delta_N(D)f\right)(a) = D\tilde{f}(a.o),$$

for $a \in A$, $f \in C^\infty(A.o)$, and where $\tilde{f}$ stands for the unique N-invariant function on G/K that coincides with f on $A.o$. Then, we define

$$\Gamma_{HC} : D \mapsto e^{-\rho}\circ\Delta_N(D)\circ e^{\rho},$$

remembering that ρ is half the sum of positive roots and thus can be seen as a function on A. Note that

$$e^{-\rho}\circ\Delta_N(D)\circ e^{\rho} = \tau_\rho\Delta_N(D),$$

where τ_ρ is the automorphism of $U(\mathfrak{a})$ defined by letting $\tau_\rho(X) = X + \rho(X)$ for every $X \in \mathfrak{a}$.

In what follows, we denote by $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$ the center of $U(\mathfrak{g}_{\mathbb{C}})$. Thus, $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$ is the algebra of G-bi-invariant operators. Differentiating the action of G on $\mathbf{S}$ gives a map $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}}) \to \mathcal{D}$. The next lemma allows to compare the isomorphism Γ_{HC} with the isomorphism $\gamma_{HC} : \mathcal{Z}(\mathfrak{g}_{\mathbb{C}}) \longrightarrow U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ that is used in [23] and also bears the name of Harish-Chandra. It is defined by $\gamma_{HC}(z) = \tau_{\rho_{\mathfrak{h}}}\mathrm{pr}(z)$, where $\mathrm{pr}(z) \in U(\mathfrak{h}_{\mathbb{C}})$ is such that $z - \mathrm{pr}(z) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}}) + U(\mathfrak{g}_{\mathbb{C}})\mathfrak{k}_{\mathbb{C}}$.

Lemma 2.1. *Assume that the restriction from $\mathfrak{h}_{\mathbb{C}}$ to $\mathfrak{a}$ induces a surjection from $U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ to $U(\mathfrak{a}_{\mathbb{C}})^W$ (thought of as functions on the respective linear spaces).*

Let $D \in \mathcal{D}$ be of degree $\bar{d}$. Then there exists $Z \in \mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$ such that Z and D coincide on (right-)K-invariant functions, and such that

$$Z - \tau_{-\rho}\Gamma_{HC}(D) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})^{\bar{d}-2} + U(\mathfrak{g}_{\mathbb{C}})\mathfrak{k}_{\mathbb{C}}.$$

Remark 2.2. The assumption is automatically satisfied when G is split. It is also satisfied when G/K is a *classical* symmetric space, that is when G is a classical group [11, p. 341]. In fact the lemma itself is Proposition II.5.32 of [11], with the difference of degree between Z and $\tau_{-\rho}\Gamma_{HC}(D)$ made precise.

Proof. Let $D \in \mathcal{D}$ be of degree $\bar{d}$, so that $\Gamma_{HC}(D) \in U(\mathfrak{a}_{\mathbb{C}})^W$ is a polynomial of degree $\leq \bar{d}$. By assumption, we can extend $\Gamma_{HC}(D)$ to an element of $U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$. Consider $Z_1 = \gamma_{HC}^{-1}\Gamma_{HC}(D)$. It is shown in [23, Cor. 4.4] that

$$Z_1 - \tau_{-\rho}\Gamma_{HC}(D) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})^{\bar{d}-2} + U(\mathfrak{g}_{\mathbb{C}})\mathfrak{k}_{\mathbb{C}}.$$

It is not completely clear that Z_1 and D coincide on K-invariant functions, but the above formula shows that $\Gamma_{HC}(Z_1) - \Gamma_{HC}(D)$ is of degree $\leq \bar{d} - 2$, and hence that $Z_1 - D$ has degree at most $\bar{d} - 2$.

By descending induction on the degree of $\Gamma_{HC}(Z) - \Gamma_{HC}(D)$, we see that we can thus construct $Z \in \mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$ such that

$$Z - \tau_{-\rho}\Gamma_{HC}(D) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})^{\bar{d}-2} + U(\mathfrak{g}_{\mathbb{C}})\mathfrak{k}_{\mathbb{C}}$$

and such that $\Gamma_{HC}(Z) - \Gamma_{HC}(D) = 0$ (which precisely means that Z and D coincide on right-K-invariant functions). $\square$

2.3. **The Helgason-Fourier transform.** In (2.1) we introduced the "Busemann function" $\mathcal{B}$. For any $\theta \in G/P_0$, $\nu \in \mathfrak{a}_{\mathbb{C}}^*$, the function

$$e_{\theta,\nu} : x \in G/K \mapsto e^{(\rho+\nu)\mathcal{B}(x,\theta)} \tag{2.4}$$

is a joint eigenfunction of $\mathcal{D}$, satisfying

$$De_{\theta,\nu} = \Gamma_{HC}(D)(\nu)e_{\theta,\nu},$$

for every $D \in \mathcal{D}$. Here we have seen $\Gamma_{HC}(D)$ as a W-invariant polynomial on $\mathfrak{a}_{\mathbb{C}}^*$.

In fact for any joint eigenfunction ψ of $\mathcal{D}$ there exists $\nu \in \mathfrak{a}_{\mathbb{C}}^*$ such that

$$D\psi = \Gamma_{HC}(D)(\nu)\psi$$

for every $D \in \mathcal{D}$ [11, Ch. II Thm. 5.18, Ch. III Lem. 3.11]. The parameter ν is called the "spectral parameter" of ψ; it is uniquely determined up to the action of W.

The Helgason–Fourier transform gives the spectral decomposition of a function $u \in C_c^\infty(\mathbf{S})$ on the "basis" $(e_{\theta,\nu})$ of eigenfunctions of $\mathcal{D}$. It is defined as

$$\mathcal{F}u(\theta, \lambda) = \int_{\mathbf{S}} u(x)e_{\theta,-i\lambda}(x)dx, \tag{2.5}$$

$(\lambda \in \mathfrak{a}^*, \theta \in G/P_0)$. It has an inversion formula:

$$u(x) = \int_{\theta\in G/P_0, \lambda\in\overline{C_\Pi}} \mathcal{F}u(\theta, \lambda)e_{\theta,i\lambda}(x)d\theta|c(\lambda)|^{-2}d\lambda.$$

Here $d\theta$ denotes the normalized K-invariant measure on $G/P_0 \sim K/M$. The function c is the so-called Harish-Chandra function, given by the Gindikin-Karpelevic formula [11, Thm. 6.14, p. 447]. The Plancherel formula reads

$$\|u\|^2_{L^2(\mathbf{S})} = \int_{\theta \in G/P_0, \lambda \in \overline{C_\Pi}} |\mathcal{F}u(\theta,\lambda)|^2 d\theta |c(\lambda)|^{-2} d\lambda.$$

Remark 2.3. For $D \in \mathcal{D}$, D acts on u by

$$Du(x) = \int_{\theta \in G/P_0, \lambda \in \overline{C_\Pi}} \Gamma_{HC}(D)(i\lambda)\mathcal{F}u(\theta,\lambda) e_{\theta,i\lambda}(x) d\theta |c(\lambda)|^{-2} d\lambda$$

3. Quantization and pseudodifferential operators

In this section we develop a pseudodifferential calculus for the symmetric space $\mathbf{S}$, inspired by the work of Zelditch [28] and based on the Helgason-Fourier transform, in other words, on the spectral decomposition of the algebra $\mathcal{D}$. We do not push the analysis as far as possible, but just state the facts we need for our purposes; for a more detailed analysis we refer to Michael Schröder's thesis [22].

3.1. **Semiclassical Helgason transform.** We now introduce a parameter $\hbar$. In the sequel it will tend to 0 at the same speed as $\|\nu\|^{-1}$; the reader may identify the two. The parameter will be assumed to go to infinity in the conditions of §1.2, the limit ν_∞ assumed to be regular.

From now on we rescale the parameter space $\mathfrak{a}^*$ of the Helgason–Fourier transform by $\hbar$. We define the semiclassical Fourier transform, $\mathcal{F}_\hbar u(\theta,\lambda) = \mathcal{F}u(\theta,\hbar^{-1}\lambda)$. Thus, for $u \in C_c^\infty(\mathbf{S})$, we rewrite equation (2.5) as:

$$\mathcal{F}_\hbar u(\theta,\lambda) = \int_{\mathbf{S}} u(x) e_{\theta,-i\hbar^{-1}\lambda}(x) dx$$

$(\lambda \in \mathfrak{a}^*, \theta \in G/P_0)$. The inversion formula now reads

$$u(x) = \int_{\theta \in G/P_0, \lambda \in \overline{C_\Pi}} \mathcal{F}_\hbar u(\theta,\lambda) e_{\theta,i\hbar^{-1}\lambda}(x) d\theta |c_\hbar(\lambda)|^{-2} d\lambda,$$

with the "semiclassical Harish-Chandra c-function",

$$|c_\hbar(\lambda)|^{-2} = \hbar^{-r}|c(\hbar^{-1}\lambda)|^{-2}.$$

Remark 3.1. By the Gindikin-Karpelevic formula [11, Thm. 6.14, p. 447], we have

$$|c(\hbar^{-1}\lambda)|^{-2} \asymp \hbar^{-\dim \mathfrak{n}}$$

uniformly for λ in a compact subset of C_Π, and thus

$$|c_\hbar(\lambda)|^{-2} \asymp \hbar^{-d}$$

where $d = \dim \mathfrak{a} + \dim \mathfrak{n} = \dim(G/K)$.

We also adjust the Plancherel formula to

$$\|u\|^2_{L^2(\mathbf{S})} = \int |\mathcal{F}_\hbar u(\theta,\lambda)|^2 d\theta |c_\hbar(\lambda)|^{-2} d\lambda. \tag{3.1}$$

From now on, we write for $\theta \in G/P_0$ and $\lambda \in \mathfrak{a}^*$,

$$e(\theta,\lambda) = e_{\theta, i\hbar^{-1}\lambda}$$

and for $x \in G/K$, $\theta \in G/P_0$ and $\lambda \in \mathfrak{a}^*$

$$e(x,\theta,\lambda) = e_{\theta, i\hbar^{-1}\lambda}(x). \tag{3.2}$$

Note that the $\hbar$-scaling of λ is implicit in this notation. For $gM \in G/M$ and $\lambda \in \mathfrak{a}^*$, we write

$$E(gM,\lambda) = e(gK, gP_0, \lambda) \tag{3.3}$$

where $gK \in G/K$ and $gP_0 \in G/P_0$ are the classes of g respectively in G/K and G/P_0. This means that we use the identification $G/M \sim G/K \times G/P_0$ to see the couple of variables (x,θ) in (3.2) as one variable in G/M.

In the calculations of Section 5, we will sometimes write $e(g, g', \lambda)$ $(g, g' \in G,\ \lambda \in \mathfrak{a}^*)$ instead of $e(gK, g'P_0, \lambda)$; and similarly, $E(g,\lambda)$ $(g \in G,\ \lambda \in \mathfrak{a}^*)$ instead of $E(gM,\lambda)$.

3.2. **Pseudodifferential calculus on Y.** The analogue of left-quantization on $\mathbb{R}^n$ in our setting associates to a function a on $G/K \times G/P_0 \times C_\Pi$ the operator which acts on $u \in C_c^\infty(G/K)$ by

$$\mathrm{Op}_\hbar^L(a)\ u(x) = \int_{\theta\in G/P_0, \lambda \in \overline{C_\Pi}} a(x,\theta,\lambda)\, \mathcal{F}_\hbar u(\theta,\lambda) e(x,\theta,\lambda) d\theta |c_\hbar(\lambda)|^{-2} d\lambda\ . \tag{3.4}$$

A similar formula was introduced by Zelditch in [28] (with $\hbar = 1$) in the case $G = SL(2,\mathbb{R})$; it is shown there that $a \mapsto \mathrm{Op}_\hbar^L(a)$ is G-equivariant. The operator $\mathrm{Op}_\hbar^L(a)$ can be defined if a belongs to a nice class of functions, and Zelditch showed that one thus gets a nice pseudodifferential calculus. We give our regularity assumptions on a below. In any case, we shall always require a to be of the form $b \circ \pi$, where b is a function on $T^*(G/K)$ and π is the map defined in (1.1); besides, we will assume that b is supported away from the singular G-orbits in $T^*(G/K)$ (which means that a is supported away from the walls in C_Π). This allows to identify a in a natural way with a function defined on (a subset of) $T^*(G/K)$.

We define *symbols of order* m on $T^*(G/K)$ (independent of $\hbar$) in the usual fashion :

$$\begin{aligned} S^m(G/K) := \big\{ & b \in C^\infty(T^*(G/K)) \ \big| \\ & \text{for every compact } F \subset G/K, \text{ for every } \alpha, \beta, \text{ there exists } C \text{ such that} \\ & |D_z^\alpha D_\xi^\beta b(x,\xi))| \le C(1+|\xi|)^{m-|\beta|} \text{ for all } x \in F,\ \xi \in T_x^*(G/K) \big\}. \end{aligned}$$

We also define *semiclassical symbols of order m and degree l* — thus called because they depend on the parameter $\hbar$:

$$S^{m,l}(G/K)=\{b_\hbar(x,\xi)=\hbar^l\sum_{j=0}^{\infty}\hbar^j b_j(x,\xi),\ b_j\in S^{m-j}\}. \tag{3.5}$$

This means that $b_\hbar(x,\xi)$ has an asymptotic expansion in powers of $\hbar$, in the sense that

$$a-\hbar^l\sum_{j=0}^{N-1}\hbar^j a_j\in \hbar^{l+N}S^{m-N}$$

for all N, uniformly in $\hbar$. In this context, we denote $S^{-\infty,+\infty}=\cap_{m\geq 0}S^{-m,m}$.

Remark 3.2. As indicated above, we define symbols on $G/K\times G/P_0\times C_\Pi$ by transporting the standard definition on $T^*(G/K)$ through the map π (1.1). We will exclusively consider the case when b vanishes in a fixed neighbourhood of the singular G-orbits in $T^*(G/K)$. In other words, b can be identified (through (1.1)) with a function on $G/K\times G/P_0\times C_\Pi$, that vanishes in a neighbourhood of $G/K\times G/P_0\times\partial C_\Pi$. While this restriction was harmless in the original case $G=SL(2,\mathbb{R})$ treated by Zelditch, here it constitutes a restriction on the scope of our result. In particular we need to assume that Λ_∞ is regular in our main theorem 1.6. There would be serious technical difficulties if one wanted to use the formula (3.4) to study the semiclassical limit in a singular direction. In particular one would need to cope with the singularity of the c-function on the walls. Our aim is not to make an exhaustive study of the pseudodifferential calculus defined by (3.4), but only to introduce a class of symbols that works for our purposes.

We now project this construction down to functions on $\mathbf{Y}$. Here we do not follow Zelditch, who defined the action of $\mathrm{Op}_\hbar(a)$ on Γ-invariant functions in a global manner, using the Helgason-Fourier decomposition of such functions. We continue to work locally, which is less elegant but sufficient for our purposes.

We identify the functions on the quotient $\mathbf{Y}=\Gamma\backslash G/K$ (respectively $T^*\mathbf{Y}$) with the Γ–invariant functions on $\mathbf{S}=G/K$ (resp. $T^*(G/K)$). If Γ has torsion, we shall use "smooth function on $\mathbf{Y}$" to mean a Γ-invariant smooth function on $\mathbf{S}$. For a compactly supported function χ on $\mathbf{S}$, we denote $\mathbf{\Pi}_\Gamma\chi(x)=\sum_\gamma\chi(\gamma.x)$. This sum is finite for any $x\in\mathbf{S}$, and hence defines a function on $\mathbf{Y}$.

