

LARGE HYPERBOLIC CIRCLES

EMILIO CORSO AND DAVIDE RAVOTTI

ABSTRACT. We consider circles of common centre and increasing radius on a compact hyperbolic surface and, more generally, on its unit tangent bundle. We establish a precise asymptotics for their rate of equidistribution. Our result holds for translates of any circle arc by arbitrary elements of $SL_2(\mathbb{R})$. Our proof relies on a spectral method pioneered by Ratner and subsequently developed by Burger in the study of geodesic and horocycle flows. We further derive statistical limit theorems, with compactly supported limiting distribution, for appropriately rescaled circle averages of sufficient regular observables. Finally, we discuss applications to the classical circle problem in the hyperbolic plane, following the approach of Duke-Rudnick-Sarnak and Eskin-McMullen.

CONTENTS

1. Introduction and main results	1
2. Preliminaries on harmonic analysis on $SL_2(\mathbb{R})$	17
3. Reduction to an ordinary differential equation	22
4. Asymptotics for joint eigenfunctions	25
5. Asymptotics for arbitrary functions	35
6. Distributional limit theorems for deviations from the average	44
7. The hyperbolic lattice point counting problem	50
References	58

1. INTRODUCTION AND MAIN RESULTS

1.1. Dilating sets in diverse geometric contexts. It is an intriguing geometric problem to understand the long-term distribution properties of a diversified range of progressively dilating shapes, when the ambient space in which they live is folded according to a prescribed rule. Formally, the framework underlying such a question can be laid down as follows. Consider a compact connected Riemannian manifold S and a Riemannian covering space N with covering projection $\pi: N \rightarrow S$. The manifold N plays the role of the ambient space, while S is to be interpreted as the manifold resulting after a certain folding procedure operated on N . A homothety on N is a diffeomorphism $h: N \rightarrow N$ transforming the Riemannian metric on N into a scalar multiple thereof; if the rescaling ratio is equal to 1, it simply reduces to a Riemannian isometry. Let now $(h_t)_{t \in \mathbb{R}_{>0}}$ be a collection of homotheties on N whose scaling factor tends to infinity as t does, and fix a Borel probability measure μ on N . The latter should be thought of as describing quantitatively the original shape, whose dilations we wish to examine. For instance, μ might be the renormalized volume measure on a finite-volume Riemannian submanifold of N , e.g. a rectifiable curve. We then let μ_t be the push-forward of μ under the homothety h_t and denote by m_t the projection of μ_t onto S , for any $t > 0$.

A mathematical formulation of our problem consists in asking for the limit points, as t goes to infinity, of the family of measures $(m_t)_{t>0}$ in the topology of weak* convergence of probability measures on S . Somewhat less pretentiously, it is already interesting to determine sufficient geometric conditions on the initial measure μ ensuring that the m_t equidistribute as t grows larger, that is, that they converge in the aforementioned topology to the renormalized volume

measure vol_S on S ; this circumstance amounts pictorially to the fact that the measures m_t (and hence, in an intuitive sense, their supports) fill up the folded space S in the most uniform way.

The Euclidean case. To the best of the authors' knowledge, the question formulated in the previous paragraph was first asked by Dennis Sullivan (cf. [55]) in the zero-curvature setting of Euclidean spaces and tori, that is, when $S = \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ and $N = \mathbb{R}^d$ for some integer $d \geq 1$, and for μ being the volume measure on a compact lower-dimensional submanifold of \mathbb{R}^n . In this case, the transformations h_t are the standard linear homotheties $x \mapsto tx$, $x \in \mathbb{R}^d$. The problem was originally addressed by Randol in [55], both in an Euclidean and in a hyperbolic setup. In the former case, the following equidistribution result is established.

Theorem 1.1 (cf. [55, Thm. 1]). *Suppose C is the smooth boundary of a compact convex subset of \mathbb{R}^d with non-empty interior, and assume its Gaussian curvature is everywhere positive. Let μ be the volume measure on C , renormalized to be a probability measure. Then the probability measures m_t on \mathbb{T}^d defined by*

$$m_t(A) = \mu(\{x \in \mathbb{R}^n : tx + \mathbb{Z}^d \in A\}), \quad A \subset \mathbb{T}^d \text{ Borel}$$

equidistribute towards the Haar measure on \mathbb{T}^d as $t \rightarrow \infty$.

Remark 1.2. (a) The result in Theorem 1.1 is actually quantitative: for every sufficiently regular function f on \mathbb{T}^d , it holds that

$$\left| \int_{\mathbb{T}^d} f \, dm_t - \int_{\mathbb{T}^d} f \, dm_{\mathbb{T}^d} \right| \ll_{f,d} t^{-(d-1)/2},$$

where $m_{\mathbb{T}^d}$ denotes the Haar probability measure on \mathbb{T}^d .

- (b) As a special case of Theorem 1.1, expanding spheres in \mathbb{R}^d equidistribute on the standard torus with a polynomial rate depending on the dimension d . The study of expanding spheres in other geometric settings shall be a driving theme of this manuscript.
- (c) In [55, Thm. 2], equidistribution is generalized to uniform measures supported on lower-dimensional rectilinear simplices in \mathbb{R}^d .

The proof of Theorem 1.1 (more generally, of its quantitative version stated in Remark 1.2(a)) relies on classical Fourier analysis on the d -dimensional torus; it is remarkably straightforward, once the decay properties at infinity of the Fourier transform of the measure μ are known. As such, it has been elaborated upon by Strichartz in [73] to prove the following generalization of Theorem 1.1. Let us say that the Fourier transform $\hat{\mu}: \mathbb{R}^d \rightarrow \mathbb{C}$ of μ decays on rays if

$$\lim_{t \rightarrow \infty} \hat{\mu}(tx) = 0 \tag{1.1}$$

for every nonzero vector $x \in \mathbb{R}^d$.

Theorem 1.3 (cf. [73, Lem. 1]). *Let μ be a Borel probability measure on \mathbb{R}^d whose Fourier transform decays on rays. Then the conclusion of Theorem 1.1 holds true.*

It actually suffices that $\hat{\mu}$ decays on integral rays, that is, that (1.1) is verified, less restrictively, for any non-zero $x \in \mathbb{Z}^d$. On the other hand, it turns out (cf. [73, Lem. 1]) that decay on arbitrary rays is equivalent to equidistribution of the projections of the μ_t onto any torus \mathbb{R}^d/Λ , where Λ is a lattice in \mathbb{R}^d .

To conclude this brief account of the state of the art on the problem in flat geometry, we mention that the question of the limiting distribution of expanding circles has been recently examined in the setting of translation surfaces by Colognese and Pollicott [14], who prove (non-effective) equidistribution towards a probability measure which is equivalent, though in general not proportional, to the area measure on the given surface.

The hyperbolic case. As already mentioned, Randol's investigations in [55] were not confined to the zero-curvature case. In the hyperbolic framework, namely when the sectional curvature is constantly equal to -1 , S is a compact connected hyperbolic d -manifold ($d \geq 2$) and N can be taken as its Riemannian universal covering manifold, that is, the d -dimensional hyperbolic space \mathbb{H}^d . A choice of the homotheties $(h_t)_{t>0}$ is determined by fixing a base point $x_0 \in \mathbb{H}^d$ and letting h_t be the map which transforms¹ each Riemannian geodesic $\gamma(s)$ ($s \in \mathbb{R}$) passing through x_0 at time 0 into the geodesic $\gamma(ts)$. In this context, the result that can be elicited from the discussion in [55] reads as follows.

Theorem 1.4 (cf. [55]). *Let S be a compact connected hyperbolic d -manifold, $\pi: \mathbb{H}^d \rightarrow S$ the universal covering map, C a $(d-1)$ -dimensional hyperbolic sphere of unit radius centered at a point $x_0 \in \mathbb{H}^d$, μ the unique isometrically-invariant² Borel probability measure on C , $(h_t)_{t>0}$ the family of homotheties $\mathbb{H}^d \rightarrow \mathbb{H}^d$ with center x_0 defined as above. Then the probability measures m_t on S defined by*

$$m_t(A) = \mu(\{x \in \mathbb{H}^d : \pi \circ h_t(x) \in A\}), \quad A \subset S \text{ Borel} \quad (1.2)$$

equidistribute towards the renormalized volume measure on S as $t \rightarrow \infty$.

Remark 1.5. Here again the result takes on a quantitative form: the rate of equidistribution of the measures m_t defined in (1.2) is exponential, as opposed to the Euclidean case, with the exponent depending on the spectral gap of the hyperbolic manifold S (cf. [55]).

Akin in spirit to the proof of Theorem 1.1, the argument leading to Theorem 1.4 is based upon the harmonic analysis of locally symmetric spaces via techniques related to the Selberg trace formula; for those, the reader is referred to Selberg's original article [63].

Lifting the question to unit tangent bundles. Let us now consider the following upgraded version of the problem explored in the foregoing paragraphs. Suppose that S , N and $(h_t)_{t>0}$ are as in the beginning of the present section, with $y_0 \in N$ being the common center of the homotheties h_t , and let C be a compact Riemannian hypersurface in N . Assume further that the Riemannian distance function on N turns it into a uniquely geodesic metric space³. If T^1N denotes the unit tangent bundle of N , then C identifies uniquely the subset \tilde{C} of T^1N consisting of all pairs (x, v) , where x is a point in C and v is the unit-length tangent vector to the unique geodesic connecting y_0 to x . Similarly, if C_t indicates the image of C under the homothety h_t for any $t > 0$, we denote by \tilde{C}_t its lift to T^1N obtained in the previous fashion. A natural question thus arises as to the asymptotic distribution of the \tilde{C}_t when projected to the unit tangent bundle of S ; in this respect, the natural choice of a measure on \tilde{C}_t is given by the push-forward of the renormalized volume measure on C_t under the canonical identification of the latter submanifold with its lift \tilde{C}_t . If the projections to S of the hypersurfaces C_t equidistribute towards the normalized volume measure on S , it may be expected that the projections of the lifts \tilde{C}_t equidistribute towards the corresponding Liouville measure on the unit tangent bundle T^1S (cf. [35, Part 1, Sec. 5.4]).

The question lends itself to a description in the language of smooth dynamical systems. If $(g_t^{(N)})_{t \in \mathbb{R}}$ is the geodesic flow on T^1N (cf. [35, Part 1, Chap. 5]), it takes a few moments to realize that, for any $t > 1$, the set \tilde{C}_t is the image of the original lift \tilde{C} under the transformation $g_{t-1}^{(N)}$; the same relation holds for the natural measures carried by the latter sets, and carries over to their projections to T^1S , using the geodesic flow $(g_t^{(S)})_{t \in \mathbb{R}}$ defined on it in place of $(g_t^{(N)})_{t \in \mathbb{R}}$.

The equidistribution problem in this formulation is treated in Margulis' thesis [48], which contains several striking developments and applications of the theory of Anosov systems to

¹This produces a well-defined assignment, as \mathbb{H}^d is a uniquely geodesic metric space (cf. [10, Part I, Chap. 1]) for the distance induced by the hyperbolic Riemannian metric.

²Here we obviously intend invariance under isometries of C equipped with the induced hyperbolic metric.

³This is certainly the case for $N = \mathbb{R}^d$ and $N = \mathbb{H}^d$, equipped respectively with the standard Euclidean metric and with the hyperbolic metric.

the large-scale geometry of negatively curved manifolds; among those, a proof is provided of equidistribution, towards the Liouville measure, of lifts of expanding circles on finite-volume hyperbolic surfaces. For further comments thereupon, as well as for the connection to the hyperbolic circle problem, the reader is referred to Section 1.5.

It is the chief purpose of the present work to provide a precise asymptotic expansion for the equidistribution rate of lifts of dilating hyperbolic circles, as well as of arbitrary sub-arcs thereof, on unit tangent bundles of compact hyperbolic surfaces; in the vein of the works of Randol [55] and Strichartz [73], we resort to a spectral approach originating in the work of Ratner [57] on quantitative mixing of geodesic and horocycle flows on Riemann surfaces of finite volume. Section 1.3 describes such results and expands on their connection to previous developments, whereas Sections 1.4 and 1.5 discuss a number of applications to statistical limit theorems and the hyperbolic lattice point counting problem.

We conclude this historical overview by pointing out that Margulis' groundbreaking contributions in [48], together with the gradual emergence of conspicuous applications to counting and Diophantine problems, spawned intensive research aimed at understanding the asymptotic distribution properties of translates of finite-volume subgroup orbits, as well as of more general subsets, on homogeneous spaces⁴; without purporting to provide an exhaustive list, we mention in this direction the works [5, 22, 24, 25, 39, 65, 66, 67, 68, 80].

1.2. The setup: circles in hyperbolic surfaces and in their unit tangent bundles.

We now set the stage for the main questions we address in the present manuscript, referring to Section 2 for the required background. Let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be a cocompact lattice, that is, a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ such that the quotient space $\Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ is compact; we indicate the latter homogeneous space with M . The group Γ acts properly discontinuously and isometrically on the Poincaré upper half-plane $\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$, endowed with the standard hyperbolic Riemannian metric, by Möbius transformations; when the projection of Γ to $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I_2\}$ is torsion-free, the quotient $S = \Gamma \backslash \mathbb{H}$ is a compact connected orientable smooth surface, inheriting a hyperbolic metric from \mathbb{H} . With respect to such a metric, there is a canonical identification of M with (possibly, a double cover of) the unit tangent bundle⁵ T^1S .

Let $(r_s)_{s \in \mathbb{R}}$ be the one-parameter flow on M defined by

$$r_s(\Gamma g) = \Gamma g \begin{pmatrix} \cos s/2 & \sin s/2 \\ -\sin s/2 & \cos s/2 \end{pmatrix} = \Gamma g \exp s\Theta, \quad \Theta = \begin{pmatrix} 0 & 1/2 \\ -1/2 & 0 \end{pmatrix} \in \mathfrak{sl}_2(\mathbb{R}) = \mathrm{Lie}(\mathrm{SL}_2(\mathbb{R})), \quad (1.3)$$

and denote by $(\phi_t^X)_{t \in \mathbb{R}}$ the geodesic flow on M , which is given algebraically by

$$\phi_s^X(\Gamma g) = \Gamma g \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} = \Gamma g \exp tX, \quad X = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \in \mathfrak{sl}_2(\mathbb{R}). \quad (1.4)$$

For any point $p = \Gamma g \in M$, the orbit of p under the flow $(r_s)_{s \in \mathbb{R}}$ is the preimage of $z = \pi(p)$ under the fibration $\pi: M \rightarrow S$. Therefore, if M identifies with T^1S , then this set consists of all unit tangent vectors to $z \in S$. For any real number $t > 0$, the time- t geodesic evolution $\phi_t^X(\{r_s(p) : s \in \mathbb{R}\})$ of the previous set coincides with the projection to M of the subset of $T^1\mathbb{H}$ given by all outward-pointing normal vectors to the hyperbolic circle in \mathbb{H} of radius t centered at (a fixed representative in \mathbb{H} of) z .

We indicate with vol the Haar probability measure on M , that is, the unique $\mathrm{SL}_2(\mathbb{R})$ -invariant Borel probability measure on M ; under the identification of M with T^1S , it coincides with

⁴It is worth noticing at this point that lifts of expanding hyperbolic circles represent a particular instance, as they are geodesic translates of orbits of the maximal compact subgroup $\mathrm{SO}_2(\mathbb{R})$ on quotients of $\mathrm{SL}_2(\mathbb{R})$: see Section 1.3.

⁵More precisely, if Γ is the preimage under the canonical projection map $\mathrm{SL}_2(\mathbb{R}) \rightarrow \mathrm{PSL}_2(\mathbb{R})$ of a cocompact lattice in $\mathrm{PSL}_2(\mathbb{R})$, then M identifies with T^1S ; else, it is a double cover thereof.

In the case where the image of Γ in $\mathrm{PSL}_2(\mathbb{R})$ has non-trivial torsion elements, then S has the structure of an orbifold (cf. [54, Chap. 13]). For the purposes of the paper, we shall never be concerned with this distinction.

the Liouville measure projecting to the normalized hyperbolic area measure on S . For any $r \in \mathbb{N} \cup \{\infty\}$, we denote by $\mathcal{C}^r(M)$ the set of complex-valued functions of class \mathcal{C}^r defined on the smooth manifold M . The supremum norm of a continuous function $f: M \rightarrow \mathbb{C}$ is denoted by $\|f\|_\infty$. For any $m \in \mathbb{N}_{\geq 1}$, $f \in \mathcal{C}^m(M)$ and $j \in \{0, \dots, m\}$ let $\nabla^j f$ be the j^{th} covariant derivative of f and $|(\nabla^j f)(p)|$ its norm at a point $p \in M$. Define then the \mathcal{C}^m -norm of f (cf. [2, Chap. 1]) as

$$\|f\|_{\mathcal{C}^m} = \sum_{j=0}^m \sup_{p \in M} |(\nabla^j f)(p)|. \quad (1.5)$$

Let $L^2(M)$ be the Hilbert space of complex-valued functions on M whose modulus is square-integrable with respect to the measure vol , and denote by

$$\langle \phi, \psi \rangle = \int_M \phi \bar{\psi} \, d \text{vol}$$

the inner product of two elements $\phi, \psi \in L^2(M)$.

Define two additional elements

$$U = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

in the Lie algebra $\mathfrak{sl}_2(\mathbb{R})$. Identifying elements of the universal enveloping algebra of $\mathfrak{sl}_2(\mathbb{R})$ with left-invariant differential operators on the space $\mathcal{C}^\infty(M)$, we define the Casimir operator as the second-order differential operator $\square = -X^2 + X - UV: \mathcal{C}^2(M) \rightarrow \mathcal{C}^0(M)$. It admits a unique maximal extension to an unbounded self-adjoint operator on $L^2(M)$; in particular, its spectrum $\text{Spec}(\square)$ consists of real numbers. As M is compact, it is well-known that $\text{Spec}(\square)$ is pure point, and is a discrete subset of \mathbb{R} . The elements of $\text{Spec}(\square)$ classify the irreducible representations strongly contained in the Koopman representation arising from the measure-preserving action of $\text{SL}_2(\mathbb{R})$ on the measure space (M, vol) , as belonging to the principal, complementary or discrete series representations if the corresponding eigenvalue μ satisfies, respectively, $\mu \geq 1/4$, $0 < \mu < 1/4$, $\mu \leq 0$. With the normalization we have chosen⁶, the action of \square on \mathcal{C}^2 -functions defined on the surface S is given by the Laplace-Beltrami operator Δ_S associated to the hyperbolic structure on S .