On $\mathbf{S}$, we fix once and for all a positive, smooth and compactly supported function χ such that $\sum_{\gamma\in\Gamma}\chi(\gamma.x)\equiv 1$. We call such a function a "smooth fundamental cutoff" or a "smooth fundamental domain". Here we have used the assumption that $\mathbf{Y}$ is compact. We also introduce $\tilde\chi\in C_c^\infty(\mathbf{S})$ which is identically 1 on the support of χ. We define the quantization of $a\in S^{m,k}\cap C^\infty(T^*\mathbf{Y})$ (supported away from singular G-orbits) to act on $u\in C^\infty(\mathbf{Y})$ by:

$$\mathrm{Op}_\hbar(a)\,u=\mathbf{\Pi}_\Gamma\tilde\chi\,\mathrm{Op}_\hbar^L(a)\chi u\in C^\infty(\mathbf{Y}). \tag{3.6}$$

We note that for any $D\in\mathcal{D}$ and for any smooth Γ-invariant u on $\mathbf{S}$ we have

$$\mathbf{\Pi}_\Gamma\left(\tilde\chi D\left(\chi u\right)\right)=\mathbf{\Pi}_\Gamma D\left(\chi u\right)=D\mathbf{\Pi}_\Gamma\chi u=Du. \tag{3.7}$$

Thus, (3.7) and Remark 2.3 imply that the formula $\mathrm{Op}_\hbar(H) = \Gamma_{HC}^{-1}[H(-i\hbar\bullet)]$ still holds on the quotient (for $H \in \mathcal{H}$).

Although we will only need it in a marginal way, one can note that our operators belong to the usual classes of pseudodifferential operators, defined using the euclidean Fourier transform in local coordinates. This can be checked by testing the action of $\mathrm{Op}_\hbar(a)$ on a local plane wave of the form $\chi(x)e^{\frac{i\xi.x}{\hbar}}$ in local euclidean coordinates and applying the stationary phase method.

3.3. **Action of $\mathrm{Op}_\hbar(H)$ on WKB states.** Fix a Hamiltonian $H \in \mathcal{H}$. The letter H will stand for several different objects which are canonically related: a function H on $T^*(G/K)$, a W-invariant polynomial function on $\mathfrak{a}^*$, and an element of $U(\mathfrak{a})^W$.

In the following lemma, all functions on G/K and G/M are lifted to functions on G, and in that sense we can apply to them any differential operator on G. If b is a function defined on $G/M = G/K \times G/P_0$, and θ is an element of G/P_0, we denote b_θ the function defined on G/K by $b_\theta(x) = b(x, \theta)$.

Lemma 3.3. *Let $H \in \mathcal{H}$ be of degree $\bar{d}$, and let b be a smooth function on G/M. Fix $\lambda \in \mathfrak{a}^*$. Then, there exist $D_m \in U(\mathfrak{n}_\mathbb{C})U(\mathfrak{a}_\mathbb{C})$ of degree $\leq m$ (depending on λ and on H) such that for any $\theta \in G/P_0$, for any $x \in G/K$,*

$$\mathrm{Op}_\hbar(H)\cdot(b_\theta e(\theta,\lambda))(x) = \left(H(\lambda)b(x,\theta) - i\hbar\,(dH(\lambda)\cdot b)(x,\theta) + \sum_{m=2}^{\bar{d}} \hbar^m D_m b(x,\theta)\right) e(x,\theta,\lambda).$$

On the right H is seen as a function on $\mathfrak{a}^*$, so its differential $dH(\lambda)$ is an element of $\mathfrak{a}$, and it acts as a differential operator of order 1 on G/M. Each operator D_m actually defines a differential operator on G/M.

Proof. By linearity, it is enough to treat the case where $H \in U(\mathfrak{a})^W$ is homogeneous of degree $\bar{d}$. In this case, we have

$$\mathrm{Op}_\hbar(H) = \hbar^{\bar{d}}\,\mathrm{Op}_1(H) = \hbar^{\bar{d}}\Gamma_{HC}^{-1}[H(-i\bullet)].$$

Consider the operator Z related to $D = \mathrm{Op}_1(H)$ by Lemma 2.1. We have

$$\mathrm{Op}_1(H)\cdot(b_\theta e(\theta,\lambda))(x) = Z\cdot(b_\theta.e(\theta,\lambda))(x).$$

In what follows we consider the point $(x,\theta) \in G/K \times G/P_0$. We choose a representative of θ in K (θ is then defined modulo M, but the calculations do not depend on the choice of this representative). We write $x = \theta naK$. This means that the pair $(x,\theta) \in G/K \times G/P_0$ corresponds to the point $\theta naM \in G/M$. All functions on G/K and G/M are lifted to functions on G, and in that sense we can apply to them any differential operator on G.

By Lemma 2.1, we have

$$Z\cdot(b_\theta e(\theta,\lambda))(x) = Z\cdot(b_\theta e(\theta,\lambda))(\theta na) = \tau_{-\rho}H(-i\bullet)\cdot(b_\theta.e(\theta,\lambda))(\theta na) + D\cdot(b_\theta.e(\theta,\lambda))(\theta na)$$

where $D \in U(\mathfrak{n}_\mathbb{C})U(\mathfrak{a}_\mathbb{C})^{\bar{d}-2}$ and $\tau_{-\rho}H(-i\bullet)$ is used as an element of $U(\mathfrak{a}_\mathbb{C}) \subset U(\mathfrak{g}_\mathbb{C})$.

We use the identity

$$e(\theta nag,\theta,\lambda) = e^{(\rho+i\hbar^{-1}\lambda)\mathcal{B}(\theta na)}e^{(\rho+i\hbar^{-1}\lambda)H_0(g)},$$

(valid for any $g \in NA$) where H_0 is defined in §2.1 and $\mathcal{B}$ is defined in (2.1). It shows that, for any $D \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})$, the term $D[e(\theta, \lambda)](\theta na)$ is of the form $Ce(\theta na, \theta, \lambda)$, where the prefactor C depends on D and $\hbar^{-1}\lambda$ but not on θ. This prefactor C is in fact polynomial in $\hbar^{-1}\lambda$.

This results in an expression :

$$Z \cdot (b_\theta e(\theta, \lambda))(x) = Z \cdot (b_\theta e(\theta, \lambda))(\theta na) = \tau_{-\rho} H(-i\bullet) \cdot (b_\theta e(\theta, \lambda))(\theta na) + \left[\sum_{m=0}^{\bar{d}-2} \hbar^{-m} D_{\bar{d}-m} b(\theta na)\right] e(\theta na, \theta, \lambda)$$

where $D_{\bar{d}-m} \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})$ depends only on λ and H.

A term in $\hbar^{-m}$ can only arise if $e(\theta, \lambda)$ is differentiated m times; but Z being of degree $\bar{d}$, we see then that $D_{\bar{d}-m}$ can be of order $\bar{d} - m$ at most. The last term, when multiplied by $\hbar^{\bar{d}}$, becomes $\sum_{m=2}^{\bar{d}} \hbar^m D_m b$. We do not know a priori if the function $D_{\bar{d}-m} b$ (defined on G) is M-invariant, but the sum $\sum_{m=0}^{\bar{d}-2} \hbar^{-m} D_{\bar{d}-m} b$ necessarily is an M-invariant function on G, since all the other terms are. Since $\hbar$ is arbitrary, we see that each D_m must necessarily send an M-invariant function to an M-invariant function.

Finally, we write

$$\begin{aligned} \tau_{-\rho} H(-i\bullet) \cdot (b_\theta e(\theta, \lambda))(\theta na) &= \left(H(-i\bullet) \cdot \left(b_\theta e(\theta, i\hbar^{-1}\lambda - \rho)\right)\right) e(\theta na, \theta, 0) \\ &= \left(\tau_{i\hbar^{-1}\lambda} H(-i\bullet) \cdot b_\theta\right) e(\theta na, \theta, \lambda). \end{aligned}$$

When multiplying by $\hbar^{\bar{d}}$, and using the Taylor expansion of H at λ, we have

$$\hbar^{\bar{d}} \tau_{i\hbar^{-1}\lambda} H(-i\bullet) = H(\lambda) - i\hbar dH(\lambda) + \sum_{m=2}^{\bar{d}} \frac{(-i\hbar)^m}{m!} d^{(m)} H(\lambda).$$

□

Definition 3.4. We will refer to a function of the form $x \mapsto b_\theta(x) e(x, \theta, \lambda)$ as a *WKB state*, using the language of semiclassical analysis.

3.4. **Definition of the symplectic lift.** Let ψ be a $\mathcal{D}$-eigenfunction, of spectral parameter ν. With the notation of Section 2.2, the state ψ satisfies

$$\mathrm{Op}_\hbar(H)\psi = H(-i\hbar\nu)\psi \tag{3.8}$$

for all $H \in \mathcal{H}$.

To ψ we attach a distribution $\tilde{\mu}_\psi$ on $T^*\mathbf{Y}$: for $a \in C_c^\infty(T^*\mathbf{Y})$ set

$$\tilde{\mu}_\psi(a) = \langle \psi, \mathrm{Op}_\hbar(a)\psi \rangle_{L^2(\mathbf{Y})}$$

As described in Section 1 we are trying to classify the weak-* limit points of the sequence of distributions $(\tilde{\mu}_\psi)$ as $\nu \to \infty$. We fix such a limit ("semiclassical measure") μ and a sequence $(\psi_j)_{j\in\mathbb{N}}$ of eigenfunctions such that the corresponding sequence $(\tilde{\mu}_{\psi_j})$ converges weakly-$*$ to μ. We assume that the spectral parameters ν_j go to infinity in the conditions of paragraph 1.2, the limit ν_∞ assumed to be regular. We let $\hbar = \hbar_j = \|\nu_j\|^{-1}$. Writing $\Lambda_j = \hbar_j \Im m(\nu_j)$ we have $\Lambda_j \longrightarrow \Lambda_\infty = \Im m(\nu_\infty) = -i\nu_\infty \in \mathfrak{a}^*$.

3.5. **Other miscellaneous notations.** From now on, we fix a Hamiltonian $H \in \mathcal{H}$. The letter H will stand for several different objects that are canonically related: a function H on $T^*(G/K)$ (G-invariant and polynomial in the fibers of the projection $T^*(G/K) \longrightarrow G/K$), a W-invariant polynomial function on $\mathfrak{a}^*$, an element of $U(\mathfrak{a})^W$.

For any $\Lambda \in \mathfrak{a}^*/W$, we denote $X_\Lambda = dH(\Lambda) \in \mathfrak{a}$. Since Λ is only defined up to an element of W, so is X_Λ. One can assume that $\alpha(X_{\Lambda_\infty}) \geq 0$ for all $\alpha \in \Delta^+$. For simplicity (and without loss of generality), we will also assume that Λ_∞ belongs to the Weyl chamber C_Π defined in §2.1.

We denote by d the dimension of G/K, r the rank, and J the dimension of N (so that $d = r + J$). Let $\tilde{J}$ be the number of roots. We index the positive roots $\alpha_1, \ldots, \alpha_{\tilde{J}}$ in such a way that $\alpha_1(X_{\Lambda_\infty}) \leq \alpha_2(X_{\Lambda_\infty}) \leq \ldots \leq \alpha_{\tilde{J}}(X_{\Lambda_\infty})$. With our previous notation, we have $\alpha_{\tilde{J}}(X_{\Lambda_\infty}) = \chi_{\max}(H)$.

We wish to prove Theorem 1.12 and thus fix some $\mathcal{K} > 0$. We denote by $j_0 = j_0(X_{\Lambda_\infty})$ the largest index j such that $\alpha_j(X_{\Lambda_\infty}) < \frac{1}{2\mathcal{K}}$.

With $w_l \in W$ the long element, we set: $\mathfrak{n}_{\rm fast} = \oplus_{j>j_0}\mathfrak{g}_{\alpha_j}$, $\mathfrak{n}_{\rm slow} = \oplus_{j\leq j_0}\mathfrak{g}_{\alpha_j}$, $\bar{\mathfrak{n}}_{\rm fast} = \oplus_{j>j_0}\mathfrak{g}_{w_l\cdot\alpha_j}$, $\bar{\mathfrak{n}}_{\rm slow} = \oplus_{j\leq j_0}\mathfrak{g}_{w_l\cdot\alpha_j}$ $J_0 = \dim \mathfrak{n}_{\rm slow} = \sum_{j\leq j_0} m_{\alpha_j}$. The spaces $\mathfrak{n}_{\rm fast}$ and $\bar{\mathfrak{n}}_{\rm fast}$ are subalgebras, in fact ideals, in $\mathfrak{n}$, $\bar{\mathfrak{n}}$ respectively; they generate subgroups $N_{\rm fast}, \overline{N}_{\rm fast}$ that are normal in $N, \overline{N}$ respectively. We note that the definition of $N_{\rm fast}, \overline{N}_{\rm fast}$ depends on X_{Λ_∞} and on $\mathcal{K}$. The action of $e^{-tX_{\Lambda_\infty}}$ by conjugacy on the group N is expanding for $t > 0$. The group $N_{\rm fast}$ corresponds to the directions for which the expansion is stronger than $\hbar^{-1/2}$ for $t = \mathcal{K}|\log\hbar|$. In the same way, the action of $e^{tX_{w_l\cdot\Lambda_\infty}}$ by conjugacy on the group $\overline{N}$ is contracting for $t > 0$. The group $\overline{N}_{\rm fast}$ corresponds to the directions for which the contraction is stronger than $\hbar^{-1/2}$ for $t = \mathcal{K}|\log\hbar|$ (this property is used in §5.5).

4. The WKB Ansatz

We now start the proof of Theorem 1.12. We first describe how the operator $\widehat{P}^\phi_{\underline{\omega}}$ (1.9) acts on the WKB states introduced in Definition 3.4. In Section 5, we will use the fact that these states form a (generalized) basis to estimate the norm of the operator.

4.1. **Goal of this section.** Fix a sequence $\underline{\omega} = (\omega_{-T}, \cdots, \omega_{-1}, \omega_0, \cdots \omega_{T-1})$, of length $2T$ such that $T\eta \leq \mathcal{K}|\log\hbar|$. Theorem 1.12 requires us to estimate the norm of the operator $\widehat{P}^\phi_{\underline{\omega}}$ acting on $L^2(\mathbf{Y})$. This operator is the same as $U^{-(T-1)\eta}\mathcal{P}$ where

$$\mathcal{P} = \widehat{P}_{\omega_{T-1}}U^\eta \ldots U^\eta \widehat{P}^{1/2}_{\omega_0}\operatorname{Op}_\hbar(\phi)\widehat{P}^{1/2}_{\omega_0}U^\eta \ldots \widehat{P}_{\omega_{-T+1}}U^\eta\widehat{P}_{\omega_{-T}}, \tag{4.1}$$

Recall that

$$U^t = \exp(i\hbar^{-1}t\operatorname{Op}_\hbar(H))$$

is the propagator of the "Schrödinger equation" generated by $\operatorname{Op}_\hbar(H)$, and that $\eta > 0$ is a fixed time step. In particular, U^t is unitary.

In what follows we estimate the norm of $\mathcal{P}$. To do so, we first describe how $\mathcal{P}$ acts on our Fourier basis $e(\theta, \lambda)$, using the technique of WKB expansion (§4.2). Then, we use the Cotlar-Stein lemma (§5) to estimate as precisely as possible the norm of $\mathcal{P}$.

The sequence $\omega_{-T}, \ldots, \omega_{T-1}$ is fixed throughout this section. Instead of working with functions on $\mathbf{Y}$ we work with functions on G/K that are Γ-invariant. For instance, the operator $\widehat{P}_\omega$ appearing in (4.1) is the multiplication by the Γ–invariant function P_ω. We assume that each connected component of the support of P_ω has diameter smaller than $\frac{\rho}{4}$ where ρ is the injectivity radius of $\mathbf{Y}$. Let Q_ω be a function in $C_c^\infty(\mathbf{S})$ such that $\mathbf{\Pi}_\Gamma Q_\omega = P_\omega$ and such that the support of Q_ω has diameter $\leq \frac{\rho}{4}$. We denote by $\widehat{Q}_\omega$ the corresponding multiplication operator. Finally we introduce Q'_ω in $C_c^\infty(\mathbf{S})$ which is identically 1 on the support of Q_ω and supported in a set of diameter $\frac{\rho}{2}$.