We are interested in quantitative equidistribution properties of the uniform probability measures supported on the circle arcs

$$\phi_t^X(\{r_s(p) : 0 \leq s \leq \theta\})$$

as t goes to infinity, for every fixed $p \in M$ and $\theta \in (0, 4\pi]$ (cf. our normalization of Θ in (1.3), a full circle corresponds to $\theta = 4\pi$). For any parameter $\theta \in (0, 4\pi]$ and any continuous function $f: M \rightarrow \mathbb{C}$, we thus define the function $k_{f,\theta}: M \times \mathbb{R} \rightarrow \mathbb{C}$ as

$$k_{f,\theta}(p, t) := \frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds, \quad p \in M, t \in \mathbb{R}. \quad (1.6)$$

In the forthcoming subsection, we provide a precise asymptotic expansion of $k_{f,\theta}(p, t)$ as t tends to infinity, first for joint eigenfunctions⁷ of the operators \square and Θ (Theorem 1.6), and then for arbitrary functions fulfilling a suitable regularity condition (Theorem 1.8).

⁶Since $\mathfrak{sl}_2(\mathbb{R})$ is a simple Lie algebra, Casimir elements in its universal enveloping algebra are uniquely determined up to real scalars.

⁷As we shall explain in Section 2.2, there exists an orthonormal basis of $L^2(M)$ consisting of such joint eigenfunctions; Theorem 1.6 is thus to be regarded as a building block for the more general Theorem 1.8.

1.3. Quantitative equidistribution of expanding translates of circle arcs. We begin with the case of joint eigenfunctions of the operators \square and Θ . Observe that the left-invariant vector field $\Theta \in \mathfrak{sl}_2(\mathbb{R})$ acts as an unbounded skew-symmetric operator on $L^2(M)$; if $\Theta f = \lambda f$ for some $f \in \mathcal{C}^1(M)$ and $\lambda \in \mathbb{C}$, then $\lambda = \frac{i}{2}n$ for some $n \in \mathbb{Z}$.

In the following statement and throughout, we associate to each $\mu \in \text{Spec}(\square)$ the unique complex number $\nu \in \mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ satisfying $1 - \nu^2 = 4\mu$.

Theorem 1.6. *There exist real constants κ_0 and $\kappa(\mu)$ for any positive Casimir eigenvalue μ such that the following assertions hold. Let $\theta \in (0, 4\pi]$, $\mu \in \text{Spec}(\square)$ and $n \in \mathbb{Z}$, and suppose $f \in \mathcal{C}^2(M)$ satisfies $\square f = \mu f$, $\Theta f = \frac{i}{2}nf$. Define $k_{f,\theta}(p, t)$ as in (1.6).*

- (1) *If $\mu > 1/4$, there exist Hölder-continuous functions $D_{\theta,\mu,n}^+ f, D_{\theta,\mu,n}^- f: M \rightarrow \mathbb{C}$ with Hölder exponent $1/2$ and*

$$\|D_{\theta,\mu,n}^\pm f\|_\infty \leq \frac{\kappa(\mu)}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1}$$

such that, for every $p \in M$ and $t \geq 1$,

$$k_{f,\theta}(p, t) = e^{-\frac{t}{2}} \cos\left(\frac{\Im\nu}{2}t\right) D_{\theta,\mu,n}^+ f(p) + e^{-\frac{t}{2}} \sin\left(\frac{\Im\nu}{2}t\right) D_{\theta,\mu,n}^- f(p) + \mathcal{R}_{\theta,\mu,n} f(p, t), \quad (1.7)$$

where

$$|\mathcal{R}_{\theta,\mu,n} f(p, t)| \leq \frac{8\kappa_0(n^2 + 1)}{\theta \Im\nu} \|f\|_{\mathcal{C}^1} e^{-t}. \quad (1.8)$$

- (2) *If $\mu = 1/4$, there exist Hölder-continuous functions $D_{\theta,1/4,n}^+ f: M \rightarrow \mathbb{C}$, with Hölder exponent $1/2 - \varepsilon$ for every $\varepsilon > 0$, and $D_{\theta,1/4,n}^- f: M \rightarrow \mathbb{C}$, with Hölder exponent $1/2$, and satisfying*

$$\|D_{\theta,1/4,n}^\pm f\|_\infty \leq \frac{\kappa(1/4)}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1},$$

such that, for every $p \in M$ and $t \geq 1$,

$$k_{f,\theta}(p, t) = e^{-\frac{t}{2}} D_{\theta,1/4,n}^+ f(p) + t e^{-\frac{t}{2}} D_{\theta,1/4,n}^- f(p) + \mathcal{R}_{\theta,1/4,n} f(p, t), \quad (1.9)$$

where

$$|\mathcal{R}_{\theta,1/4,n} f(p, t)| \leq \frac{4\kappa_0}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1} (t + 1) e^{-t}.$$

- (3) *If $0 < \mu < 1/4$, there exist functions $D_{\theta,\mu,n}^+ f, D_{\theta,\mu,n}^- f: M \rightarrow \mathbb{C}$, respectively Hölder-continuous with Hölder exponent $\frac{1-\nu}{2}$ and of class \mathcal{C}^1 , and satisfying*

$$\|D_{\theta,\mu,n}^\pm f\|_\infty \leq \frac{\kappa(\mu)}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1},$$

such that, for every $p \in M$ and $t \geq 1$,

$$k_{f,\theta}(p, t) = e^{-\frac{1+\nu}{2}t} D_{\theta,\mu,n}^+ f(p) + e^{-\frac{1-\nu}{2}t} D_{\theta,\mu,n}^- f(p) + \mathcal{R}_{\theta,\mu,n} f(p, t), \quad (1.10)$$

where

$$|\mathcal{R}_{\theta,\mu,n} f(p, t)| \leq \frac{4\kappa_0}{\theta\nu(1-\nu)(1+\nu)}(n^2 + 1) \|f\|_{\mathcal{C}^1} e^{-t}. \quad (1.11)$$

- (4) *If $\mu = 0$, there exists a function $G_{\theta,n} f: M \times \mathbb{R}_{>0} \rightarrow \mathbb{C}$, with $G_{\theta,n} f(\cdot, t)$ of class \mathcal{C}^1 for any $t > 0$, $G_{\theta,n} f(p, \cdot)$ continuous for every $p \in M$ and*

$$\sup_{p \in M, t > 0} |G_{\theta,n} f(p, t)| \leq \frac{\kappa_0}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1}$$

such that, for every $p \in M$ and $t \geq 1$,

$$k_{f,\theta}(p, t) = \int_M f \, d\text{vol} + e^{-t} \int_1^t -G_{\theta,n} f(p, \xi) \, d\xi + \mathcal{R}_{\theta,0,n} f(p, t), \quad (1.12)$$

where

$$|\mathcal{R}_{\theta,0,n}f(p,t)| \leq \frac{8e\pi + \kappa_0}{\theta}(n^2 + 1) \|f\|_{\mathcal{C}^1} e^{-t}. \quad (1.13)$$

(5) If $\mu < 0$ then, for every $p \in M$ and $t \geq 1$,

$$|k_{f,\theta}(p,t)| \leq \frac{1}{\theta} \left(\frac{4\kappa_0}{(\nu-1)(\nu+1)} + \frac{2e\pi(3+\nu)}{\nu} \right) (n^2 + 1) \|f\|_{\mathcal{C}^1} e^{-t}. \quad (1.14)$$

Remark 1.7. The asymptotic expansions for $k_{f,\theta}(p,t)$ are easily transferred to the case of large negative values of the time parameter t , by noticing that

$$k_{f,\theta}(p,-t) = k_{f \circ r_\pi, \theta}(r_\pi(p), t)$$

for any $p \in M$ and $t \geq 1$.

Taking advantage of Theorem 1.6 and of standard harmonic analysis on the group $\mathrm{SL}_2(\mathbb{R})$, we establish an asymptotic expansion of $k_{f,\theta}(p,t)$ for all sufficiently regular, but otherwise arbitrary test functions f . Define the Laplacian on M to be the second-order linear differential operator $\Delta = \square - 2\Theta^2$. For any $s \in \mathbb{R}_{>0}$, let $W^s(M)$ be the Sobolev space of order s on the manifold M , that is, the Hilbert-space completion of the complex vector space $\mathcal{C}^\infty(M)$ of smooth functions on M endowed with the inner product

$$\langle \phi, \psi \rangle_{W^s} = \langle (1 + \Delta)^s \phi, \psi \rangle, \quad \phi, \psi \in \mathcal{C}^\infty(M).$$

As M is compact, the well-known Sobolev Embedding Theorem (which we recall in Theorem 2.4) ensures the existence of a continuous embedding $W^s(M) \hookrightarrow \mathcal{C}^r(M)$ whenever $s \in \mathbb{R}_{>0}$ and $r \in \mathbb{N}$ are such that $s > r + 3/2$; explicitly, there is a constant $C_{r,s} \in \mathbb{R}_{>0}$ (which for definiteness we take equal to the operator norm of the corresponding embedding) such that $\|f\|_{\mathcal{C}^r} \leq C_{r,s} \|f\|_{W^s}$ for any $f \in W^s(M)$. Hereinafter, an element $f \in W^s(M)$ for $s > 3/2$ is always identified with its unique continuous representative.