We decompose

$$\mathcal{P} = \mathcal{S}^* \mathcal{U}_\phi \tag{4.2}$$

where

$$\mathcal{U}_\phi = \mathrm{Op}_\hbar(\phi) \widehat{P}_{\omega_0}^{1/2} U^\eta \widehat{P}_{\omega_{-1}} \ldots U^\eta \widehat{P}_{\omega_{-T+1}} U^\eta \widehat{P}_{\omega_{-T}} \tag{4.3}$$

and

$$\mathcal{S} = \widehat{P}_{\omega_0}^{1/2} \ldots U^{-\eta} \widehat{P}_{\omega_{T-2}} U^{-\eta} \widehat{P}_{\omega_{T-1}}. \tag{4.4}$$

4.2. **The WKB Ansatz for the Schrödinger propagator .** We recall some standard calculations, essentially done in [3]. Here the formulae take a special form, due to the fact that the functions $e(\theta, \lambda)$ are eigenfunctions of $\mathrm{Op}_\hbar(H)$.

On $\mathbf{S}$, let us try to solve

$$-i\hbar \frac{\partial \tilde{u}}{\partial t} = \mathrm{Op}_\hbar(H) \tilde{u},$$

in other words

$$\tilde{u}(t) = U^t \tilde{u}(0),$$

with initial condition the WKB state $\tilde{u}(0, x) = a_\hbar(0, x) e(x, \theta, \lambda)$. We only consider $t \geq 0$. We assume that $a_\hbar$ is compactly supported and has an asymptotic expansion in all the C^l norms as $a_\hbar \sim \sum_{m \geq 0} \hbar^m a_m$. We look for approximate solution up to order $\hbar^{\tilde{M}}$, in the form of an Ansatz

$$u(t, x) = e^{\frac{itH(\lambda)}{\hbar}} e(x, \theta, \lambda) a_\hbar(t, x) = e^{\frac{itH(\lambda)}{\hbar}} e(x, \theta, \lambda) \sum_{m=0}^{\tilde{M}-1} \hbar^m a_m(t, x).$$

Let us rather write

$$u(t, x) = e^{\frac{itH(\lambda)}{\hbar}} e(x, \theta, \lambda) a_\hbar(t, x, \theta, \lambda) = e^{\frac{itH(\lambda)}{\hbar}} e(x, \theta, \lambda) \sum_{m=0}^{\tilde{M}-1} \hbar^m a_m(t, x, \theta, \lambda) \tag{4.5}$$

to keep track of the dependence on θ and λ; the pair (x, θ) then represents an element of $G/K \times G/P_0 = G/M$, and a_m can be seen as a function on $\mathbb{R}_+ \times G/M \times \mathfrak{a}^*$. Identifying

powers of $\hbar$, and using Lemma 3.3, we find the conditions:
(4.6)

$$\begin{cases} \frac{\partial a_0}{\partial t}(t,(x,\theta),\lambda) = \left(dH(\lambda)\cdot a_0\right)(t,(x,\theta),\lambda) \quad \text{(0-th transport equation)} \\ \\ \frac{\partial a_m}{\partial t}(t,(x,\theta),\lambda) = \left(dH(\lambda)\cdot a_m\right)(t,(x,\theta),\lambda) + i\sum_{l=2}^{\bar d}\sum_{l+k=m+1} D_l a_k(t,(x,\theta),\lambda) \\ \quad (m\text{-th transport equation}) \,. \end{cases}$$

The equations (4.6) can be solved explicitly by

$$a_0(t,(x,\theta),\lambda) = a_0(0,(x,\theta)e^{tX_\lambda},\lambda),$$

in other words

$$a_0(t) = R(e^{tX_\lambda})a_0(0),$$

where R here denotes the action of A on functions on G/M by right translation, and $X_\lambda = dH(\lambda)$. For $m \geq 1$,

$$a_m(t) = R(e^{tX_\lambda})a_m(0) + \int_0^t R(e^{(t-s)X_\lambda})\left(i\sum_{l=2}^{\bar d}\sum_{l+k=m+1} D_l a_k(s)\right)ds.$$

If we now define u by (4.5), u solves

$$-i\hbar\frac{\partial u}{\partial t} = \mathrm{Op}_\hbar(H)u - e^{\frac{itH(\lambda)}{\hbar}}e(\theta,\lambda)\left[\sum_{l=2}^{\bar d}\sum_{m=\tilde M+1-l}^{\tilde M-1}\hbar^{m+l}D_l a_m(t,\cdot,\theta,\lambda)\right]$$

and thus

$$\begin{aligned} \|u(t)-U^t u(0)\|_{L^2(\mathbf{S})} &\leq \int_0^t \left[\sum_{l=2}^{\bar d}\sum_{m=\tilde M+1-l}^{\tilde M-1}\hbar^{m+l-1}\|D_l a_m(s)\|_{L^2(\mathbf{S})}\right]ds \\ &\leq te^{(2\tilde M+\bar d-2)t\max_{\alpha\in\Delta^+}\alpha(X_\lambda)^-}\left[\sum_{l=2}^{\bar d}\sum_{m=\tilde M+1-l}^{\tilde M-1}\hbar^{m+l-1}\sum_{j=0}^{m}\|a_{m-j}(0)\|_{C^{2j+l}}\right] \\ &\leq Ct\hbar^{\tilde M}e^{(2\tilde M+\bar d-2)t\max_{\alpha\in\Delta^+}\alpha(X_\lambda)^-}\left[\sum_{m=0}^{\tilde M-1}\|a_m(0)\|_{C^{2(\tilde M-m)+\bar d-2}}\right]. \end{aligned}$$

Since D_m belongs to $U(\mathfrak{n}_\mathbb{C})U(\mathfrak{a}_\mathbb{C})$, in the co-ordinates (x,θ) it only involves differentiation with respect to x. We also recall that D_m is of order m. To write the last inequalities we have used the following estimate on the flow $R(e^{tX_\lambda})$ (for $t \geq 0$) :

$$\left\|\frac{d^N}{dx^N}a((x,\theta)e^{tX_\lambda})\right\| \leq e^{-tN\min_{\alpha\in\Delta^+}\alpha(X_\lambda)}\left\|\frac{d^N}{dx^N}a(x,\theta)\right\|$$

and we have denoted $y^- = \max(-y,0)$ for $y\in\mathbb{R}$.

Remark 4.1. This calculation will later on be used only for $\lambda \in \mathrm{supp}(\phi)$, where ϕ is the cut-off function in Theorem 1.12. By assumption ϕ is supported on a neighbourhood of size ϵ of Λ_∞, and by the conventions chosen in §3.5, we have $\alpha(\Lambda_\infty) \geq 0$ for $\alpha \in \Delta^+$. We thus have $\alpha(X_\lambda) \geq -\epsilon$ for all $\alpha \in \Delta^+$ and $\lambda \in \mathrm{supp}(\phi)$. We see that our approximation method makes sense if

$$\hbar^{\tilde{M}} e^{(2\tilde{M}+\bar{d}-2)t\epsilon} \ll 1. \tag{4.7}$$

The integer $\tilde{M}$ can be taken arbitrarily large, it will be fixed at the end of this Section 4 (depending on the $\mathcal{K}$ for which we want to prove Theorem 1.12) and t will be of order $\mathcal{K}|\log \hbar|$. We can choose ϵ (depending on $\mathcal{K}$) such that condition (4.7) holds : take ϵ such that $1 - \bar{d}\mathcal{K}\epsilon > 0$.

Remark 4.2. On the quotient $\mathbf{Y} = \Gamma\backslash\mathbf{S}$, the same method allows to find an approximate solution of $U^t \mathbf{\Pi}_\Gamma u(0)$ in the form $\mathbf{\Pi}_\Gamma u(t)$ (where $\mathbf{\Pi}_\Gamma$ is the periodization operator used in Section 3.2). The same bound

$$\|\mathbf{\Pi}_\Gamma u(t) - U^t \mathbf{\Pi}_\Gamma u(0)\|_{L^2(\mathbf{Y})} \leq Ct\hbar^{\tilde{M}} e^{\epsilon t(2\tilde{M}+\bar{d}-2)} \left[\sum_{m=0}^{\tilde{M}-1} \|a_m(0)\|_{C^{2(\tilde{M}-m)+\bar{d}-2}}\right] \tag{4.8}$$

holds, provided that the projection $\mathbf{S} \longrightarrow \mathbf{Y}$ is injective when restricted to the support of $a_\hbar(t)$. If the support of $a_\hbar(0)$ has diameter strictly smaller than the injectivity radius of $\mathbf{Y}$, this condition will be satisfied in a time interval $t \in [0, T_0]$ with $T_0 > 0$. The time T_0 depend only on λ (and is uniform for λ in the support of ϕ) and, of course, on the Hamiltonian H. In the statement of Theorem 1.12, η being "small enough" means that we must take $\eta < T_0$.

We can iterate the previous WKB construction T times to get the following description of the action of $\mathcal{U}_\phi$ on $\mathbf{\Pi}_\Gamma Q'_{\omega_{-T}} e(\theta, \lambda)$ (the induction argument to control the remainders at each step is the same as in [3] and we will not repeat it here):

Proposition 4.3.
(4.9)

$$\mathcal{U}_\phi \mathbf{\Pi}_\Gamma Q'_{\omega_{-T}} e(\theta, \lambda) = \mathbf{\Pi}_\Gamma \left[e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta, \lambda) A^{(T)}(\bullet, \theta, \lambda) \right] + \mathcal{O}_{L^2(\mathbf{Y})}(\hbar^{\tilde{M}}) \|Q'_{\omega_{-T}} e(\theta, \lambda)\|_{L^2(\mathbf{S})}$$

where

$$A^{(T)}(x, \theta, \lambda) = \sum_{m=0}^{\tilde{M}-1} \hbar^m a_m^{(T)}(x, \theta, \lambda). \tag{4.10}$$

The function $a_0^{(T)}(x, \theta, \lambda)$ *is equal to*

$$a_0^{(T)}(x, \theta, \lambda) = \phi(\lambda)\, P_{\omega_0}^{1/2}(x) P_{\omega_{-1}}((x, \theta)e^{\eta X_\lambda}) P_{\omega_{-2}}((x, \theta)e^{2\eta X_\lambda}) \ldots Q_{\omega_{-T}}((x, \theta)e^{T\eta X_\lambda}), \tag{4.11}$$

where we have lifted the functions P_ω *(originally defined on* G/K*) to* $G/M = G/K \times G/P_0$*. The functions* $a_m^{(T)}$ *have the same support as* $a_0^{(T)}$*. Moreover, if we consider* $a_m^{(T)}$ *as a* *function of (x,θ), that is, as a function on G/M, we have the following bound*

$$\|Z_\alpha^\ell a_m^{(T)}\| \leq C_{m,\ell,Z_\alpha}(T) \sup_{j=0,\dots T}\{e^{-(\ell+2m)j\eta\,\alpha(X_\lambda)}\}$$

if Z_α belongs to $\mathfrak{g}_\alpha$. The prefactor $C_{m,\ell,Z_\alpha}(T)$ is a polynomial in T.

From (4.11), we see that we only need to take into account the energy parameters λ in the support of ϕ, hence ϵ-close to Λ_∞. In particular, with the conventions of §3.5, we have $\alpha(X_{\Lambda_\infty}) \geq 0$ for $\alpha \in \Delta^+$, and $\alpha(X_\lambda) \geq -\epsilon$. Thus, the last bound of the proposition reads

$$\|Z_\alpha^\ell a_m^{(T)}\| \leq C_{m,\ell,Z_\alpha}(T) e^{(\ell+2m)T\eta\,\epsilon}$$

for $\alpha \in \Delta^+$.

We chose η (the time step) small enough to ensure the following: there exists $\gamma = \gamma_{\omega_{-T},\dots,\omega_0} \in \Gamma$ (independent of θ or λ) such that

(4.12)
$$a_0^{(T)}(x,\theta,\lambda) = \phi(\lambda)\, Q_{\omega_0}^{1/2}\circ\gamma^{\;-1}(x) P_{\omega_{-1}}((x,\theta)e^{\eta X_\lambda}) P_{\omega_{-2}}((x,\theta)e^{2\eta X_\lambda})\dots Q_{\omega_{-T}}((x,\theta)e^{T\eta X_\lambda}).$$

This means that the function $a_0^{(T)}(\bullet,\theta,\lambda)$ is supported in a single connected component of the support of $P_{\omega_0}^{1/2}$.

We will also use the following variant:

Proposition 4.4. *Let $\gamma = \gamma_{\omega_{-T},\dots,\omega_0}$.*

$$\mathcal{U}_\phi(Q'_{\omega_{-T}}\circ\gamma \quad e(\theta,\lambda)) = e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta,\lambda)(x) A^{(T)}\circ\gamma\;(x,\theta,\lambda) + \mathcal{O}(\hbar^{\tilde M})\|Q'_{\omega_{-T}}\circ\gamma\; e(\theta,\lambda)\|_{L^2(\mathbf{S})}$$

where

$$A^{(T)}(x,\theta,\lambda) = \sum_{m=0}^{\tilde M-1} \hbar^m a_m^{(T)}(x,\theta,\lambda).$$

Remark 4.5. For the operator $\mathcal{S}$ (4.4), analogous results can be obtained if we replace everywhere λ by $w_l \cdot \lambda$, η by $-\eta$, and the label ω_{-j} by ω_{+j}.

Remark 4.6. Let $u, v \in L^2(\mathbf{Y})$. We explain how the previous Ansatz can be used to estimate the scalar product $\langle v, \mathcal{U}_\phi u\rangle_{L^2(\mathbf{Y})}$ (up to a small error). This is done by decomposing u and v, locally, into a combination of the functions $e(\theta,\lambda)$ (using the Helgason-Fourier transform), and inputting our Ansatz into this decomposition.