Set

$$\varepsilon_0 = \begin{cases} 1 & \text{if } 1/4 \in \mathrm{Spec}(\square), \\ 0 & \text{if } 1/4 \notin \mathrm{Spec}(\square). \end{cases} \quad (1.15)$$

Theorem 1.8. *There exist real constants C_{Spec} and C'_{Spec} , depending only on the spectrum of the Casimir operator on $L^2(M)$, such that the following holds. Let $\theta \in (0, 4\pi]$ and $s > 11/2$ be real numbers, and suppose $f \in W^s(M)$. Then there exist continuous functions $D_{\theta,\mu}^+ f, D_{\theta,\mu}^- f: M \rightarrow \mathbb{C}$ for any positive Casimir eigenvalue μ , with*

$$\sum_{\mu \in \mathrm{Spec}(\square) \cap \mathbb{R}_{>0}} \|D_{\theta,\mu}^+ f\|_\infty + \|D_{\theta,\mu}^- f\|_\infty \leq \frac{C'_{\mathrm{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s}, \quad (1.16)$$

such that, for every $p \in M$ and $t \geq 1$,

$$\begin{aligned} \frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds &= \int_M f \, d \mathrm{vol} \\ &+ e^{-\frac{t}{2}} \left(\sum_{\mu \in \mathrm{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} t\right) D_{\theta,\mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2} t\right) D_{\theta,\mu}^- f(p) \right) \\ &+ \sum_{\mu \in \mathrm{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2} t} D_{\theta,\mu}^+ f(p) + e^{-\frac{1-\nu}{2} t} D_{\theta,\mu}^- f(p) \\ &+ \varepsilon_0 \left(e^{-\frac{t}{2}} D_{\theta,1/4}^+ f(p) + t e^{-\frac{t}{2}} D_{\theta,1/4}^- f(p) \right) + \mathcal{R}_\theta f(p,t), \end{aligned} \quad (1.17)$$

where

$$|\mathcal{R}_\theta f(p,t)| \leq \frac{C_{\mathrm{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s} (t+1) e^{-t}. \quad (1.18)$$

We record here below the ensuing effective equidistribution statement, in which we single out the two main terms of the asymptotic expansion, thereby highlighting the dependence of the latter on the spectral gap of the underlying hyperbolic surface $S = \Gamma \backslash \mathbb{H}$, defined as

$$\mu_* = \inf(\text{Spec}(\square) \cap \mathbb{R}_{>0}) = \inf(\text{Spec}(\Delta_S) \setminus \{0\}) ; \quad (1.19)$$

its corresponding parameter is denoted by ν_* .

Notation. We adopt the classical Landau notation $o(\eta(t))$ for $\eta: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ tending to zero at infinity, to indicate a function $\lambda: \mathbb{R}_{>0} \rightarrow \mathbb{C}$ such that $|\lambda(t)|/\eta(t) \rightarrow 0$ as $t \rightarrow \infty$.

Corollary 1.9. *Let $\theta, s, C_{1,s-3}, C'_{\text{Spec}}$ and f be as in Theorem 1.8. Then there exists a function $D_\theta^{\text{main}} f: M \rightarrow \mathbb{C}$ with*

$$\|D_\theta^{\text{main}} f\|_\infty \leq \frac{C'_{\text{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s} \quad (1.20)$$

such that, for every $p \in M$ and $t \geq 1$,

$$\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds = \int_M f \, d \text{vol} + D_\theta^{\text{main}} f(p) t^{\varepsilon_0} e^{-\frac{1-\Re \nu_*}{2} t} + o(e^{-\frac{1-\Re \nu_*}{2} t}). \quad (1.21)$$

Remark 1.10. Using the result of Dyatlov, Faure and Guillarmou in [19] we can relate the coefficients $D_{\theta,\mu}^- f$, for $0 < \mu < 1/4$, to the resonant and co-resonant states in the first band for the geodesic flow (in other words, with the invariant distributions for the unstable and stable horocycle flow, respectively). Let u_μ be a resonant state for the geodesic flow associated to the Pollicott-Ruelle resonance $-\frac{1-\nu}{2} \in (-1/2, 0)$, and let u_μ^* be its dual (we refer to [19, Sec. 1 and 2] for the relevant definitions). In the language of Flaminio-Forni's work [27], u_μ is an invariant distribution for the unstable horocycle flow, and u_μ^* is an invariant distribution for the stable horocycle flow. By virtue of [19, Thm. 4], we have, for any sufficiently small $\varepsilon > 0$ and for arbitrary $f, g \in \mathcal{C}^\infty(M)$,

$$\langle f \circ \phi_t^X, g \rangle = \int_M f \, d \text{vol} \int_M \bar{g} \, d \text{vol} + \sum_{\mu \in \text{Spec}(\square) \cap (0, 1/4)} \langle f, u_\mu \rangle \langle u_\mu^*, g \rangle e^{-\frac{1-\nu}{2} t} + O\left(e^{-\left(\frac{1}{2}-\varepsilon\right)t}\right).$$

Thus,

$$\begin{aligned} \langle k_{f,\theta}(\cdot, t), g \rangle &= \frac{1}{\theta} \int_0^\theta \langle f \circ \phi_t^X, g \circ r_{-s} \rangle \, ds \\ &= \int_M f \, d \text{vol} \int_M \bar{g} \, d \text{vol} + \sum_{\mu \in \text{Spec}(\square) \cap (0, 1/4)} \left(\frac{1}{\theta} \int_0^\theta \langle f, u_\mu \rangle \langle u_\mu^*, g \circ r_{-s} \rangle \, ds \right) e^{-\frac{1-\nu}{2} t} + O\left(e^{-\left(\frac{1}{2}-\varepsilon\right)t}\right). \end{aligned}$$

On the other hand, by Theorem 1.8,

$$\langle k_{f,\theta}(\cdot, t), g \rangle = \int_M f \, d \text{vol} \int_M \bar{g} \, d \text{vol} + \sum_{\mu \in \text{Spec}(\square) \cap (0, 1/4)} \langle D_{\theta,\mu}^- f, g \rangle e^{-\frac{1-\nu}{2} t} + O\left(e^{-\left(\frac{1}{2}-\varepsilon\right)t}\right).$$

Equating coefficients, we conclude that the functions $D_{\theta,\mu}^- f$, $0 < \mu < 1/4$, coincide as distributions with the corresponding averaged resonant states, namely

$$D_{\theta,\mu}^- f = \langle f, u_\mu \rangle \cdot \frac{1}{\theta} \int_0^\theta u_\mu^* \circ r_s \, ds. \quad (1.22)$$

Of particular interest is the full average $\theta = 4\pi$: in this case, $D_{4\pi,\mu}^- f$ is a multiple of $\frac{1}{4\pi} \int_0^{4\pi} u_\mu^* \circ r_s \, ds$, which is an eigenfunction of the Laplacian of eigenvalue μ , see [19, p. 931].

From (1.22), we also deduce that $D_{\theta,\mu}^- f$ is identically zero if and only if $\langle f, u_\mu \rangle = 0$, or, in other words, if and only if f lies in the kernel of the invariant distribution for the unstable horocycle flow with Casimir parameter μ (see [27, Sec. 3]).

In order to prove an analogous relation between the other coefficients $D_{\theta,\mu}^\pm f$ and the horocycle invariant distributions, we would need an asymptotic expansion of the correlations for the

geodesic flow as in [30, Thm. 4.3], but with an explicit dependence of the coefficients in terms of the stable and unstable horocycle-invariant distributions.

Remark 1.11. We collect here below further comments about Theorem 1.6, Theorem 1.8 and Corollary 1.9.

- (1) The Hölder-continuity claims concerning the coefficients $D_{\theta,\mu,n}^{\pm}$ appearing in Theorem 1.6 tacitly involve the choice of a distance function d on M . It is intended that d is the Riemannian distance function induced on the connected manifold M by a Riemannian metric g . The Hölder-continuity property, as well as the Hölder exponent, of $D_{\theta,\mu,n}^{\pm}$ is independent of the choice of such a g , as any two Riemannian metrics on a compact manifold induce Lipschitz-equivalent metrics (see [42, Lem. 13.28]).
- (2) We point out the analogy of Theorem 1.6 with the asymptotics of horocycle ergodic integrals for Casimir eigenfunctions established in [59, Thm. 1] (in the same way, Theorem 1.8 mirrors [59, Thm. 2]). This similarity stems computationally from the application of the exact same spectral method in both circumstances; as a matter of fact, it comes as no surprise from a geometric standpoint, for large hyperbolic circles approximate orbits of the unstable horocycle flow.
- (3) For $\theta = 4\pi$, we recover the qualitative equidistribution of expanding circles towards the uniform measure vol obtained by Margulis in [48].
- (4) The equidistribution rate in Corollary 1.9 matches exactly the mixing rate of the geodesic flow on M obtained by Ratner in [57, Thm. 2].
- (5) As suggested by the underlying geometric picture, the various upper bounds in the statements of Theorem 1.6 and 1.8 indicate that the speed of equidistribution improves as θ increases to 4π , that is, as the length of the initial circle arc gets larger.

As a special case of Theorem 1.8, we obtain an asymptotic expansion for the equidistribution rate of Θ -invariant functions; in other words, we provide the precise asymptotic behaviour of large circles on the compact hyperbolic surface S , identified, here and afterwards whenever convenient, with the double coset space $\Gamma \backslash \text{SL}_2(\mathbb{R}) / \text{SO}_2(\mathbb{R})$.