In more detail, we note that $P_{\omega_{-T}} = P_{\omega_{-T}}\mathbf{\Pi}_\Gamma Q'^2_{\omega_{-T}}$, so that $\mathcal{U}_\phi u = \mathcal{U}_\phi \mathbf{\Pi}_\Gamma Q'^2_{\omega_{-T}} u$. We use the Helgason-Fourier decomposition to write

$$Q'^2_{\omega_{-T}} u(x) = Q'_{\omega_{-T}}(x) \int_{\theta\in G/P_0, \lambda\in\overline{C_\Pi}} \mathcal{F}_\hbar\left(Q'_{\omega_{-T}} u\right)(\theta,\lambda) e(\theta,\lambda)(x) d\theta |c_\hbar(\lambda)|^{-2} d\lambda.$$

Applying Cauchy-Schwarz, the Plancherel formula (3.1), and the asymptotics of the c-function (Remark 3.1), we note that

$$\int_{\phi(\lambda)\neq 0} \left|\mathcal{F}_\hbar\left(Q'_{\omega_{-T}} u\right)(\theta,\lambda)\right| d\theta |c_\hbar(\lambda)|^{-2} d\lambda = \mathcal{O}(\hbar^{-d/2})\|u\|_{L^2(\mathbf{Y})}.$$

We write

$$(4.13)\quad \langle v, \mathcal{U}_\phi u\rangle_{L^2(\mathbf{Y})} = \left\langle v, \mathcal{U}_\phi \mathbf{\Pi}_\Gamma Q'^2_{\omega_{-T}} u\right\rangle_{L^2(\mathbf{Y})}$$

$$= \int_{\phi(\lambda)\neq 0} \mathcal{F}_\hbar \left(Q'_{\omega_{-T}} u\right)(\theta,\lambda) \left\langle v, \mathcal{U}_\phi \mathbf{\Pi}_\Gamma Q'_{\omega_{-T}} e(\theta,\lambda)\right\rangle d\theta |c_\hbar(\lambda)|^{-2} d\lambda + \mathcal{O}(\hbar^\infty)\|u\|_{L^2(\mathbf{Y})}\|v\|_{L^2(\mathbf{Y})}.$$

We now use Proposition 4.3 to replace $\mathcal{U}_\phi$ by the Ansatz,

$$\begin{aligned}
\left\langle v, \mathcal{U}_\phi \mathbf{\Pi}_\Gamma Q'_{\omega_{-T}} e(\theta,\lambda)\right\rangle_{L^2(\mathbf{Y})} &= \left\langle v, e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta,\lambda) \;\; A^{(T)}(\bullet,\theta,\lambda)\right\rangle_{L^2(\mathbf{S})} + \mathcal{O}(\hbar^{\tilde{M}})\|v\|_{L^2(\mathbf{Y})} \\
&= \left\langle Q'_{\omega_0} \circ \gamma^{-1} \,.\, v,\; e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta,\lambda) \;\; A^{(T)}(\bullet,\theta,\lambda)\right\rangle_{L^2(\mathbf{S})} + \mathcal{O}(\hbar^{\tilde{M}})\|v\|_{L^2(\mathbf{Y})} \\
&= \left\langle Q'_{\omega_0} v,\; e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta,\lambda) \circ\gamma \quad A^{(T)}(\gamma\bullet,\theta,\lambda)\right\rangle_{L^2(\mathbf{S})} + \mathcal{O}(\hbar^{\tilde{M}})\|v\|_{L^2(\mathbf{Y})}
\end{aligned}$$

where $\gamma = \gamma_{\omega_{-T},\ldots,\omega_0}$ is the element of Γ appearing in (4.12). Thus,

(4.14)

$$\langle v, \mathcal{U}_\phi u\rangle_{L^2(\mathbf{Y})} = \int_{\phi(\lambda)\neq 0} \mathcal{F}_\hbar \left(Q'_{\omega_{-T}} u\right)(\theta,\lambda) \left\langle Q'_{\omega_0} v, e^{\frac{iT\eta H(\lambda)}{\hbar}} e(\theta,\lambda)\circ\gamma \quad A^{(T)}(\gamma\bullet,\theta,\lambda)\right\rangle_{L^2(\mathbf{S})} d\theta |c_\hbar(\lambda)|^{-2} d\lambda$$
$$+ \mathcal{O}(\hbar^{\tilde{M}-d/2})\|v\|_{L^2(\mathbf{Y})}\|u\|_{L^2(\mathbf{Y})}.$$

In this last line we see that replacing the exact expression of $\mathcal{U}_\phi$ by the Ansatz induces an error of $\mathcal{O}(\hbar^{\tilde{M}-d/2})\|v\|_{L^2(\mathbf{Y})}\|u\|_{L^2(\mathbf{Y})}$. We will take $\tilde{M}$ very large, depending on the constant $\mathcal{K}$ in Theorem 1.12, so that the error $\mathcal{O}(\hbar^{\tilde{M}-d/2})$ is negligible compared to the bound announced in the theorem.

5. The Cotlar–Stein argument.

We now use the previous approximations of $\mathcal{U}_\phi$ (4.3) and $\mathcal{S}$ (4.4) to estimate the norm of $\mathcal{P}$ (4.1). This is done in a much subtler manner than in [1, 3], because we want to eliminate the slowly expanding/contracting directions.

5.1. **The Cotlar-Stein lemma.**

Lemma 5.1. *Let E, F be two Hilbert spaces. Let $(A_n) \in \mathcal{L}(E,F)$ be a countable family of bounded linear operators from E to F. Assume that for some $R > 0$ we have*

$$\sup_n \sum_\ell \|A_n^* A_\ell\|^{\frac12} \leq R$$

and

$$\sup_n \sum_\ell \|A_n A_\ell^*\|^{\frac12} \leq R$$

Then $A = \sum_n A_n$ converges strongly and A is a bounded operator with $\|A\| \leq R$.

We refer for instance to [7] for the proof.

5.2. **A non-stationary phase lemma.** Let Ω be an open set in a smooth manifold, and consider a vector field Z, a smooth function J and a measure $\mathcal{M}$, all three defined on Ω, with the property that for every smooth compactly supported function f one has $\int_\Omega (Zf) d\mathcal{M} = \int_\Omega f J d\mathcal{M}$.

We require asymptotics for integrals of the form

$$I_\hbar = \int_\Omega e^{\frac{iS(x)}{\hbar}} a(x) d\mathcal{M}(x) \ , \tag{5.1}$$

where S, a are smooth functions on Ω, with a compactly supported.

The following lemma is a variant of integration by parts.

Lemma 5.2. *Let $S \in C^\infty(\Omega, \mathbb{R})$ and $a \in C_c^\infty(\Omega)$. Assume that ZS does not vanish, and let D_Z be the operator*

$$D_Z a = Z\left(\frac{a}{ZS}\right) - \frac{aJ}{ZS}.$$

Then, with $I_\hbar$ as in (5.1) *above, we have*

$$I_\hbar = i\hbar \int e^{\frac{iS(x)}{\hbar}} D_Z a(x) d\mathcal{M}(x) \ ,$$

and hence (iterating n times),

$$I_\hbar = (i\hbar)^n \int e^{\frac{iS(x)}{\hbar}} D_Z^n a(x) d\mathcal{M}(x),$$

where D_Z^n has the form

$$D_Z^n a = \sum_{m \geq n, k+m \leq 2n, \sum l_j \leq n} J_{k,(l_j),m} \frac{Z^k a Z^{l_1} S \ldots Z^{l_r} S}{(ZS)^m}$$

for some smooth functions $J_{k,(l_j),m}(x)$ which only depend on $Z, J, \mathcal{M}$ but not on a or S.

5.3. **Study of several phase functions.** In this paragraph we study the critical points of several functions; this will be useful when applying the stationary phase method.

5.3.1. *Sum of two Helgason phase functions.* We refer to §2.1 for the notation pertaining to the structure of the group G, and in particular for the definition of the function H_0.

Proposition 5.3. *(i) Let $g_1 P_0, g_2 P_0 \in G/P_0$. Let $\lambda, \nu \in \overline{C_\Pi}$ be two elements of the closed nonnegative Weyl chamber. Consider the function on G/K,*

$$gK \mapsto \lambda.H_0(g_1^{-1} gK) + \nu.H_0(g_2^{-1} gK). \tag{5.2}$$

Then, this map has critical points if and only if $\nu = -\operatorname{Ad}(w_1).\lambda$.

(ii) Let $\lambda, \nu \in C_\Pi$ be two (regular) elements of the positive Weyl chamber. Let $g_1 P_0, g_2 P_0 \in G/P_0$, and assume that $g_1^{-1} g_2 \in P_0 w_1 P_0$ (we don't assume here that the conclusion of (i) is satisfied). Write $g_1^{-1} g_2 = b_1 w_1 b_2$ with $b_1, b_2 \in P_0$.

Then, the set of critical points for variations of the form

$$t \mapsto \lambda.H_0(e^{tX} g_1^{-1} gK) + \nu.H_0(e^{tX} g_2^{-1} gK),$$

with $X \in \mathfrak{n}$ is precisely $\{gK, g \in g_1 b_1 A\}$. Moreover, these critical points are non-degenerate.

Remark 5.4. The set of critical points in (ii) is $\{gK, g \in g_1 P_0, gw_l \in g_2 P_0\}$, that is, the flat in G/K determined by the two points $g_1 P_0, g_2 P_0$ by (2.3).

Proof. (i) It is enough to consider the case $g_1 = e$. By the Bruhat decomposition, we know that there exists a unique $w \in W$ such that $g_2 \in P_0 w P_0$, that is, $g_2 = b_1 w b_2$ for some $b_1, b_2 \in P_0$. The map (5.2) has the same critical points as the map

$$gK \mapsto \lambda.H_0(gK) + \nu.H_0(w^{-1}b_1^{-1}gK), \tag{5.3}$$

and they are the images under $gK \mapsto b_1 gK$ of the critical points of

$$gK \mapsto \lambda.H_0(gK) + \nu.H_0(w^{-1}gK). \tag{5.4}$$

For $X \in \mathfrak{a}$ the derivative at $t = 0$ of

$$t \mapsto \lambda.H_0(e^{tX}gK) + \nu.H_0(w^{-1}e^{tX}gK) \tag{5.5}$$

is $\lambda(X) + \nu(\mathrm{Ad}(w^{-1})X)$. Thus, for the map (5.4) to have critical points, we must have

$$\lambda(X) + \nu(\mathrm{Ad}(w^{-1})X) = 0$$

for every $X \in \mathfrak{a}$. Letting X vary over the dual basis to a positive basis of $\mathfrak{a}^*$, we see that $\nu = -\mathrm{Ad}(w).\lambda$ is in the nonnegative Weyl chamber, and this is only possible if $\nu = -\mathrm{Ad}(w_l).\lambda$ where w_l is the long element of the Weyl group (this does not necessarily mean that $w = w_l$ if λ is not regular).

(ii) Here we assume that ν and λ are regular, and that we are in the "generic" case where $g_1^{-1}g_2 \in P_0 w_l P_0$. Starting from (5.4), we now consider variations of the form

$$t \mapsto \lambda.H_0(e^{tX}gK) + \nu.H_0(w_l^{-1}e^{tX}gK) \tag{5.6}$$

for $X \in \mathfrak{n}$. The term $\lambda.H_0(e^{tX}gK)$ is constant, and it remains to deal with $\nu.H_0(w_l^{-1}e^{tX}gK)$. Write $g = w_l anK$, $n \in N, a \in A$, and denote $Y = \mathrm{Ad}(w_l).X \in \bar{\mathfrak{n}}$, $Y' = \mathrm{Ad}(a^{-1})Y$. We have

$$\nu.H_0(w_l^{-1}e^{tX}gK) = \nu.H_0(e^{tY}anK) = \nu(a) + \nu.H_0(e^{tY'}nK) = \nu(a) + \nu.H_0(n^{-1}e^{tY'}nK).$$

Hence

$$\frac{d}{dt}\nu.H_0(e^{tY}anK) = \nu.H_0(\mathrm{Ad}(n^{-1})Y').$$

We see that the set of critical points of (5.6) is the set of those points gK, with $g = w_l anK$ such that n satisfies $\nu.H_0(\mathrm{Ad}(n^{-1})Y') = 0$ for all $Y' \in \bar{\mathfrak{n}}$. Since ν is regular, one can check that this implies $n = e$. This proves the first assertion of (ii).

Finally, assume that we are at a critical point, that is, $gK = aK$ in (5.6). We calculate the second derivative at $t = 0$ of $t \mapsto \nu.H_0(w_l^{-1}e^{tX}aK)$ when $X \in \mathfrak{n}$. We keep the same notation as above for Y and Y'.

Let $U = Y' - \theta(Y') \in \mathfrak{k}$. By the Baker-Campbell-Hausdorff formula, we have

$$e^{tY'} = e^{t\theta(Y') + \frac{t^2}{2}[Y',\theta(Y')] + \mathcal{O}(t^3)}e^{tU} = e^{t\theta(Y')}e^{\frac{t^2}{2}[Y',\theta(Y')] + \mathcal{O}(t^3)}e^{tU}. \tag{5.7}$$

Remember that $\theta(Y') \in \mathfrak{n}$, and that the function H_0 is left-N-invariant. This calculation shows that the second derivative of $t \mapsto \nu.H_0(w_l^{-1}e^{tX}aK)$ is the quadratic form

$$X \mapsto \nu\left([Y', \theta(Y')]\right),$$

where $Y' = \mathrm{Ad}(a^{-1})\,\mathrm{Ad}(w_l).X$. This is a non-degenerate quadratic form if ν is regular. $\square$

5.3.2. *Variations with respect to $\overline{N}$.* In this section we need the decomposition $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m} \oplus \bar{\mathfrak{n}}$. We will denote $\pi_{\mathfrak{n}}$, $\pi_{\mathfrak{a}}$, $\pi_{\bar{\mathfrak{n}}}$ the corresponding projections. We note that $\pi_{\mathfrak{a}} = H_0$, since $\bar{\mathfrak{n}} \subset \mathfrak{n} + \mathfrak{k}$.

We also recall from §3.5 our two decompositions $\bar{\mathfrak{n}} = \sum_{j \leq j_0} \mathfrak{g}_{w_l \cdot \alpha_j} \oplus \sum_{j > j_0} \mathfrak{g}_{w_l \cdot \alpha_j} = \bar{\mathfrak{n}}_{\mathrm{slow}} \oplus \bar{\mathfrak{n}}_{\mathrm{fast}}$ and $\mathfrak{n} = \sum_{j \leq j_0} \mathfrak{g}_{\alpha_j} \oplus \sum_{j > j_0} \mathfrak{g}_{\alpha_j} = \mathfrak{n}_{\mathrm{slow}} \oplus \mathfrak{n}_{\mathrm{fast}}$. The space $\mathfrak{n}_{\mathrm{fast}}$ is an ideal of $\mathfrak{n}$, and we denote the associated (normal) Lie subgroup by N_{fast}.

Lemma 5.5. *Fix $n \in N$ and $a \in A$. Then there exist two neighbourhoods V_1, V_2 of 0 in $\bar{\mathfrak{n}}$, and a diffeomorphism $\Psi = \Psi_{na} \colon V_1 \longrightarrow V_2$ such that*

$$e^{-Y_1} n a e^{Y_2} \in NA, Y_1 \in V_1, Y_2 \in V_2 \iff Y_2 = \Psi(Y_1).$$

Moreover, the differential at 0 of Ψ (denoted Ψ_0') preserves the subalgebra $\bar{\mathfrak{n}}_{slow}$ defined in §3.5. Finally, if we write $e^{-Y} n a e^{\Psi(Y)} = n(Y)a(Y)$ and a_0' for the differential of $a(Y)$ at $Y = 0$, we have

$$a_0'.Y = \pi_{\mathfrak{a}}[\mathrm{Ad}(na)\Psi_0'(Y)].$$

Proof. We apply the implicit function theorem. For $Y_1 = 0$, the differential of $Y_2 \mapsto nae^{Y_2}(na)^{-1}$ at 0 is $Y_2 \mapsto \mathrm{Ad}(na).Y_2$. What we need to check is the equivalence of $\pi_{\bar{\mathfrak{n}}}[\mathrm{Ad}(na).Y_2] = 0$ and $Y_2 = 0$, which is the case since $\mathrm{Ad}(na)$ preserves $\mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m}$. So the existence of Ψ is proved, in addition the differential Ψ_0' is defined by

$$Y = \pi_{\bar{\mathfrak{n}}}[\mathrm{Ad}(na).\Psi_0'.Y]$$

for $Y \in \bar{\mathfrak{n}}$. Since $\mathrm{Ad}(na)$ preserves the space $\mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m} \oplus \bar{\mathfrak{n}}_{\mathrm{slow}}$ (without preserving the decomposition, of course), $\Psi_0'.Y$ must belong to $\bar{\mathfrak{n}}_{\mathrm{slow}}$ if Y does.

The last formula is simply obtained by differentiating $e^{-Y} n a e^{\Psi(Y)} = n(Y)a(Y)$. $\square$

In the next lemma, recall that ϑ is the Cartan involution.

Lemma 5.6. *(i) The set*

$$\{n \in N, H_0(\mathrm{Ad}(n)Y) = 0 \qquad \forall Y \in \vartheta\mathfrak{n}_{slow}\}$$

is, near the identity, a submanifold of N tangent to $\mathfrak{n}_{fast}$.