Let $d_{\mathbb{H}}$ denote the distance function on \mathbb{H} arising from the hyperbolic Riemannian metric. For any $z \in \mathbb{H}$ and $t > 0$, let $C_{\mathbb{H}}(z, t) = \{z' \in \mathbb{H} : d_{\mathbb{H}}(z, z') = t\}$ be the hyperbolic circle of radius t centered at z , and denote by $C_S(z, t)$ its projection to S . With $m_{C_S(z,t)}$ we indicate the projection to S of the unique isometrically-invariant Borel probability measure supported on $C_{\mathbb{H}}(z, t)$. Finally, let m_S be the hyperbolic area measure on S , normalized to be a probability measure.

Theorem 1.12. *Let $\theta \in (0, 4\pi]$, $s > 9/2$, $\tilde{f}: S \rightarrow \mathbb{C}$ a function such that the $\text{SO}_2(\mathbb{R})$ -invariant function $f: M \rightarrow \mathbb{C}$ defined by $f(\Gamma g) = f(\Gamma g \text{SO}_2(\mathbb{R}))$ for any $g \in \text{SL}_2(\mathbb{R})$ is in $W^s(M)$. Then, for every $z = \Gamma g \text{SO}_2(\mathbb{R}) \in S$, $p = \Gamma g \in M$ and $t \geq 1$,*

$$\begin{aligned} \int_S \tilde{f} \, dm_{C_S(z,t)} &= \int_S \tilde{f} \, dm_S + e^{-\frac{t}{2}} \left(\sum_{\mu \in \text{Spec}(\Delta_S), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} t\right) D_{4\pi, \mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2} t\right) D_{4\pi, \mu}^- f(p) \right) \\ &+ \sum_{\mu \in \text{Spec}(\Delta_S), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2} t} D_{4\pi, \mu}^+ f(p) + e^{-\frac{1-\nu}{2} t} D_{4\pi, \mu}^- f(p) \\ &+ \varepsilon_0 \left(e^{-\frac{t}{2}} D_{4\pi, 1/4}^+ f(p) + t e^{-\frac{t}{2}} D_{4\pi, 1/4}^- f(p) \right) + \mathcal{R}_{4\pi} f(p, t). \end{aligned}$$

In particular, the functions $\mathcal{R}_{4\pi} f$ and $D_{4\pi, \mu}^{\pm} f$, $\mu \in \text{Spec}(\Delta_S) \cap \mathbb{R}_{>0}$, defined as in Theorem 1.8, are $\text{SO}_2(\mathbb{R})$ -invariant.

Clearly, an analogous statement holds for arbitrary sub-arcs of the circles $C_S(z, t)$.

Remark 1.13. It is worth highlighting that, in the case of an $\text{SO}_2(\mathbb{R})$ -invariant observable f , we require mildly less restrictive assumptions on its Sobolev regularity; this will become relevant in Section 7 when we deal with the error rate for the hyperbolic lattice point counting problem.

Furthermore, we emphasize that Theorem 1.12 improves upon the equidistribution results in [55] (which, on the other hand, apply to any dimension) in a twofold way: it demands less restrictive conditions on the regularity of test functions and, more importantly, it refines Randol's upper bound on the equidistribution rate by giving a precise asymptotic expansion.

Leveraging the explicit dependence of the asymptotics in Theorem 1.8 on the upper bound θ of the domain of parametrization of the circle arc under consideration, we deduce a sufficient quantitative condition for the equidistribution of shrinking pieces of expanding circles; this is in the vein of Strömbergsson's results in [74], where the analogous question is investigated for shrinking portions of closed horocycles on non-closed hyperbolic surfaces of finite volume.

Corollary 1.14. *Let $p \in M$, $\theta_1, \theta_2: \mathbb{R}_{>0} \rightarrow (0, 4\pi)$ two functions with $\theta_1(t) \leq \theta_2(t)$ for any $t > 0$, and consider the circle sub-arcs $\gamma_t = \{\phi_t^X \circ r_s(p) : \theta_1(t) \leq s \leq \theta_2(t)\}$. For any $t > 0$, let μ_t be the normalized restriction to γ_t of the unique isometrically-invariant measure on the corresponding full circle. Assume that there exist $t_0 > 0$ and a function $\eta: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ with $\eta(t) \rightarrow \infty$ as $t \rightarrow \infty$ such that $\theta_2(t) - \theta_1(t) \geq \eta(t)e^{-\frac{1-\mathfrak{R}\nu^*}{2}t}$ for any $t \geq t_0$. Then the circle arcs γ_t equidistribute as $t \rightarrow \infty$: more precisely, the measures μ_t converge in the weak* topology, as $t \rightarrow \infty$, towards the uniform measure vol on M .*

1.4. Statistical limit theorems for deviations from the average. The asymptotics in Theorem 1.8 affords the means to examine the long-term statistical behaviour of the averages of a given observable along expanding circle arcs. Historically, a momentous discovery in the twentieth century was the realization that the long-term evolution of deterministic systems frequently obeys the same statistical laws governing the asymptotic behaviour of random processes. Specifically, this feature is a typical characteristic of dynamical systems with hyperbolic behaviour, among which geodesic flows on negatively curved compact manifolds feature prominently; we refer the reader to the survey in the introduction to [17], as well as to the references therein, for an extensive discussion of the topic.

Because of the exponential mixing properties of the geodesic flow $(\phi_t^X)_{t \in \mathbb{R}}$ on M (cf. [57, Thm. 2]), at first it stands to reason to expect, for a real-valued function f on M with finite first moment with respect to the volume measure vol , the distribution of the deviations from the average

$$\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} ,$$

when the base point p is randomly chosen according to the law vol , (something which is henceforth indicated with $p \sim \text{vol}$), to mimic for large values of T the law of the empirical mean

$$\frac{1}{N} \sum_{n=1}^N X_n$$

of an increasing number N of independent real-valued random variables X_n . More precisely, since the hyperbolic length of a circle of radius T is proportional to e^T (see the explanation below (7.7)), a full analogue of the classical Central Limit Theorem in this case would affirm that the random variables

$$e^{\frac{T}{2}} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right) , \quad p \sim \text{vol}$$

converge in law, as T tends to infinity, to a normally distributed random variable. However, the geometric resemblance of large hyperbolic circles to orbits of the unstable horocycle flow, for which similar phenomena occur (cf. [11, Thm. 1.4, 1.5] and [59, Thm. 4]) accounts both for the emergence of other types of limiting distributions, and for possibly different renormalization factors, depending on the spectral properties of the observable under consideration.

In order to minimize the difference with the classical probabilistic setup of sums of independent random variables, we shall state all the results in this subsection for real-valued observables, though the extension to complex-valued ones is immediate.

Theorem 1.15. *Let $\theta \in (0, 4\pi]$, $s > 11/2$, $f \in W^s(M)$ a real-valued function. Assume that*

$$\mu_f = \inf\{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0} : D_{\theta, \mu}^- f \text{ does not vanish identically on } M\}$$

is finite, and let ν_f be the unique complex number in $\mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ satisfying $1 - \nu_f^2 = 4\mu_f$.

(1) *If $0 < \mu_f < 1/4$, the random variables*

$$e^{\frac{1-\nu_f}{2}T} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right), \quad p \sim \text{vol}$$

converge in distribution to $D_{\theta, \mu_f}^- f$ as $T \rightarrow \infty$.

(2) *If $\mu_f = 1/4$, the random variables*

$$T^{-1} e^{\frac{T}{2}} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right), \quad p \sim \text{vol}$$

converge in distribution to $D_{\theta, 1/4}^- f$ as $T \rightarrow \infty$.

(3) *If $\mu_f > 1/4$, the random variables*

$$e^{\frac{T}{2}} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right), \quad p \sim \text{vol}$$

converge in distribution, as $T \rightarrow \infty$, to the quasi-periodic motion

$$\varepsilon_0 D_{\theta, 1/4}^+ f(p) + \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^+ f + \sum_{\mu \in \text{Spec}(\square), \mu \geq \mu_f} \sin\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^- f$$

on the set of real-valued random variables defined on the probability space (M, vol) .

Observe the remarkable fact that the limiting distributions appearing in the statement of Theorem 1.15 are compactly supported on the real line, owing to the fact that the coefficients $D_{\theta, \mu}^\pm f$ are bounded. This stands in stark contrast with the classical versions of the Central Limit Theorem in probability theory, where in non-trivial situations the distribution of errors is governed by the fully supported Gaussian distribution. Furthermore, it is straightforward to check, at least when the Casimir components of f are eigenfunctions of Θ and using the explicit expressions of the coefficients $D_{\theta, \mu}^\pm f$ in (4.8), (4.9), (4.12) and (4.14), that the limit law is non-trivial⁸, that is, not a Dirac mass. In the general case, the coefficients are given by infinite superpositions of the previous ones; though we shall refrain from a detailed verification, there is no reason to expect cancellation phenomena to come about and produce limiting random variables which are constant almost-surely.

Theorem 1.15 is a consequence of its quantitative version which we presently discuss. For f fulfilling the conditions in Theorem 1.15, let $\mathbf{P}_{\theta, f}$ denote the law of the random variable $D_{\theta, \mu_f}^- f$ when $\mu_f \leq 1/4$, and $\mathbf{P}_{\theta, f}(T)$ the law of the random variable

$$\varepsilon_0 D_{\theta, 1/4}^+ f(p) + \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^+ f + \sum_{\mu \in \text{Spec}(\square), \mu \geq \mu_f} \sin\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^- f, \quad T > 0,$$

when $\mu_f > 1/4$. Furthermore, for any $T \geq 1$ we let $\mathbf{P}_{\theta, f}^{\text{circ}}(T)$ be:

(1) the law of

$$e^{\frac{1-\nu_f}{2}T} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right) \quad \text{if } 0 < \mu_f < 1/4,$$

(2) the law of

$$T^{-1} e^{\frac{T}{2}} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right) \quad \text{if } \mu_f = 1/4,$$

⁸As soon as f is not almost-surely constant, obviously.