(ii) Let $\nu \in \mathfrak{a}^*$. For $T = \sum_{\alpha \in \Delta^+} T_\alpha \in \mathfrak{n}$ (with $T_\alpha \in \mathfrak{g}_\alpha$), for $\beta \in \Delta^+$ and $Z_\beta \in \mathfrak{g}_\beta$ we have

$$\nu.H_0\left(\mathrm{Ad}(\exp(T))\vartheta(Z_\beta)\right) = -\langle \nu, \beta \rangle \langle T_\beta, Z_\beta \rangle + o(\|T\|)_{T \longrightarrow 0} \|Z_\beta\|.$$

Proof. The differential of $T \mapsto H_0(\mathrm{Ad}(\exp(T))Y)$ is $T \mapsto H_0([T,Y])$ $(T \in \mathfrak{n})$. Take $Y = \vartheta(Z_\beta)$ for some $\beta \in \Delta^+$ and $Z_\beta \in \mathfrak{g}_\beta$. We have $H_0([T,Y]) = -\langle T_\beta, Z_\beta\rangle H_\beta$ where $H_\beta \in \mathfrak{a}$ is the coroot [14, Ch. VI §5, Prop. 6.52].

In particular, having $H_0([T,Y]) = 0$ for all $Y \in \vartheta\mathfrak{n}_{\text{slow}}$ is equivalent to $T \in \mathfrak{n}_{\text{fast}}$. The first claim follows from the implicit function theorem.

For the second claim, note that $\nu(H_\beta) = \langle \nu, \beta\rangle$. The remainder is uniform over β since there are finitely many roots.

□

5.4. **First decomposition of $\mathcal{P}$.** We want to use the Cotlar-Stein lemma to estimate the norm of the operator $\mathcal{P}$, defined in (4.2). To do so, we will decompose $\mathcal{P}$ into pieces. Our first decomposition of $\mathcal{P}$ is obtained by covering G/P_0 by a finite number of sets Ω_i described below. We use the fact that there is a neighbourhood Ω of eP_0 in G/P_0 that is diffeomorphic to a neighbourhood of e in $\overline{N}$, via the map

$$\begin{array}{ccc} \overline{N} & \longrightarrow & G/P_0 \\ \bar{n} & \mapsto & \bar{n}P_0. \end{array}$$

Using compactness, we can find an open cover of G/P_0 by a finite number of open sets Ω_i such that, for every i, there exists $g_i \in G$ with $\Omega_i \subset g_i\Omega$. Introduce a family of smooth functions χ_{Ω_i} on G/P_0 such that χ_{Ω_i} is supported inside Ω_i and $\sum_i \chi_{\Omega_i} \equiv 1$. We then define the pseudodifferential operators

$$\mathcal{Q}_i u(x) = \int_{k\in G/P_0, \nu\in\overline{C_\Pi}} \mathcal{F}_\hbar u(k, w_l\cdot\nu) Q'_{\omega_0}(x)\chi_{\Omega_i}(k) e(x,k,w_l\cdot\nu)\, dk\, |c_\hbar(\nu)|^{-2} d\nu,$$

and

$$\mathcal{P}_i u = \mathbf{\Pi}_\Gamma \mathcal{S}^* \mathcal{Q}_i^* \mathcal{U}_\phi u \tag{5.8}$$

where $\mathcal{S}$ and $\mathcal{U}_\phi$ were defined in (4.4), (4.3), $\mathbf{\Pi}_\Gamma$ is the periodization operator and Q'_{ω_0} is the cut-off function introduced in §3.2.

Obviously, $\mathcal{P} = \sum_i \mathcal{P}_i$. The sum over i is finite, and we now fix i. The variable k stays in Ω_i.

The reason for using $w_l \cdot \nu$ in the definition of $\mathcal{Q}_i$ – in other words, for working in the negative Weyl chamber – becomes apparent in the following remark, which is only valid for $\nu \in \overline{C_\Pi}$.

Remark 5.7. Let $\gamma = \gamma_{\omega_{T-1},\ldots,\omega_0}$ defined as in (4.12).

Proposition (4.4) (and Remark 4.5) can be generalized by writing

$$\mathcal{Q}_i\mathcal{S}\left(Q'_{\omega_{T-1}}\circ\gamma\ \ e(k,w_l\cdot\nu)\right) = e^{\frac{-iT\eta H(\nu)}{\hbar}} e(k,w_l\cdot\nu) B^{(T)}\circ\gamma\ (x,k,w_l\cdot\nu) \\ + \mathcal{O}_{L^2(\mathbf{S})}(\hbar^{\tilde{M}}) \|Q'_{\omega_T}\circ\gamma\ \ e(k,w_l\cdot\nu)\|_{L^2(\mathbf{S})}, \tag{5.9}$$

where now

$$B^{(T)}(x,k,w_1\cdot\nu)=\sum_{m=0}^{\tilde{M}-1}\hbar^m b_m^{(T)}(x,k,w_1\cdot\nu), \tag{5.10}$$

(5.11)

$$b_0^{(T)}(x,k,w_1\cdot\nu)=\chi_{\Omega_i}(k)P_{\omega_0}^{1/2}(x)P_{\omega_1}((x,k)e^{-\eta X_{w_1\cdot\nu}})P_{\omega_2}((x,k)e^{-2\eta X_{w_1\cdot\nu}})\ldots Q_{\omega_{T-1}}((x,k)e^{-(T-1)\eta X_{w_1\cdot\nu}})$$
$$=\chi_{\Omega_i}(k)Q_{\omega_0}^{1/2}\circ\gamma^{-1}(x)P_{\omega_1}((x,\theta)e^{-\eta X_{w_1\cdot\nu}})P_{\omega_2}((x,k)e^{-2\eta X_{w_1\cdot\nu}})\ldots Q_{\omega_{T-1}}((x,k)e^{-(T-1)\eta X_{w_1\cdot\nu}})$$

and the next terms have the same support as the leading one (their derivatives are bounded the same way as in Proposition 4.3).

In the next paragraphs we will concentrate our attention on brackets of the form:

$$\left\langle Q'_{\omega_{T-1}}\circ\gamma_2\quad e(k,w_1\cdot\nu)\,,\,\mathcal{S}^*\mathcal{Q}_i^*\mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1\quad e(\theta,\lambda)\right\rangle_{L^2(\mathbf{S})}, \tag{5.12}$$

for $\lambda,\nu\in C_\Pi$, $\theta,k\in G/P_0$. We take $\gamma_1=\gamma_{\omega_{-T},\ldots,\omega_0}$ and $\gamma_2=\gamma_{\omega_{T-1},\ldots,\omega_0}$ as defined in (4.12). These are none other than the matrix elements of the operator $\mathcal{P}_i$ (5.8) in the Fourier basis $e(\theta,\lambda)$.

5.5. **Second decomposition of $\mathcal{P}$.** The index i being fixed, we will apply the Cotlar-Stein lemma to bound the norm of $\mathcal{P}_i$ (5.8). We decompose $\mathcal{P}_i$ as a sum of countably many operators, and this decomposition is technically more involved.

We have assumed that we have a diffeomorphism from a relatively compact subset of $\overline{N}$ onto Ω_i: $\bar{n}_1\mapsto g_i\bar{n}_1P_0$. We can write the Haar measure on Ω_i as $dk=\mathrm{Jac}(\bar{n}_1)d\bar{n}_1$, where Jac is a smooth function on $\overline{N}$ (we suppress from the notation its dependence on g_i). Accordingly, for $k=g_i\bar{n}_1P_0\in\Omega_i$, we now write $e(g_i\bar{n}_1,w_1\cdot\nu)$ for $e(k,w_1\cdot\nu)$.

Let us look at the scalar product (5.12), that can also be written as

$$\left\langle\mathcal{Q}_i\mathcal{S}Q'_{\omega_{T-1}}\circ\gamma_2\,e(g_i\bar{n}_1,w_1\cdot\nu),\ \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1\,e(\theta,\lambda)\right\rangle. \tag{5.13}$$

We only need to consider the generic case where $\theta\in g_i\bar{n}_1P_0w_1P_0$, that is, θ is of the form $g_i\bar{n}_1n_1w_1P_0$ (with $n_1\in N$). In addition, we may always assume that λ and ν are regular, since in the end they will have to belong to the support of the cut-off function ϕ. Proposition 5.3 (ii) tells us that the stationary points of the phase function

$$gK\mapsto\lambda.H_0(\theta^{-1}gK)+\nu.H_0(k^{-1}gK),\qquad k=g_i\bar{n}_1P_0$$

with respect to variations

$$(g_i\bar{n}_1n_1)e^{tX}(g_i\bar{n}_1n_1)^{-1}gK,\qquad X\in\mathfrak{n},$$

are the points of the form $gK = g_i\bar{n}_1 n_1 a_1 K$ with $a_1 \in A$. Thus the set of critical points is of codimension J. The stationary phase method then gives:

$$(5.14)\quad \left\langle \mathcal{Q}_i \mathcal{S} Q'_{\omega_{T-1}} \circ \gamma_2\, e(g_i\bar{n}_1, w_l\cdot\nu),\ \mathcal{U}_\phi Q'_{\omega_{-T}} \circ \gamma_1\, e(\theta,\lambda)\right\rangle = \hbar^{J/2}\int_{a_1\in A} d(\lambda,a_1) C_\hbar\left(g_i\bar{n}_1 n_1 a_1 M, \lambda, w_l\cdot\nu\right)\overline{e(g_i\bar{n}_1 n_1 a_1, g_i\bar{n}_1, w_l\cdot\nu)}\; e(g_i\bar{n}_1 n_1 a_1, g_i\bar{n}_1 n_1 w_l, \lambda)\, da_1$$

with an asymptotic expansion valid up to any order,

$$C_\hbar\left(g_i\bar{n}_1 n_1 a_1 M, \lambda, w_l\cdot\nu\right) \sim \sum \hbar^m c_m\left(g_i\bar{n}_1 n_1 a_1 M, \lambda, w_l\cdot\nu\right)$$

and

$$(5.15)\quad c_0\left(g_i\bar{n}_1 n_1 a_1 M, \lambda, w_l\cdot\nu\right) = \left(A^{(T)}\circ\gamma_1(g_i\bar{n}_1 n_1 a_1 w_l M, \lambda)\right)\left(\bar{B}^{(T)}\circ\gamma_2(g_i\bar{n}_1 n_1 a_1 M, w_l\cdot\nu)\right)$$

(and the next terms have the same support as the leading one). The functions $A^{(T)}$ and $B^{(T)}$ are the ones appearing in (4.10) and (5.10). They are functions on $G/M\times\mathfrak{a}^*$; to simplify the notation we will see them as (M-invariant) functions on $G\times\mathfrak{a}^*$.

The term $d(\lambda,a_1)$ is the prefactor involving the hessian of the phase function in the application of the method of stationary phase, it is a smooth function.

Using the notation (3.3), the expression (5.14) can also be written as

$$(5.16)\quad \left\langle \mathcal{Q}_i \mathcal{S} Q'_{\omega_{T-1}} \circ \gamma_2\, e(g_i\bar{n}_1, w_l\cdot\nu),\ \mathcal{U}_\phi Q'_{\omega_{-T}} \circ \gamma_1\, e(\theta,\lambda)\right\rangle = \hbar^{J/2}\int_{a_1\in A} d(\lambda,a_1) C_\hbar\left(g_i\bar{n}_1 n_1 a_1, \lambda, w_l\cdot\nu\right)\overline{E(g_i\bar{n}_1 n_1 a_1, w_l\cdot\nu)}\; E(g_i\bar{n}_1 n_1 a_1 w_l, \lambda)\, da_1.$$

We see from (5.15) that the asymptotics of our scalar product only takes into account the elements $g_i\bar{n}_1 n_1 a_1$ with

$$A^{(T)}\circ\gamma_1(g_i\bar{n}_1 n_1 a_1 w_l, \lambda)\bar{B}^{(T)}\circ\gamma_2(g_i\bar{n}_1 n_1 a_1, w_l\cdot\nu) \neq 0.$$

From (4.10) we see that we must have $\phi(\lambda)\neq 0$.

In the next lemma, we recall that $\Omega_i \subset g_i\Omega$, an open subset of G/P_0 defined at the beginning of §5.4. We fix $\epsilon_0 > 0$ such that the exponential map is a diffeomorphism from $B(0, 10\epsilon_0)$ in $\mathfrak{g}$ onto its image in G.

Lemma 5.8. *Assume that the diameters of Ω and of $\operatorname{supp} Q_{\omega_0}$ are smaller than ϵ_0. Then there exist $n_0\in N$ and $a_0\in A$ such that*

$$B^{(T)}\circ\gamma_2(g_i\bar{n}_1 n_1 a_1, w_l\cdot\nu)\neq 0$$

implies $n_1 a_1 = n_0 a_0 g$, where $g\in NA$ is ϵ_0-close to identity.

Proof. Note from the expression of $B^{(T)} \circ \gamma_2$ that, if it is not 0, we must have

$$g_i \bar{n}_1 n_1 a_1 \in \operatorname{supp} Q_{\omega_0}.$$

The element g_i varies in a finite set and $\bar{n}_1$ varies over Ω which is of diameter $\leq \epsilon_0$. We also assume that $\operatorname{supp} Q_{\omega_0}$ is of diameter $\leq \epsilon_0$, so that n_1 and a_1 must both vary in sets of diameter $\leq \epsilon_0$. In other words, $n_1 a_1$ stays in some ball of diameter $\leq \epsilon_0$ in NA. □

Since $\bar{n}_1$ stays in Ω, it follows that $\bar{n}_1 n_1 a_1 M$ itself is ϵ_0-close to $n_0 a_0 M$ in G/M. From now on we write $g_i \bar{n}_1 n_1 a_1 M = g_i n_0 a_0 g M$, where $g \in G$ varies in a neighbourhood of e of diameter $\leq \epsilon_0$. We will always choose a representative $g \in \exp(\mathfrak{n} \oplus \mathfrak{a} \oplus \bar{\mathfrak{n}})$. By G-equivariance we may assume $g_i n_0 a_0 = 1$, which we do from now on.

In the next proposition, we show that the support of the function c_0 (5.15) is rather small; and especially small in the "fast" expanded or contracted directions.

Proposition 5.9. *(Contracting and expanding foliations)*

(1) *Let ν be such that $\alpha_j(X_\nu) > 0$ for all $\alpha_j \in \Delta^+$ with $j > j_0$ (this is of course the case if ν is close enough to Λ_∞). Suppose we have gM and $g'M$ both ϵ_0-close to eM such that $B^{(T)} \circ \gamma_2 \ (g, w_1 \cdot \nu) \neq 0$ and $B^{(T)} \circ \gamma_2 \ (g', w_1 \cdot \nu) \neq 0$, then $g'^{-1} g = \exp(X + \sum_{\alpha \in \Delta^+} Y_\alpha + \sum_{\alpha \in \Delta^+} Y_{w_1 . \alpha})$ with $X \in \mathfrak{a}, Y_\alpha \in \mathfrak{g}_\alpha$, $\|X\|, \|Y_\alpha\| \leq \epsilon_0$, and $\|Y_{w_1 . \alpha_j}\| \leq \epsilon_0 e^{-T\eta(w_1 . \alpha_j)(X_{w_1 \cdot \nu})} = \epsilon_0 e^{-T\eta \alpha_j(X_\nu)}$ for $j > j_0$.*

(2) *Similarly, assume that $\alpha_j(X_\lambda) > 0$ for all $\alpha_j \in \Delta^+$ with $j > j_0$. Suppose we have g and g' both ϵ_0-close to eM such that $A^{(T)} \circ \gamma_1(g w_1, \lambda) \neq 0$ and $A^{(T)} \circ \gamma_1(g' w_1, \lambda) \neq 0$. Then $g'^{-1} g = \exp(X + \sum_\alpha Y_\alpha)$ with $X \in \mathfrak{a}, Y_\alpha \in \mathfrak{g}_\alpha$, $\|X\|, \|Y_\alpha\| \leq \epsilon_0$, $\|Y_{\alpha_j}\| \leq \epsilon_0 e^{-T\eta\alpha_j(X_\lambda)}$ for $j > j_0$.*

Actually, the claim holds for all j, but we will only use it for $j > j_0$. The other directions will receive a different treatment in §5.6.