(3) and the law of

$$e^{\frac{T}{2}} \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right) \quad \text{if } \mu_f > 1/4.$$

Denote by d_{LP} the Lévi-Prokhorov distance on the set of Borel probability measures on \mathbb{R} (cf. Section 6.1), which induces on the latter the topology of weak convergence. Recall also that μ_* denotes the spectral gap of $S = \Gamma \backslash \mathbb{H}$, with associated parameter ν_* .

Proposition 1.16. *Let the assumptions be as in Theorem 1.15, and the constants C_{Spec} and C'_{Spec} be as in Theorem 1.8.*

(1) *If $0 < \mu_f < 1/4$, then there is an explicit constant $\eta_f > 0$, depending only on μ_f and on $\text{Spec}(\square)$, such that*

$$d_{LP}(\mathbf{P}_{\theta,f}^{\text{circ}}(T), \mathbf{P}_{\theta,f}) \leq \frac{C'_{\text{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s} T e^{-\eta_f T}$$

for every $T \geq 1$.

(2) *If $\mu_f = 1/4$, then there exists a constant C_{pos} , depending only on $\text{Spec}(\square) \cap \mathbb{R}_{>0}$, such that*

$$d_{LP}(\mathbf{P}_{\theta,f}^{\text{circ}}(T), \mathbf{P}_{\theta,f}(T)) \leq \frac{C_{\text{pos}} C_{1,s-3}}{\theta} \|f\|_{W^s} T^{-1}$$

for every $T \geq 1$.

(3) *If $\mu_f > 1/4$, then:*

(a) *when $\mu_* < 1/4$,*

$$d_{LP}(\mathbf{P}_{\theta,f}^{\text{circ}}(T), \mathbf{P}_{\theta,f}(T)) \leq \frac{C'_{\text{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s} e^{-\frac{\nu_*}{2} T}$$

for every $T \geq 1$;

(b) *when $\mu_* \geq 1/4$,*

$$d_{LP}(\mathbf{P}_{\theta,f}^{\text{circ}}(T), \mathbf{P}_{\theta,f}(T)) \leq \frac{C_{\text{Spec}} C_{1,s-3}}{\theta} \|f\|_{W^s} (T+1) e^{-\frac{T}{2}}$$

for every $T \geq 1$.

Entirely analogous deductions, of which we omit the details (cf. Section 6.1), can be made in the case where the $D_{\theta,\mu}^- f$ vanish everywhere for any Casimir eigenvalue $\mu > 0$ but $D_{\theta,\mu}^+ f$ is not identically zero for at least one such μ .

Remark 1.17. As the proof of Proposition 1.16 shall clearly illustrate (see, in particular, Lemma 6.1), the assumption that the base point p is sampled according to the uniform measure vol is immaterial, as far as the validity of Proposition 1.16 and Theorem 1.15 is concerned. It is possible to replace the measure vol with any other Borel probability measure μ on M , without affecting the quantitative rate of convergence, provided that the laws of the limiting random variables are modified accordingly.

In the remaining case when the $D_{\theta,\mu}^\pm f$ vanish identically on M for any positive Casimir eigenvalue μ , the explicit expressions of the coefficients appearing in Theorem 1.6 and 1.8 enable us to rule out the existence of a non-trivial distributional limit in the case of full circles, that is, when $\theta = 4\pi$. More precisely, we establish the following result.

Theorem 1.18. *Let $s > 11/2$, $f \in W^s(M)$ a real-valued function. Assume that, for any positive Casimir eigenvalue μ , the functions $D_{4\pi,\mu}^\pm f$ defined as in Theorem 1.8 vanish identically on M . Then, for any collection $(B_T)_{T>0}$ of positive real numbers such that $B_T \rightarrow \infty$ as $T \rightarrow \infty$, the distributional limit of the random variables*

$$\frac{e^T \left(\frac{1}{4\pi} \int_0^{4\pi} f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} \right)}{B_T}, \quad p \sim \text{vol}$$

as $T \rightarrow \infty$ exists and is almost surely equal to zero.

Theorem 1.18 results from approximating averages of f along expanding circle arcs with the difference of the ergodic integrals of f along two geodesic orbits, which in the case of complete circles happen to coincide; details are carried out in Section 6.2.

Remark 1.19. We hasten to add that the argument we conduct enables to show that Theorem 1.18 holds true for arbitrary circle arcs (that is, it is possible to replace 4π with an arbitrary $\theta \in (0, 4\pi]$) provided that, in addition to the vanishing hypothesis on the $D_{4\pi, \mu}^\pm$ for $\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}$, the derivative Uf_0 along the stable horocycle flow of the projection f_0 of f onto the Casimir eigenspace of eigenvalue 0 is assumed to be a coboundary for the geodesic flow (cf. Section 6.2).

1.5. Asymptotics for arbitrary translates of compact orbits and the circle problem in the hyperbolic plane. The celebrated Gauss circle problem asks for the precise asymptotic behaviour of the discrepancy between the number of integer points in a disk of radius R in the Euclidean plane and the area of the disk, as R tends to infinity. More precisely, define the integer-point counting function

$$\mathcal{N}(R) = |\{(m, n) \in \mathbb{Z}^2 : m^2 + n^2 \leq R^2\}|, \quad R \in \mathbb{R}_{>0},$$

where $|A|$ denotes, here and henceforth, the cardinality of a finite set A . Tessellating the Euclidean plane with \mathbb{Z}^2 -translates of $[0, 1)^2$, which is a fundamental domain for the \mathbb{Z}^2 -action by translations on \mathbb{R}^2 , leads to the main term πR^2 , equal to the area of the disk of radius R , for the asymptotics of $\mathcal{N}(R)$, as well as to the upper bound (due to Gauss himself)

$$|\mathcal{N}(R) - \pi R^2| \leq 2\pi(\sqrt{2}R + 1)$$

for the discrepancy. Despite considerable successive improvements on Gauss' original bound, for the history of which we refer to the comprehensive survey [33], it is a notoriously unsolved problem to attain the conjectural sharpest upper bound, deemed to be of the order of $R^{1/2+\varepsilon}$ for any $\varepsilon > 0$.

We consider the analogous question in the hyperbolic plane. For $\Gamma < \text{SL}_2(\mathbb{R})$ a cocompact lattice, we examine the asymptotics of the function

$$N(R) = |\{z \in \Gamma \cdot i : d_{\mathbb{H}}(z, i) \leq R\}| \tag{1.23}$$

as R tends to infinity, where $\Gamma \cdot i$ denotes the (discrete) orbit⁹ of i under the Γ -action on the hyperbolic plane \mathbb{H} and we recall that $d_{\mathbb{H}}$ is the hyperbolic distance function on \mathbb{H} .

Remark 1.20. Upon replacing Γ by a conjugate, there is no loss of generality in choosing $i \in \mathbb{H}$ as the base point: an elementary algebraic computation, together with the fact that $\text{SL}_2(\mathbb{R})$ acts by $d_{\mathbb{H}}$ -isometries, leads to the equality

$$|\{z \in \Gamma \cdot w : d_{\mathbb{H}}(z, w) \leq R\}| = |\{z \in g^{-1}\Gamma g \cdot i : d_{\mathbb{H}}(z, i) \leq R\}|$$

for any $w = g \cdot i \in \mathbb{H}$ and $g \in \text{SL}_2(\mathbb{R})$.

There is a compact fundamental domain for the action of Γ on \mathbb{H} , and a transposition of Gauss' tessellation argument to this setup provides a rationale for the heuristics concerning the main term of the asymptotics, which once again should be proportional to the hyperbolic area measure of the ball $B_R = \{z \in \mathbb{H} : d_{\mathbb{H}}(z, i) \leq R\}$, which we denote by $m_{\mathbb{H}}(B_R)$. However, a consequence of the peculiar features of hyperbolic geometry is that boundary effects become relevant, as opposed to the Euclidean setting: more precisely, the growth rate of the length of the boundary ∂B_R turns out to be equal to the growth rate of $m_{\mathbb{H}}(B_R)$. The error rate resulting from the tessellation approach is consequently of the same order of the main term, and as such meaningless.

⁹We thus count the number of actual lattice-orbit elements; we remark that, in the literature, the count of lattice elements $\gamma \in \Gamma$ such that $d_{\mathbb{H}}(\gamma \cdot i, i) \leq R$ often appears instead; the two quantities are proportional by a factor $|\text{Stab}_{\Gamma}(i)|$.

As for its Euclidean counterpart, the circle problem in the hyperbolic plane has been the subject of intensive research over the course of the twentieth century, with fundamental contributions due to Delsarte ([15]), Selberg ([64]), Margulis ([47]), Lax and Phillips ([41]) and Phillips and Rudnick ([52], see its introduction for a detailed history of the problem). To a large extent, the state of the art concerning the best estimate on the error term $|N(R) - c_\Gamma m_{\mathbb{H}}(B_R)|$, where c_Γ is an explicit constant which we identify in Theorem 1.24, is represented by Selberg's upper bound

$$|N(R) - c_\Gamma m_{\mathbb{H}}(B_R)| \leq e^{(\sup\{2/3, (1+\Re\nu_*)/2\})R}, \quad (1.24)$$

where recall that $(1 - \nu_*^2)/4$ is the spectral gap of $S = \Gamma \backslash \mathbb{H}$. The estimate¹⁰ in (1.24) (which is equally valid for non-uniform lattices $\Gamma < \mathrm{SL}_2(\mathbb{R})$) is obtained by means of a deep analysis of a class of integral operators commuting with hyperbolic isometries (cf. [64]).