Proof. Assume that the term $B^{(T)} \circ \gamma_2 \ (g, w_1 \cdot \nu)$ does not vanish. The evolution equation (5.11) shows that we must have[5]

- $g e^{-(T-1)\eta X_{w_1 \cdot \nu}} M \in \gamma_2^{-1} . \operatorname{supp} Q_{\omega_{T-1}}$;
- $gM \in \operatorname{supp} Q_{\omega_0}$.

If gM and $g'M$ both satisfy the two conditions above, then we see that $g'^{-1} g$ must be ϵ_0-close to identity. Also, $e^{(T-1)\eta X_{w_1 \cdot \nu}} g'^{-1} g e^{-(T-1)\eta X_{w_1 \cdot \nu}}$ must stay in the fixed compact set

$$M [\operatorname{supp} Q_{\omega_{T-1}}]^{-1} \operatorname{supp} Q_{\omega_{T-1}} M \subset G.$$

Let us write $g'^{-1} g = \exp(X + \sum_{\alpha \in \Delta^+} Y_\alpha + \sum_{\alpha \in \Delta^+} Y_{w_1 . \alpha})$ as in part (1) of the claim. Writing the action of A in the co-ordinate system gives the claim. The proof of the second part is similar. □

Finally we write $gM = \bar{n} n a M$ with $\bar{n} \in \overline{N}, n \in N, a \in A$ all ϵ_0-close to identity. We decompose $n = e^Y n_{\text{fast}}$, and $\overline{n} = e^{\overline{Y}} \overline{n}_{\text{fast}}$, $Y \in \mathfrak{n}_{\text{slow}} \simeq \mathbb{R}^{J_0}$, $\overline{Y} \in \bar{\mathfrak{n}}_{\text{slow}} \simeq \mathbb{R}^{J_0}$ both ϵ_0-close to 0 (we fix a vector space isomorphism that sends the root spaces to the coordinate axes of

[5]Here the Q_ω are treated as functions on G/M that factor through G/K.

$\mathbb{R}^{J_0}$); and $n_{\rm fast} \in N_{\rm fast}$, $\bar{n}_{\rm fast} \in \overline{N}_{\rm fast}$ both ϵ_0-close to 1. The quantity ϵ_0 is fixed, but can be chosen as small as we wish. Note that the previous Proposition restricts $n_{\rm fast}$ and $\bar{n}_{\rm fast}$ to sets of measure $\prod_{j>j_0} \epsilon_0 e^{-T\eta m_{\alpha_j}\alpha_j(X_\lambda)}$ and $\prod_{j>j_0} \epsilon_0 e^{-T\eta m_{\alpha_j}\alpha_j(X_\nu)}$, respectively.

On the other hand, in the "slowly" expanded or contracted directions, the support of the function c_0 (5.15) is still of size $> \hbar^{1/2}$ (this is how we chose to define the "slow" directions). This leaves room for breaking the support of c_0 into pieces of diameter $\hbar^{1/2}$ in these directions. The size $\hbar^{1/2}$ is critical in the application of the stationary phase method.

In what follows we break $\mathcal{P}_i$ into countably many pieces along the "slow" directions,

$$\mathcal{P}_i = \sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{J_0}\times\mathbb{Z}^{J_0}\times\mathbb{Z}^r\times\mathbb{Z}^r} \mathcal{P}_{(\bar{y},y,t,\lambda_0)}$$

to which we shall apply the Cotlar-Stein lemma. The dependence on the index i is now suppressed from the notation.

On $\mathbb{R}^\ell$ ($\ell = J_0$ or $\ell = r$) we choose a smooth nonnegative compactly supported function χ^ℓ such that

$$\sum_{y\in\mathbb{Z}^\ell} \chi^\ell(Y-y) \equiv 1 \tag{5.17}$$

and such that $\chi^\ell(Y).\chi^\ell(Y+y) = 0$ for all $Y \in \mathbb{R}^\ell$ and $y \in \mathbb{Z}^\ell$ with $\|y\| > 2$.

Let $(\bar{y}, y) \in \mathbb{Z}^{J_0} \times \mathbb{Z}^{J_0}$ and let $(t, \lambda_0) \in \mathbb{Z}^r \times \mathbb{Z}^r$. Denote 2^+ a fixed real number > 2. Define $\chi^\hbar_{(\bar{y},y)}(\overline{Y}, Y) = \chi^{J_0}(\hbar^{-1/2^+}\overline{Y} - \bar{y})\chi^{J_0}(\hbar^{-1/2^+}Y - y)$; and $\chi^\hbar_{\lambda_0}(\lambda) = \chi^r(\hbar^{-1/2^+}\lambda - \lambda_0)$ and $\chi^\hbar_t(a) = \chi^r(\hbar^{-1/2^+}a - t)$. Also define $\chi^\hbar_{(\bar{y},y,t)}(gM) = \chi^\hbar_{(\bar{y},y,t)}(\overline{Y}, Y)\chi^\hbar_t(a)$ if gM is an element of G/M that is decomposed as $gM = e^{\overline{Y}}\bar{n}_{\rm fast}e^Y n_{\rm fast}aM$, as described above.

We define an operator $\mathcal{S}_{(\bar{y},y,t,\lambda_0)} : L^2(G/K) \longrightarrow L^2(G/K)$ by

$$\mathcal{S}_{(\bar{y},y,t,\lambda_0)}\left(e(k, w_{\rm l}\cdot\nu)\right)(x) \stackrel{\rm def}{=} e^{\frac{-iT\eta H(\nu)}{\hbar}} e(x,k,w_{\rm l}\cdot\nu)\; \chi^\hbar_{(\bar{y},y,t)}(x,k)\,\chi^\hbar_{\lambda_0}(\nu)\,B^{(T)}\!\circ\gamma_2\;(x,k,w_{\rm l}\cdot\nu)$$

for all $k \in G/P_0$, $\nu \in \overline{C_\Pi}$. We then define

$$\mathcal{P}_{(\bar{y},y,t,\lambda_0)} \stackrel{\rm def}{=} \mathbf{\Pi}_\Gamma Q'_{\omega_{T-1}} \circ \gamma_2\; \mathcal{S}^*_{(\bar{y},y,t,\lambda_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}} \circ \gamma_1.$$

We have

$$\|\mathcal{P}_i - \sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{2J_o+2r}} \mathcal{P}_{(\bar{y},y,t,\lambda_0)}\|_{L^2(\mathbf{Y})\longrightarrow L^2(\mathbf{Y})} = \mathcal{O}(\hbar^{\tilde{M}-d/2}),$$

which can be checked by noting that the sum $\sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{2J_o+2r}} \mathcal{S}_{(\bar{y},y,t,\lambda_0)}$ gives back our Ansatz (5.9) for $\mathcal{Q}_i\mathcal{S}$, and by arguing as in (4.13) that the difference between $\mathcal{Q}_i\mathcal{S}$ and the Ansatz is of order $\mathcal{O}(\hbar^{\tilde{M}-d/2})$. Again we choose $\tilde{M}$ large enough so that the error $\mathcal{O}(\hbar^{\tilde{M}-d/2})$ is negligible compared to the bound announced in Theorem 1.12.

The scalar product (5.13) now appears as a sum of the terms

$$\left\langle \mathcal{S}_{(\bar{y},y,t,\lambda_0)} e(\bar{n}P_0, w_{\rm l}\cdot\nu),\; \mathcal{U}_\phi Q'_{\omega_{-T}} \circ \gamma_1\, e(\theta,\lambda)\right\rangle$$

over all $(\bar{y}, y, t, \lambda_0) \in \mathbb{Z}^{2J_o+2r}$. We need only consider the generic case where $\theta \in \bar{n}P_0w_1P_0$, that is, θ is of the form $\theta = \bar{n}nw_1P_0$ (with $n \in N$); the non-generic case has zero measure. It follows from the previous discussions that this scalar product is non-negligible only if $\bar{n}$ and n stay in some sets of diameters $\leq \epsilon_0$; and, without loss of generality, we have assumed they are both ϵ_0-close to identity. As in (5.16), we have by the stationary phase method

$$\begin{aligned}(5.18)\quad &\left\langle \mathcal{S}_{(\bar{y},y,t,\lambda_0)} e(\bar{n}P_0, w_1\cdot\nu),\ \mathcal{U}_\phi Q'_{\omega_{-T}} \circ \gamma_1\, e(\theta,\lambda)\right\rangle \\ &\qquad = \hbar^{J/2} \int_{a\in A} d(\lambda,a)\, C_{\hbar}^{(\bar{y},y,t,\lambda_0)}\left(\bar{n}naM, \lambda, w_1\cdot\nu\right) \overline{E(\bar{n}na, w_1\cdot\nu)} E(\bar{n}naw_1,\lambda) da\end{aligned}$$

where

$$C_{\hbar}^{(\bar{y},y,t,\lambda_0)}\left(\bar{n}na, \lambda, w_1\cdot\nu\right) = \sum \hbar^m c_m\left(\bar{n}na, \lambda, w_1\cdot\nu\right) \tag{5.19}$$

and

$$c_0\left(\bar{n}na, \lambda, w_1\cdot\nu\right) = A^{(T)}\circ\gamma_1(\bar{n}naw_1,\lambda)\bar{B}^{(T)}\circ\gamma_2(\bar{n}na, w_1.\nu)\chi^{\hbar}_{(\bar{y},y,t)}(\bar{n}na)\chi^{\hbar}_{\lambda_0}(\nu) \tag{5.20}$$

The expansion (5.19) holds in all C^ℓ norms over compact sets. The function $c_m\left(\bar{n}na, \lambda, w_1\cdot\nu\right)$ can be expressed in terms of $2m$ derivatives of

$$A^{(T)}\circ\gamma_1(\bar{n}naw_1,\lambda)\bar{B}^{(T)}\circ\gamma_2(\bar{n}n'a', w_1.\nu)\chi^{\hbar}_{(\bar{y},y,t)}(\bar{n}n'a')\chi^{\hbar}_{\lambda_0}(\nu)$$

at $n' = n$, $a' = a$.

Remember the notation $\bar{n} = e^{\overline{Y}}\bar{n}_{fast}$, $n = e^{Y}n_{fast}$. By Proposition 5.9, and by definition of the cut-off functions χ^{J_0}, χ^r, our scalar product is non-negligible only if $\overline{Y}, Y$ stay in a set of measure $\hbar^{J_0/2^+}$, and $n_{fast}, \bar{n}_{fast}$ stay in a set of measure $\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\alpha_j(X_\lambda)}$ and $\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\alpha_j(X_\nu)}$, respectively.

5.6. **Norm of $\mathcal{P}^*_{(\bar{x},x,s,\nu_0)}\mathcal{P}_{(\bar{y},y,t,\lambda_0)}$.** We are now ready to check the first assumption of the Cotlar-Stein lemma, that is, to bound from above the norm of $\mathcal{P}^*_{(\bar{x},x,s,\nu_0)}\mathcal{P}_{(\bar{y},y,t,\lambda_0)}$.

Let $u, v \in L^2(\Gamma\backslash G/K)$. We write

$$\begin{aligned}&\left\langle \mathcal{P}_{(\bar{x},x,s,\nu_0)}v,\ \mathcal{P}_{(\bar{y},y,t,\lambda_0)}u\right\rangle_{\Gamma\backslash G/K} \\ &\quad = \left\langle Q'_{\omega_{T-1}}\circ\gamma_2\mathcal{S}^*_{(\bar{x},x,s,\nu_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, v,\ Q'_{\omega_{T-1}}\circ\gamma_2\mathcal{S}^*_{(\bar{y},y,t,\lambda_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, u\right\rangle_{G/K} \\ &\qquad = \left\langle \mathcal{S}^*_{(\bar{x},x,s,\nu_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, v,\ \mathcal{S}^*_{(\bar{y},y,t,\lambda_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, u\right\rangle_{G/K} + \mathcal{O}(\hbar^\infty)\|u\|\|v\|.\end{aligned}$$

We develop fully this scalar product using the Fourier transform.

$$(5.21)\quad \left\langle \mathcal{S}^*_{(\bar{x},x,s,\nu_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, v, \mathcal{S}^*_{(\bar{y},y,t,\lambda_0)}\mathcal{U}_\phi Q'^2_{\omega_{-T}}\circ\gamma_1\, u\right\rangle_{G/K}$$

$$= \int d\theta d\theta' |c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, u\right)(\theta,\lambda)\overline{\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, v\right)(\theta',\lambda')}$$
$$\left\langle \mathcal{S}^*_{(\bar{x},x,s,\nu_0)}\mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta',\lambda'), \mathcal{S}^*_{(\bar{y},y,t,\lambda_0)}\mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta,\lambda)\right\rangle_{G/K}$$

$$= \int d\theta d\theta' dk |c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'|c_\hbar(\nu)|^{-2}d\nu\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, u\right)(\theta,\lambda)\overline{\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, v\right)(\theta',\lambda')}$$
$$\left\langle \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta',\lambda'), \mathcal{S}_{(\bar{x},x,s,\nu_0)}e(k,w_1\cdot\nu)\right\rangle\left\langle \mathcal{S}_{(\bar{y},y,t,\lambda_0)}e(k,w_1\cdot\nu)\mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta,\lambda)\right\rangle_{G/K}$$

$$= \int d\theta d\theta' \operatorname{Jac}(\bar{n})d\bar{n} |c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'|c_\hbar(\nu)|^{-2}d\nu\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, u\right)(\theta,\lambda)\overline{\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, v\right)(\theta',\lambda')}$$
$$\left\langle \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta',\lambda'), \mathcal{S}_{(\bar{x},x,s,\nu_0)}e(\bar{n}P_0,w_1\cdot\nu)\right\rangle\left\langle \mathcal{S}_{(\bar{y},y,t,\lambda_0)}e(\bar{n}P_0,w_1\cdot\nu), \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta,\lambda)\right\rangle_{G/K}$$

Finally, in equation (5.21), we write $\theta = \bar{n}nw_1P_0$ and $\theta' = \bar{n}n'w_1P_0$ (we can do so on a set of full measure). We have shown in (5.18) that

(5.22)
$$\left\langle \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta',\lambda'), \mathcal{S}_{(\bar{x},x,s,\nu_0)}e(\bar{n}P_0,w_1\cdot\nu)\right\rangle\left\langle \mathcal{S}_{(\bar{y},y,t,\lambda_0)}e(\bar{n}P_0,w_1\cdot\nu), \mathcal{U}_\phi Q'_{\omega_{-T}}\circ\gamma_1 e(\theta,\lambda)\right\rangle_{G/K}$$
$$= \hbar^J \int_{a\in A} d(\lambda,a)\, C_\hbar^{(\bar{y},y,t,\lambda_0)}(\bar{n}naM,\lambda,w_1\cdot\nu)\,\overline{E(\bar{n}na,w_1\cdot\nu)}E(\bar{n}naw_1,\lambda)da$$
$$\int_{a'\in A} d(\lambda',a')\,\bar{C}_\hbar^{(\bar{x},x,s,\nu_0)}(\bar{n}n'a'M,\lambda',w_1\cdot\nu)\,E(\bar{n}n'a',w_1\cdot\nu)\overline{E(\bar{n}n'a'w_1,\lambda')}da'.$$

Already we can note that $C_\hbar^{(\bar{y},y,t,\lambda_0)}(\bar{n}naM,\lambda,w_1\cdot\nu)\,\bar{C}_\hbar^{(\bar{x},x,s,\nu_0)}(\bar{n}n'a'M,\lambda',w_1\cdot\nu)$ can only be non zero if $\chi^\hbar_{(\bar{y},y,t)}(\bar{n}naM)\chi^\hbar_{(\bar{x},x,s)}(\bar{n}n'a'M)\neq 0$, and from the properties of χ^{J_0} this can happen only for $\|\bar{x}-\bar{y}\|\leq 2$. For the same reason, it can only be non zero if $\|\nu_0-\lambda_0\|\leq 2$.