It was Margulis' realization (see [48]) that lattice point counting problems of the type we are examining are intimately interwoven with questions of equidistribution of translates of subgroup orbits on homogeneous spaces. Subsequently, this novel perspective was profitably pursued and vastly generalized in the works of Duke, Rudnick and Sarnak [18] and Eskin and McMullen [24]. Specializing to our current setup, it turns out that the distribution properties of translates of $\mathrm{SO}_2(\mathbb{R})$ -orbits¹¹ on $M = \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ lead to meaningful information on the growth rate of $N(R)$, in a way that is amenable to quantitative refinements (see Section 7 for an extensive treatment of the connection in its quantitative form).

We are thus lead to study effective equidistribution properties of $\mathrm{SO}_2(\mathbb{R})$ -orbits on M , which can be readily derived from Theorem 1.8 via the standard Cartan decomposition for $\mathrm{SL}_2(\mathbb{R})$. For any $g_0 \in \mathrm{SL}_2(\mathbb{R})$, define the right-translation map $R_{g_0}: M \rightarrow M$ by $R_{g_0}(\Gamma g) = \Gamma g g_0$ for any $g \in \mathrm{SL}_2(\mathbb{R})$. For every $p \in M$, we indicate with $m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ the unique $\mathrm{SO}_2(\mathbb{R})$ -invariant Borel probability measure on M which is fully supported on the (compact) $\mathrm{SO}_2(\mathbb{R})$ -orbit of the point p ; furthermore, for any $g \in \mathrm{SL}_2(\mathbb{R})$, the notation $g_* m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ stands for the push-forward of $m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ under the action of g , which clearly depends only on the left coset of g modulo $\mathrm{SO}_2(\mathbb{R})$. Making use of the mixing properties of the global $\mathrm{SL}_2(\mathbb{R})$ -action on M via a clever thickening argument, Margulis proved (see [47, 48]) that arbitrary translates of $m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ equidistribute towards the $\mathrm{SL}_2(\mathbb{R})$ -invariant measure vol ; this amounts to the fact that, for any continuous function $f: M \rightarrow \mathbb{C}$,

$$\int_M f \, dg_* m_{\mathrm{SO}_2(\mathbb{R}) \cdot p} \longrightarrow \int_M f \, d \mathrm{vol}$$

as $g \mathrm{SO}_2(\mathbb{R})$ tends to infinity in the quotient $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$.

In order to phrase a quantitative version of the previous statement conveniently, we introduce the notation $\|g\|_{\mathrm{op}}$ to indicate operator norm of an element $g \in \mathrm{SL}_2(\mathbb{R})$ with respect to the standard Euclidean norm on \mathbb{R}^2 ; such a specific choice, while obviously immaterial, arises naturally over the course of the proof.

Theorem 1.21. *Let $C_{\mathrm{Spec}}, C'_{\mathrm{Spec}}$ be as in Theorem 1.8, $s > 11/2$ a real number. There exists a real constant $C_{s, \mathrm{rot}}$, depending only on s and on M , such that the following holds: if $f \in W^s(M)$, then there exist, for any positive Casimir eigenvalue μ , functions $D_\mu^+ f, D_\mu^- f: M \times \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ with*

$$\sum_{\mu \in \mathrm{Spec}(\square) \cap \mathbb{R}_{>0}} \sup_{p \in M, g \in \mathrm{SL}_2(\mathbb{R})} |D_\mu^\pm f(p, g)| \leq C_{s, \mathrm{rot}} C'_{\mathrm{Spec}} C_{1, s-3} \|f\|_{W^s},$$

¹⁰As a matter of fact, Selberg's asymptotics is more accurate than the one recorded here, featuring a main term which involves, beside the hyperbolic area of the balls, additional terms depending on Laplace eigenfunctions for small eigenvalues; see [52, Eq. 1.13].

¹¹We consider here the canonical left action $g \cdot \Gamma g' = \Gamma g' g^{-1}$ of $\mathrm{SL}_2(\mathbb{R})$ on M .

such that, for every $p \in M$ and every $g \in \mathrm{SL}_2(\mathbb{R})$ with $\|g\|_{\mathrm{op}} \geq \sqrt{e}$,

$$\begin{aligned} \int_M f \, dg_* m_{\mathrm{SO}_2(\mathbb{R}) \cdot p} &= \int_M f \, d \mathrm{vol} \\ &+ \|g\|_{\mathrm{op}}^{-1} \left(\sum_{\mu \in \mathrm{Spec}(\square), \mu > 1/4} \cos(\Im \nu \log \|g\|_{\mathrm{op}}) D_\mu^+ f(p, g) + \sin(\Im \nu \log \|g\|_{\mathrm{op}}) D_\mu^- f(p, g) \right) \\ &+ \sum_{\mu \in \mathrm{Spec}(\square), 0 < \mu < 1/4} \|g\|_{\mathrm{op}}^{-(1+\nu)} D_\mu^+ f(p, g) + \|g\|_{\mathrm{op}}^{-(1-\nu)} D_\mu^- f(p, g) \\ &+ \varepsilon_0 (\|g\|_{\mathrm{op}}^{-1} D_{1/4}^+ f(p, g) + 2 \|g\|_{\mathrm{op}}^{-1} \log \|g\|_{\mathrm{op}} D_{1/4}^- f(p, g)) + \mathcal{R} f(p, g), \end{aligned}$$

where

$$|\mathcal{R} f(p, g)| \leq C_{\mathrm{Spec}} C_{1, s-3} C_{s, \mathrm{rot}} \|f\|_{W^s} (2 \log \|g\|_{\mathrm{op}} + 1) \|g\|_{\mathrm{op}}^{-2}.$$

Remark 1.22. (1) The problem of effective equidistribution of translates of finite-volume orbits, in the vastly more general context of affine symmetric spaces, was thoroughly explored by Benoist and Oh in [5]. Their approach relies crucially on effective bounds for the mixing rates¹² of the relevant global action, and systematically develops quantitative versions of the geometric properties which play a major role in the original work [24] of Eskin and McMullen. In the specific instance of the affine symmetric space being the hyperbolic plane \mathbb{H} , Theorem 1.21 improves upon [5, Thm. 1.10] in that it quantifies the exponent governing the equidistribution rate and spells out additional terms in the asymptotic expansion.

(2) Just as in the case of Theorem 1.8 and Corollary 1.9, the asymptotic expansion in Theorem 1.21 delivers at once the (optimal) effective equidistribution bound

$$\left| \int_M f \, dg_* m_{\mathrm{SO}_2(\mathbb{R}) \cdot p} - \int_M f \, d \mathrm{vol} \right| \leq D^{\mathrm{main}} f(p, g) (\log \|g\|_{\mathrm{op}})^{\varepsilon_0} \|g\|_{\mathrm{op}}^{-(1-\Re \nu_*)}$$

for every $p \in M$ and $g \in \mathrm{SL}_2(\mathbb{R})$ with $\|g\|_{\mathrm{op}} \geq \sqrt{e}$, where the function $D^{\mathrm{main}}: M \times \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ is uniformly bounded in terms of an appropriate Sobolev norm of f and of spectral data depending only on M . Following the thread of the observations expressed in Remark 1.11, it is instructive to compare it with the decay rates for matrix coefficients of unitary representations of $\mathrm{SL}_2(\mathbb{R})$ computed by Venkatesh in [79, Sec. 9.1.2].

Theorem 1.21 affords a precise asymptotic formula for the averaged counting of points on translates of Γ -orbits inside balls of increasing radius, that is, for quantities of the form¹³

$$\int_{\mathrm{SL}_2(\mathbb{R})/\Gamma} \frac{|g\Gamma \cdot i \cap B_R|}{m_{\mathbb{H}}(B_R)} \psi(g\Gamma) \, d \mathrm{vol}(g\Gamma) \quad (1.25)$$

as $\|g\|_{\mathrm{op}}$ tends to infinity, where ψ is a sufficiently regular function on $\mathrm{SL}_2(\mathbb{R})/\Gamma$. Prior to the statement of the result, we shall fix advantageous normalizations for the various invariant measures involved (see Section 7.1 for the details).

Let $m_{\mathrm{SL}_2(\mathbb{R})}$ be the unique choice of Haar measure on $\mathrm{SL}_2(\mathbb{R})$ such that, if $m_{\mathrm{SO}_2(\mathbb{R})}$ is the probability Haar measure on the compact subgroup $\mathrm{SO}_2(\mathbb{R})$, then $m_{\mathrm{SL}_2(\mathbb{R})}$ is the (formal) product (cf. Proposition 7.2) of $m_{\mathrm{SO}_2(\mathbb{R})}$ and of the $\mathrm{SL}_2(\mathbb{R})$ -invariant measure on the homogeneous space $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$ which corresponds, under the canonical identification of the latter space with

¹²Specializing to our setup, effective mixing for the $\mathrm{SL}_2(\mathbb{R})$ -action on finite-volume quotients $\Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ was first worked out in detail, to the best of our knowledge, by Kleinbock and Margulis in [37, Sec. 2.4] (see also [57]).