Now we try to show that (5.21) decays fast when $\|x-y\|$ gets large. Using (5.22), the last integral in (5.21) appears as a function of the pair $(\bar{n}na,\bar{n}n'a')$. We have an oscillatory integral of the form (5.1), with a phase

$$(5.23)\quad S(\bar{n}na,\bar{n}n'a') = \lambda.\mathcal{B}(\bar{n}naw_1M) + (w_1\cdot\nu)[\mathcal{B}(\bar{n}n'a') - \mathcal{B}(\bar{n}na)] - \lambda'.\mathcal{B}(\bar{n}n'a'w_1M)$$
$$= \lambda.\mathcal{B}(\bar{n}naw_1M) + (w_1\cdot\nu)[a'-a] - \lambda'.\mathcal{B}(\bar{n}n'a'w_1M),$$

where $\mathcal{B}$ is the function on G/M defined in (2.1) or (2.2). We want to do "integration by parts with respect to $\bar{n}$" and use Lemma 5.2. However, because the derivatives of S with respect to $\bar{n}$ are tricky to compute, it is preferable to use a vector field Z with the property that $Z.\mathcal{B}(\bar{n}naw_1M) = 0$ and $Z.\mathcal{B}(\bar{n}n'a'w_1M) = 0$.

Consider a variation of the form

$$\Psi^\tau : (\bar{n}na,\bar{n}n'a') \mapsto (\bar{n}ne^{\tau Y}a, \bar{n}n'a'a^{-1}e^{\Psi(\tau Y)}a) = \bar{n}n(e^{\tau Y}a, n^{-1}n'a'a^{-1}e^{\Psi(\tau Y)}a),$$

for $Y \in \bar{\mathfrak{n}}$, $\Psi = \Psi_{n^{-1}n'a'a^{-1}}$ defined in Lemma 5.5, and $\tau \in \mathbb{R}$. By definition of Ψ, the two elements $\bar{n}ne^{\tau Y}a$ and $\bar{n}n'a'a^{-1}e^{\Psi(\tau Y)}a$ are in the same NA orbit, for all τ. Such a variation preserves the terms $\mathcal{B}(\bar{n}n'a'w_1M)$ and $\mathcal{B}(\bar{n}naw_1M)$. We call Z the vector field $\frac{d\Psi^\tau}{d\tau}_{|\tau=0}$. We note that each term of the product

$$\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, u\right)(\bar{n}nw_1P_0,\lambda)\overline{\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, v\right)(\bar{n}n'w_1P_0,\lambda')}E(\bar{n}naw_1,\lambda)\overline{E(\bar{n}n'a'w_1,\lambda')}$$

is invariant under Ψ^τ. The function $C_\hbar^{(\bar{y},y,t,\lambda_0)}(\bar{n}naM)\,\bar{C}_\hbar^{(\bar{x},x,s,\nu_0)}(\bar{n}n'aM)$ satisfies

$$\|Z^m C_\hbar^{(\bar{y},y,t,\lambda_0)}(\bar{n}naM)\,\bar{C}_\hbar^{(\bar{x},x,s,\nu_0)}(\bar{n}n'a'M)\| \leq C(m)\hbar^{-m/2^+}, \tag{5.24}$$

by the expression (5.19) and (5.20) of $C_\hbar^{(\bar{y},y,t,\lambda_0)}$ and $C_\hbar^{(\bar{x},x,s,\nu_0)}$: the growth of the derivatives comes from the cut-off $\chi_\hbar$.

Now we want to apply the integration by parts lemma 5.2, so we need to understand $ZS = Z\left((w_1\cdot\nu)(\mathcal{B}(\bar{n}n'a') - \mathcal{B}(\bar{n}na))\right)$.

Lemmas 5.5 and 5.6 tell us that if we write $n^{-1}n' = \exp(T)$ with $T = \sum_{\alpha\geq 0} T_\alpha$, and if Y is chosen such that then $\vartheta\Psi_0'(Y) \in \mathfrak{g}_\beta$ for some $\beta \in \Delta^+$, we have

$$ZS(\bar{n}na, \bar{n}n'a') = -\langle w_1\cdot\nu, \beta\rangle\langle \mathrm{Ad}(a^{-1}a')T_\beta, \vartheta\Psi_0'(Y)\rangle + o(\|T\|)\|\Psi_0'(Y)\|. \tag{5.25}$$

We are interested in those n, n', a, a' such that $C_\hbar^{(\bar{y},y,t,\lambda_0)}(\bar{n}naM)\,\bar{C}_\hbar^{(\bar{x},x,s,\nu_0)}(\bar{n}n'aM) \neq 0$. This implies $\|\log(n^{-1}n') - \hbar^{1/2^+}(x-y)\| \leq 8\hbar^{1/2^+}$. We also have $\|\log(a^{-1}a')\| \leq \epsilon_0$. We thus choose β such that $\|T_\beta\| \geq \frac{1}{2}\left(\hbar^{1/2^+}\|x-y\| - 8\hbar^{1/2^+}\right)$, and Y such that $\vartheta\Psi_0'(Y) = \frac{T_\beta}{\|T_\beta\|}$.

We now apply Lemma 5.2 to the last expression of integral (5.21), integrating by parts $\tilde{N}$-times using the vector field Z. If $\|x-y\| \geq 16$ we gain a factor $\hbar^{\tilde{N}}\left(\hbar^{1/2^+}\|x-y\| - 8\hbar^{1/2^+}\right)^{-\tilde{N}}$ which comes from the non-stationarity of the phase, but we lose a factor $\hbar^{-\tilde{N}/2^+}$ which comes from (5.24). This yields that $\left\langle \mathcal{P}_{(\bar{x},x,s,\nu_0)}v, \mathcal{P}_{(\bar{y},y,t,\lambda_0)}u\right\rangle_{\mathbf{Y}}$ is bounded from above by

$$\frac{C(\tilde{N})\hbar^{\tilde{N}(1-2/2^+)}}{\max(16,\|x-y\|)^{\tilde{N}}}\hbar^J$$
$$\int \mathrm{Jac}(n)dn\,\mathrm{Jac}(n')dn'\,\mathrm{Jac}(\bar{n})d\bar{n}da\,da'\chi^\hbar_{(\bar{y},y,t)}(\bar{n}naM)\chi^\hbar_{(\bar{x},x,s)}(\bar{n}n'a'M)|c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'|c_\hbar(\nu)|^{-2}d\nu$$
$$\chi^\hbar_{\lambda_0}(\nu)\chi^\hbar_{\nu_0}(\nu)\left|\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, u\right)(\bar{n}nw_1P_0,\lambda)\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\, v\right)(\bar{n}n'w_1P_0,\lambda')\right|$$

for an arbitrarily large integer $\tilde{N}$. For any $\bar{n}, n, n'$, just by looking at the size of the support of the function, we have

$$\int da\,da'\chi^\hbar_{(\bar{y},y,t)}(\bar{n}naM)\chi^\hbar_{(\bar{x},x,s)}(\bar{n}n'a'M) = \mathcal{O}(\hbar^{2r/2^+}),$$

so the previous bound becomes

$$\frac{\hbar^{\tilde N(1-2/2^+)}}{\max(16,\|x-y\|)^{\tilde N}}\hbar^J\hbar^{2r/2^+}$$
$$\int \mathrm{Jac}(n)dn\,\mathrm{Jac}(n')dn'\,\mathrm{Jac}(\bar n)d\bar n|c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'|c_\hbar(\nu)|^{-2}d\nu$$
$$\chi^\hbar_{\lambda_0}(\nu)\chi^\hbar_{\nu_0}(\nu)\left|\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,u\right)(\bar n n w_1 P_0,\lambda)\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,v\right)(\bar n n' w_1 P_0,\lambda')\right|$$

We also apply integration by parts with respect to ν, a and a'. Here we do not have to use Lemma 5.2 but just apply integration by parts in its "usual" form : we can note that the phase S (5.23) has the form $S(\bar n n a, \bar n n' a') = (w_1\cdot\lambda)a + (w_1\cdot\nu)[a'-a] - (w_1\cdot\lambda')a' + s(\bar n n,\lambda,\lambda')$ for a function s that does not depend on a, a', ν. We see that $\tilde N$ integrations by parts in (5.21) with respect to the variable ν allows to gain a factor $\frac{\hbar^{\tilde N}}{\|a-a'\|}$ which comes from the non-stationarity of the phase, but we lose a factor $\hbar^{-\tilde N/2^+}$ from (5.24). This yields a gain of $\frac{\hbar^{\tilde N(1-1/2^+)}}{\|a-a'\|^{\tilde N}}$, which is less than $\frac{\hbar^{\tilde N(1-2/2^+)}}{\|t-s\|^{\tilde N}}$ if $\|t-s\|$ is large enough and if $C_\hbar^{(\bar y,y,t,\lambda_0)}(\bar n n a M)\,\bar C_\hbar^{(\bar x,x,s,\nu_0)}(\bar n n' a M)\neq 0$.

Similarly, integrations by parts with respect to a allow to gain a factor $\frac{\hbar^{\tilde N(1-1/2^+)}}{\|\lambda-\nu\|^{\tilde N}}$; and integrations by parts with respect to a' allow to gain a factor $\frac{\hbar^{\tilde N(1-1/2^+)}}{\|\lambda'-\nu\|^{\tilde N}}$. In particular, the contribution to (5.21) of those λ,λ',ν with $\|\lambda'-\nu\|\geq\hbar^{1/2}$ or $\|\lambda-\nu\|\geq h^{1/2}$ is $\mathcal{O}(\hbar^\infty)$.

We find that $\left\langle\mathcal{P}_{(\bar x,x,s,\nu_0)}v,\mathcal{P}_{(\bar y,y,t,\lambda_0)}u\right\rangle_{\mathbf{Y}}$ is bounded from above by

$$\frac{1}{\max(16,\|x-y\|)^{\tilde N}}\frac{1}{\max(16,\|t-s\|)^{\tilde N}}\hbar^J\hbar^{2r/2^+} \tag{5.26}$$
$$\int \mathrm{Jac}(n)dn\,\mathrm{Jac}(n')dn'\,\mathrm{Jac}(\bar n)d\bar n|c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'|c_\hbar(\nu)|^{-2}d\nu$$
$$\chi^\hbar_{\lambda_0}(\nu)\chi^\hbar_{\nu_0}(\nu)\left|\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,u\right)(\bar n n w_1 P_0,\lambda)\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,v\right)(\bar n n' w_1 P_0,\lambda')\right|.$$

In this integral, λ',λ,ν are all ϵ-close to Λ_∞ (they are all in the support of ϕ), and each of them runs over a set of volume $\hbar^{r/2^+}$; $\bar n$ runs over a set of measure $\hbar^{J_0/2^+}\prod_{j>j_0}e^{-T\eta m_{\alpha_j}\alpha_j(X_\nu)}$, n runs over a set of measure $\hbar^{J_0/2^+}\prod_{j>j_0}e^{-T\eta m_{\alpha_j}\alpha_j(X_\lambda)}$, and n' runs over a set of measure $\hbar^{J_0/2^+}\prod_{j>j_0}e^{-T\eta m_{\alpha_j}\alpha_j(X_{\lambda'})}$.

Using Cauchy-Schwarz and the Plancherel formula we find that the integral

$$\int \mathrm{Jac}(n)dn\,\mathrm{Jac}(n')dn'|c_\hbar(\lambda)|^{-2}d\lambda|c_\hbar(\lambda')|^{-2}d\lambda'$$
$$\left|\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,u\right)(\lambda,\bar n n w_1 P_0)\mathcal{F}_\hbar\left(Q'_{\omega_{-T}}\circ\gamma_1\,v\right)(\bar n n' w_1 P_0,\lambda')\right|$$

is bounded by $\hbar^{-d}\hbar^{J_0/2^+}\hbar^{r/2^+}\prod_{j>j_0}e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}\|u\|_{L^2(\mathbf{Y})}\|v\|_{L^2(\mathbf{Y})}$.

The integral $\int \mathrm{Jac}(\bar{n})d\bar{n}|c_\hbar(\nu)|^{-2}d\nu$ adds another factor $\hbar^{-d}\hbar^{J_0/2^+}\hbar^{r/2^+}\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}$. Overall we find that

$$\begin{aligned}\|\mathcal{P}^*_{(\bar{x},x,s,\nu_0)}\mathcal{P}_{(\bar{y},y,t,\lambda_0)}\| \\ \leq \frac{1}{\max(16,\|x-y\|)^{\tilde{N}}}\frac{1}{\max(16,\|t-s\|)^{\tilde{N}}}\hbar^{J+4r/2^+-2d+2J_0/2^+} \\ \prod_{j>j_0} e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}\end{aligned}$$

and it vanishes for $\|\bar{x}-\bar{y}\|>2$ or $\|\nu_0-\lambda_0\|>2$.

Choosing $\tilde{N}$ large enough, we can sum over all $(\bar{y},y,t,\lambda_0)$, and we find

$$\sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{2J_0+2r}}\|\mathcal{P}^*_{(\bar{x},x,s,\nu_0)}\mathcal{P}_{(\bar{y},y,t,\lambda_0)}\|^{1/2}\leq \hbar^{J/2+2r/2^+-d+J_0/2^+}\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}.$$

Remembering that $J=d-r$ and that 2^+ could be chosen arbitrarily close to 2, we get

$$\sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{2J_0+2r}}\|\mathcal{P}^*_{(\bar{x},x,s,\nu_0)}\mathcal{P}_{(\bar{y},y,t,\lambda_0)}\|^{1/2}\leq \hbar^{\frac{J_0-J}{2^-}}\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}$$

where 2^- is smaller than, but arbitrarily close to 2.

5.7. **Norm of $\mathcal{P}_{(\bar{x},x,s,\nu_0)}\mathcal{P}^*_{(\bar{y},y,t,\lambda_0)}$.** Using a similar calculation reversing the roles of $\overline{N}$ and N, we get the same bound,

$$\sum_{(\bar{y},y,t,\lambda_0)\in\mathbb{Z}^{2J_0+2r}}\|\mathcal{P}_{(\bar{x},x,s,\nu_0)}\mathcal{P}^*_{(\bar{y},y,t,\lambda_0)}\|^{1/2}\leq \hbar^{\frac{J_0-J}{2^-}}\prod_{j>j_0} e^{-T\eta m_{\alpha_j}\inf_{\nu\in\mathrm{supp}(\phi)}\alpha_j(X_\nu)}$$

Using the Cotlar-Stein lemma and the fact that the $\alpha(X_\nu)$ coincide with the Lyapunov exponents $\chi_j(H)$ on the energy layer $\mathcal{E}_\nu$, we get Theorem 1.12.

6. Measure Rigidity

In this section we prove Theorems 1.8, 1.9 and 1.10. The proofs combine our entropy bounds with the measure classification results of [8, 9] and the orbit classification results of [17, 26] which give information about A-invariant and ergodic measures that have a large entropy.

Proposition 6.1. *(Measure rigidity theory) Let G be a split group, and let μ be an ergodic A-invariant measure on $\mathbf{X}=\Gamma\backslash G$.*

(1) [8, Lem. 6.2] *there exist constants $s_\alpha(\mu)\in[0,1]$ associated to the roots $\alpha\in\Delta$, such that for any $a\in A$,*

$$h_{KS}(\mu,a)=\sum_{\alpha\in\Delta}s_\alpha(\mu)\left(\log\alpha(a)\right)^+.$$

Here $t^+=\max\{0,t\}$ for $t\in\mathbb{R}$. Furthermore, $s_\alpha(\mu)=1$ if and only if μ is invariant by the root subgroup U_α.

(2) [8, Prop. 7.1] *Assume that $s_\alpha(\mu), s_\beta(\mu) > 0$ for two roots $\alpha, \beta \in \Delta$ such that $\alpha + \beta \in \Delta$. Then $s_{\alpha+\beta}(\mu) = 1$.*

(3) [8, Thm. 4.1(iv)] *If G is locally isomorphic to SL_n and $s_\alpha(\mu) > 0$ for all α, then μ is G-invariant.*

(4) [9, Cor. 3.4] *In the case $G = SL_n$, we have $s_\alpha(\mu) = s_{-\alpha}(\mu)$ for all roots α.*

We do not know if (4) holds in general.

Now let μ be an A-invariant probability measure with ergodic decomposition $\mu = \int_{\mathbf{X}} \mu_x d\mu(x)$. For each subset $R \subset \Delta$ let $\mathbf{X}_R$ be the set of x such that the ergodic component μ_x satisfies $\{\alpha, s_\alpha(\mu_x) > 0\} = R$. Write $w_R = \mu(\mathbf{X}_R)$ and if $w_R > 0$, let $\mu_R = \frac{1}{w_R} \int_{\mathbf{X}_R} \mu_x d\mu(x)$, so that $\mu = \sum_{R \subset \Delta} w_R \mu_R$. Fixing $a \in A$, from Proposition 6.1(1) we have for each R separately that

$$h_{KS}(\mu_R, a) \leq \sum_{\alpha \in R} (\log \alpha(a))^+ ,$$

(this is in fact an avatar of the Ruelle-Pesin inequality). Averaging with respect to R we get

$$h_{KS}(\mu, a) \leq \sum_{R \subset \Delta} w_R \sum_{\alpha \in R} (\log \alpha(a))^+ .$$

By Proposition 6.1(2), $w_R = 0$ unless R is closed under the addition of roots, so we may assume that only such R are included in the sum. In the case $G = SL_n$, parts (4) and (3) show, respectively, that it is enough to consider those R which are symmetric and that $\mu_\Delta = \mu_{Haar}$.

Proposition 6.2. *Let $G = SL_3(\mathbb{R})$, Γ a lattice in G, and μ an A-invariant probability measure on $\Gamma\backslash G$, such that $h_{KS}(\mu, a) \geq \frac{1}{2} h_{KS}(\mu_{Haar}, a)$ for $a = e^X$, $X = \mathrm{diag}(2, -1, -1)$, $\mathrm{diag}(-1, 2, -1)$, and $\mathrm{diag}(-1, -1, 2)$. Then $w_\Delta \geq \frac{1}{4}$, that is, the Haar component has weight at least $\frac{1}{4}$.*

Proof. The possible sets R are $\Delta, \emptyset, \{\alpha, -\alpha\}$. In the case of SL_n the roots are indexed by $\{ij, 1 \leq i, j \leq n, i \neq j\}$: α_{ij} is defined by $\alpha_{ij}(X) = X_{ii} - X_{jj}$. Consider $a = \mathrm{diag}(e^1, e^1, e^{-2})$. Then $h_{KS}(\mu_{Haar}, a) = 6$ (since $s_\alpha = 1$ for all α), $h_{KS}(\mu_\emptyset, a) = 0$, $h_{KS}(\mu_{12}, a) = 0$, $h_{KS}(\mu_{13}, a) \leq 3$, $h_{KS}(\mu_{23}, a) \leq 3$. Thus,

$$3 \leq h_{KS}(\mu, a) \leq 3w_{13} + 3w_{23} + 6w_\Delta. \tag{6.1}$$

This implies

$$w_\Delta - w_{12} \geq 1 - (w_\Delta + w_{12} + w_{13} + w_{23}) \geq 0.$$

By symmetry it follows that $w_\Delta \geq w_{13}$ and $w_\Delta \geq w_{23}$. Returning to (6.1), it follows that $3 \leq 12 w_\Delta$.

In fact, if $h_{KS}(\mu, a) \geq \left(\frac{1}{3} + \epsilon\right) h_{KS}(\mu_{Haar}, a)$ for $a = e^X$, $X = \mathrm{diag}(2, -1, -1)$, $\mathrm{diag}(-1, 2 - 1)$, or $\mathrm{diag}(-1, -1, 2)$, then $w_\Delta \geq \frac{3}{2}\epsilon$. $\square$

Putting together Theorem 1.7 and Proposition 6.2 gives Theorem 1.8.

For SL_4 the analogue of Proposition 6.2 is given below. Theorem 1.9 is an immediate corollary.

Proposition 6.3. *Let $G = SL_4(\mathbb{R})$, μ an A-invariant probability measure on $\Gamma\backslash G$, such that $h_{KS}(\mu, a) \geq \left(\frac{1}{2} + \epsilon\right) h_{KS}(\mu_{Haar}, a)$ for $a = e^X$, X in the Weyl orbit of* $\mathrm{diag}(3, -1, -1, -1)$*. Then $w_\Delta \geq 2\epsilon$. If $\epsilon = 0$ and there is no Haar component, then each ergodic component is the Haar measure on a closed orbit of the group* $\begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}$ *(or one of its 4 images under the Weyl group), and the components invariant by any of these 4 subgroups have total weight $\frac{1}{4}$.*

Theorem 1.8 and its analogue for $G = SL_4$ apply to any lattice Γ. On the other hand for $G = SL_n$ with n large some quotients $\Gamma\backslash G$ support ergodic invariant measures of large entropy other than Haar measure, so our entropy bound is not strong enough to obtain a Haar component. However, for some lattices Γ there are further restrictions on the set of ergodic components, so that non-Haar measures have much smaller entropy. This is the case where Γ is a lattice associated to a division algebra.

We give here a quick outline of the construction, refering the reader to [26] and its references (or [18]) for a detailed discussion. Let F be a central simple algebra of degree n over $\mathbb{Q}$ and assume that F *splits* over $\mathbb{R}$, that is that $F \otimes_{\mathbb{Q}} \mathbb{R} \simeq M_n(\mathbb{R})$. Next, let $\mathcal{O} \subset F$ be an *order*, that is a subring whose additive group is generated by a basis for F over $\mathbb{Q}$. Finally, let $\mathcal{O}^1 \subset SL_n(\mathbb{R})$ denote the subgroup of elements of $\mathcal{O}$ with determinant 1 ("reduced norm 1"). Such $\mathcal{O}^1$ are in fact lattices; any lattice $\Gamma < SL_n(\mathbb{R})$ commensurable with some $\mathcal{O}^1$ is said to be of *inner type*. We simply say that they are *associated* to the algebra F. Our Theorem 1.7 applies when the lattice is co-compact, which is the case if and only if F is a division algebra.

We shall need the fact that those measure rigidity results of [9] which are stated specifically for $SL_n(\mathbb{Z})$ apply, in fact, to any lattice of inner type, since the proof of Lemma 5.2 of that paper carries over to the more general situation. We give the argument here:

Lemma 6.4. *Let $\Gamma < SL_n(\mathbb{R})$ be a lattice of inner type. Then there is no $\gamma \in \Gamma$, diagonalizable in $SL_n(\mathbb{R})$, such that ± 1 are not eigenvalues of γ and all eigenvalues of γ are simple except for precisely one which occurs with multiplicity two.*

Proof. Say that Γ is associated to the central simple algebra F, and let $\mathcal{O}$ be an order in F such that $\Gamma \cap \mathcal{O}^1$ has finite index in Γ.

Assume by contradiction that there exists γ as in the statement, and choose r so that $\gamma^r \in \mathcal{O}$. Since $\mathcal{O}$ is a ring with a finitely generated additive group, the Cayley-Hamilton Theorem shows that γ^r is integral over $\mathbb{Z}$. It follows that every eigenvalue of γ^r, hence of γ, is an algebraic integer. The fact that $\det(\gamma) = 1$ now shows that the rational eigenvalues of γ must be integral divisors of 1, so by assumption all eigenvalues of γ are irrational. Let $f(x) \in \mathbb{R}[x]$ be the characteristic polynomial of γ, when γ is thought of as an element of $SL_n(\mathbb{R})$. We will show $f(x) \in \mathbb{Q}[x]$. Then the multiplicity the eigenvalues of f would be Galois invariant giving the desired contradiction. For the last claim extend scalars to $\mathbb{C}$ and note that the usual proof that the reduced trace and norm belong to $\mathbb{Q}$ applies to the entire characteristic polynomial. $\square$

Proposition 6.5. *Let $n \geq 3$ and let t be the largest proper divisor of n. Let $G = SL_n(\mathbb{R})$ and let $\Gamma < G$ be a lattice of inner type. Let μ be an A-invariant probability measure on $\Gamma\backslash G$ such that $h_{KS}(\mu, a) \geq \frac{1}{2} h_{KS}(\mu_{Haar}, a)$ for $a = e^X$, X a Weyl conjugate of* $\mathrm{diag}(n-1, -1, \cdots, -1)$*. Then $w_\Delta \geq \frac{\frac{(n+1)}{2} - t}{n-t} > 0$. In other words, μ must contain an ergodic component proprtional to Haar measure.*

Theorem 1.10 follows.

Proof. As above, let μ_x be an ergodic component of μ that has positive entropy with respect to e^X. By [9, Thm. 1.3] (replacing Lemma 5.2 of that paper with Lemma 6.4 above) μ_x must be *algebraic*: there exists a closed subgroup H containing A, and a closed orbit zH in $\Gamma\backslash G$, such that μ_x is the H-invariant measure on zH. By [17] (the arguments are contained in the proof of Lemma 4.1 and Lemma 6.2) and [26] (see Thm 1.2 and §4.2), H must be reductive, and conjugate to the connected component of identity in $GL_k(\mathbb{R})^l \cap SL_n(\mathbb{R})$; where $n = kl$ and $GL_k(\mathbb{R})^l$ denotes the block-diagonal embedding of l copies of $GL_k(\mathbb{R})$ into $GL_n(\mathbb{R})$.

By the discussion following Proposition 6.1 we see that for such lattices Γ the possible sets R are obtained by partitioning n into l subsets $B_1, B_2, \ldots, B_l$ of equal size k, and letting

$$R = \{\alpha_{ij}, 1 \leq i, j \leq n, \exists u \text{ such that } i \in B_u \text{ and } j \in B_u\}.$$

Consider $a = \mathrm{diag}(e^{n-1}, e^{-1}, \ldots, e^{-1})$. Then $h_{KS}(\mu_{Haar}, a) = n(n-1)$, and for every subset R defined as above by a non-trivial partition, we have $h_{KS}(\mu_R) \leq n(t-1)$. The inequality $h_{KS}(\mu, a) \geq \frac{1}{2} h_{KS}(\mu_{Haar}, a)$ now shows that

$$w_\Delta(n-1) + \sum_{R \neq \Delta} w_R(t-1) \geq \frac{n-1}{2}.$$

In other words we have

$$w_\Delta(n-1) + (1 - w_\Delta)(t-1) \geq \frac{n-1}{2},$$

which is equivalent to the statement of the theorem. □

References

[1] N. Anantharaman, *Entropy and the localization of eigenfunctions*, Ann. of Math. (2) 168 (2008), no. 2, 435–475.

[2] N. Anantharaman, *Exponential decay for products of Fourier integral operators*, arXiv:1007.2066 (2010), to be published in Methods and Applications of Analysis.

[3] N. Anantharaman, S. Nonnenmacher, *Half-delocalization of eigenfunctions for the Laplacian on an Anosov manifold* Ann. Inst. Fourier (Grenoble) **57**, no. 7, 2465£2523 (2007).

[4] N. Anantharaman, H. Koch and S. Nonnenmacher, *Entropy of eigenfunctions*, New Trends of Mathematical Physics, selected contributions of the 15th International Congress on Mathematical Physics, Springer (2009), 1–22.

[5] J. Bourgain, E. Lindenstrauss, *Entropy of quantum limits,* Comm. Math. Phys. **233**, 153–171 (2003).

[6] U. Bunke, M. Olbrich, *On Quantum Ergodicity for Vector Bundles*, Acta Appl. Math. **90**, 19–41 (2006)

[7] M. Dimassi, J. Sjöstrand, *Spectral asymptotics in the semi-classical limit*, London Mathematical Society Lecture Note Series, 268. Cambridge University Press, Cambridge, 1999.

[8] M. Einsiedler, A. Katok, *Invariant measures on G/Γ for split simple Lie groups G*. Dedicated to the memory of Jürgen K. Moser. Comm. Pure Appl. Math. 56 (2003), no. 8, 1184–1221.

[9] M. Einsiedler, A. Katok, E. Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture.* Ann. of Math. (2) 164 (2006), no. 2, 513–560.

[10] R. Godement, *Introduction à la théorie des groupes de Lie*,Tome 2. Publications Mathématiques de l'Université Paris VII **12** Université de Paris VII, U.E.R. de Mathématiques, Paris, 1982.

[11] S. Helgason, *Groups and geometric analysis. Integral geometry, invariant differential operators, and spherical functions.* Pure and Applied Mathematics, 113. Academic Press Inc., Orlando, 1984.

[12] S. Helgason, *Geometric analysis on symmetric spaces.* Second edition. Mathematical Surveys and Monographs, 39. American Mathematical Society, Providence, RI, 2008. .

[13] J. Hilgert *An ergodic Arnold-Liouville theorem for locally symmetric spaces.* Twenty years of Bialowieza: a mathematical anthology, 163–184, World Sci. Monogr. Ser. Math., 8, World Sci. Publ., Hackensack, NJ, 2005.

[14] A. Knapp, *Lie groups beyond an introduction.* Progress in Mathematics, 140. Birkhäuser Boston, Inc., Boston, MA, 1996.

[15] A. Knapp, *Representation theory of semisimple groups.* An overview based on examples. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1986.

[16] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Annals of Math. **163**, 165-219 (2006)

[17] E. Lindenstrauss, B. Weiss, *On sets invariant under the action of the diagonal group.* Ergodic Theory Dynam. Systems 21 (2001), no. 5, 1481–1500.

[18] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory.* Translated from the 1991 Russian original by Rachel Rowen. Pure and Applied Mathematics, 139. Academic Press, Inc., Boston, MA, 1994. xii+614 pp.

[19] G. Rivière, *Entropy of semiclassical measures in dimension 2*, to appear in Duke Math. J.

[20] G. Rivière, *Entropy of semiclassical measures for surfaces of non-positive curvature*, preprint (2009).

[21] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Commun. Math. Phys. **161**, 195–213 (1994).

[22] M. Schröder, PattersonÐSullivan distributions for symmetric spaces of the noncompact type, PhD dissertation, Universität Paderborn, 2010.

[23] L. Silberman, A. Venkatesh, *On quantum unique ergodicity for locally symmetric spaces,* Geom. Funct. Anal. **17** no. 3, 960–998 (2007).

[24] L. Silberman, A. Venkatesh, *Entropy bounds for Hecke eigenfunctions on division algebras,* to appear in Geom. Funct. Anal.

[25] Soundararajan, *Quantum Unique Ergodicity* Ann. of Math. (2) **172**, no. 2, 1529-£1538 (2010).

[26] G. Tomanov, *Actions of maximal tori on homogeneous spaces.* Rigidity in dynamics and geometry (Cambridge, 2000), 407–424, Springer, Berlin, 2002.

[27] S.A. Wolpert, *The modulus of continuity for $\Gamma_0(m)/\mathbb{H}$ semi-classical limits*, Commun. Math. Phys. **216**, 313–323 (2001)

[28] S. Zelditch, *Pseudodifferential analysis on hyperbolic surfaces,* J. Funct. Anal. **68**, no. 1, 72Ð105 (1986).

Laboratoire de Mathématique, Université d'Orsay Paris XI, 91405 Orsay Cedex, France
E-mail address: Nalini.Anantharaman@math.u-psud.fr

Department of Mathematics, University of British Columbia, Vancouver BC V6T 1Z2, Canada
E-mail address: lior@math.ubc.ca