¹³As it is customary in the literature addressing such themes, we shall choose to work with spaces of left cosets whenever dealing with the lattice point counting problem. The map $g\Gamma \mapsto \Gamma g^{-1}$ establishes an $\mathrm{SL}_2(\mathbb{R})$ -equivariant diffeomorphism, between $\mathrm{SL}_2(\mathbb{R})/\Gamma$ and $\Gamma \backslash \mathrm{SL}_2(\mathbb{R})$; every object we have defined on $\Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ shall thus be identified (without altering notation) with the corresponding object in $\mathrm{SL}_2(\mathbb{R})/\Gamma$ without further comment.

\mathbb{H} , to the hyperbolic area measure $m_{\mathbb{H}}$. We then indicate with $m_{\mathrm{SL}_2(\mathbb{R})/\Gamma}$ the unique $\mathrm{SL}_2(\mathbb{R})$ -invariant measure on $\mathrm{SL}_2(\mathbb{R})/\Gamma$ such that $m_{\mathrm{SL}_2(\mathbb{R})}$ is the product of $m_{\mathrm{SL}_2(\mathbb{R})/\Gamma}$ and the counting measure on the discrete group Γ . Observe that $m_{\mathrm{SL}_2(\mathbb{R})/\Gamma}$ is a scalar multiple of the probability measure vol , the multiplying factor being equal to the covolume

$$\mathrm{covol}_{\mathrm{SL}_2(\mathbb{R})}(\Gamma) = m_{\mathrm{SL}_2(\mathbb{R})/\Gamma}(\mathrm{SL}_2(\mathbb{R})/\Gamma)$$

of the lattice Γ inside $\mathrm{SL}_2(\mathbb{R})$. We denote similarly by $\mathrm{covol}_{\mathrm{SO}_2(\mathbb{R})}(\Gamma \cap \mathrm{SO}_2(\mathbb{R}))$ the volume¹⁴ of the compact quotient $\mathrm{SO}_2(\mathbb{R})/(\Gamma \cap \mathrm{SO}_2(\mathbb{R}))$ with respect to the $\mathrm{SO}_2(\mathbb{R})$ -invariant measure induced by $m_{\mathrm{SO}_2(\mathbb{R})}$ and the counting measure on $\Gamma \cap \mathrm{SO}_2(\mathbb{R})$.

Lastly, recall that B_R is the closed hyperbolic ball in \mathbb{H} of radius $R > 0$ centered at i .

Proposition 1.23. *Let $C_{\mathrm{Spec}}, C'_{\mathrm{Spec}}$ be as in Theorem 1.8. Suppose given a real number $s > 11/2$ and a function $\psi \in W^s(\mathrm{SL}_2(\mathbb{R})/\Gamma)$. There exist, for any positive Casimir eigenvalue μ , functions $\beta_{\psi, \mu}^+, \beta_{\psi, \mu}^- : \mathbb{R}_{>0} \rightarrow \mathbb{C}$ with*

$$\sum_{\mu \in \mathrm{Spec}(\square) \cap \mathbb{R}_{>0}} \sup_{R > 0} |\beta_{\psi, \mu}^{\pm}(R)| \leq \frac{C'_{\mathrm{Spec}} C_{1, s-3}}{2\pi} \|\psi\|_{W^s} \quad (1.26)$$

such that, for every $R \geq 1$,

$$\begin{aligned} \frac{1}{\mathrm{covol}_{\mathrm{SO}_2(\mathbb{R})}(\Gamma \cap \mathrm{SO}_2(\mathbb{R}))} \int_{\mathrm{SL}_2(\mathbb{R})/\Gamma} \frac{|g\Gamma \cdot i \cap B_R|}{m_{\mathbb{H}}(B_R)} \psi(g\Gamma) \, dm_{\mathrm{SL}_2(\mathbb{R})/\Gamma}(g\Gamma) = \\ \frac{1}{\mathrm{covol}_{\mathrm{SL}_2(\mathbb{R})}(\Gamma)} \int_{\mathrm{SL}_2(\mathbb{R})/\Gamma} \psi \, dm_{\mathrm{SL}_2(\mathbb{R})/\Gamma} \\ + e^{-\frac{R}{2}} \sum_{\mu \in \mathrm{Spec}(\square), \mu > 1/4} \beta_{\psi, \mu}^+(R) + \beta_{\psi, \mu}^-(R) \\ + \sum_{\mu \in \mathrm{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2}R} \beta_{\psi, \mu}^+(R) + e^{-\frac{1-\nu}{2}R} \beta_{\psi, \mu}^-(R) \\ + \varepsilon_0 \left(e^{-\frac{R}{2}} \beta_{\psi, 1/4}^+(R) + R e^{-\frac{R}{2}} \beta_{\psi, 1/4}^-(R) \right) + \gamma_{\psi}(R), \end{aligned} \quad (1.27)$$

where

$$|\gamma_{\psi}(R)| \leq \frac{5C_{\mathrm{Spec}} C_{1, s-3}}{4\pi} \|\psi\|_{W^s} (R+1) e^{-R} \quad (1.28)$$

Recall now our definition of the counting function $N(R)$ in (1.23). An optimization argument involving approximate identities on the homogeneous space $\mathrm{SL}_2(\mathbb{R})/\Gamma$ enables us to derive information on the asymptotic behaviour of the error term in the pointwise counting problem discussed at the beginning of this subsection.

Theorem 1.24. *Let $\Sigma : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ be the function defined by*

$$\Sigma(R) = \frac{\mathrm{covol}_{\mathrm{SO}_2(\mathbb{R})}(\Gamma \cap \mathrm{SO}_2(\mathbb{R}))}{\mathrm{covol}_{\mathrm{SL}_2(\mathbb{R})}(\Gamma)} m_{\mathbb{H}}(B_R), \quad R > 0.$$

Set $\eta_* = \frac{1}{13}(1 - \Re\nu_*)$. Then, for every $\varepsilon > 0$,

$$N(R) = \Sigma(R) + o(e^{(1-\eta_*+\varepsilon)R})$$

as R tends to infinity.

Remark 1.25. As it will emerge in the proof of Theorem 1.24, which is detailed in Section 7.2, the appearance of the quantity $1/13$ in the exponent is ultimately an outgrowth of the minimal Sobolev regularity of the test function f we need to impose in Theorem 1.12, which in turn is needed because of the upper bounds in (1.16) and (1.18) depending on the Sobolev norm

¹⁴A straightforward application of the formula in (7.2) shows that this equals the reciprocal of the cardinality of the finite group $\Gamma \cap \mathrm{SO}_2(\mathbb{R})$.

$\|f\|_{W^s}$ for some $s > 9/2$. In this regard, we observe the following: suppose that, in the latter two bounds, the norm $\|f\|_{W^s}$ can be replaced by $\|f\|_{W^1}$, as it is the case for joint eigenfunctions of \square and Θ (cf. Theorem 1.6); then Theorem 1.24 would hold with $1/6$ in place of $1/13$.

1.6. Outline of the proofs and layout of the article. The method we employ to prove Theorem 1.6 was originally devised by Ratner in [57], who realized that the problem of finding mixing rates for geodesic and horocycle flows can be reduced to solving a family of linear second-order ordinary differential equations¹⁵. This ingenious and yet fairly elementary approach has been further developed by Burger in [12] to prove polynomial bounds for the equidistribution of horocycle orbits in compact quotients of $\mathrm{SL}_2(\mathbb{R})$. Later, Strömbergsson [75] and Edwards [21] exploited the same idea to study effective equidistribution properties of unipotent orbits in more general settings. In the same spirit, the second author recently provided (see [59]), using Ratner's strategy, an alternative proof of Flaminio-Forni's asymptotic expansion for horocycle ergodic integrals [27].

We begin in Section 2 with an overview of the required notions concerning hyperbolic surfaces, Sobolev spaces and harmonic analysis on the Lie group $\mathrm{SL}_2(\mathbb{R})$, nailing down notation to be employed throughout the manuscript. Assuming that a \mathcal{C}^2 -observable f is a joint eigenfunction of the Casimir operator and of the vector field Θ , we then show in Section 3 that the behaviour of the circle-arc averages $k_{f,\theta}(p, \cdot)$ (see (1.6)), viewed as functions of the time t for a fixed base point $p \in \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$, fulfill a second order linear ODE, solving which explicitly leads to the proof of Theorem 1.6 presented in Section 4; incidentally, we may arrange computations so that the latter takes on the same form of the ODE satisfied by time rescalings of horocycle averages in [59] (see, in particular, [59, Prop. 8]), which accounts for the similarity between Theorem 1.6 with [59, Thm. 1]. In Section 5, the asymptotic expansion of Theorem 1.8 is deduced from Theorem 1.6 taking advantage of a few elementary facts from the classical harmonic analysis of $\mathrm{SL}_2(\mathbb{R})$. Additional regularity on f is required in order to ensure the absolute convergence of the expansion in (1.17); see, in particular, Section 5.1. The asymptotics for arbitrary translates in Theorem 1.21 is derived from Theorem 1.8 in Section 5.5. Building on Theorem 1.8 once more, we establish in Sections 6.1 and 6.2 the distributional limit Theorems 1.15 and 1.18 for the random variable $k_{f,\theta}(p, t)$, appropriately centered and normalized, when the initial point p is taken randomly with respect to the uniform measure. These limit theorems mirror those for horocycle ergodic integrals, for which we refer the reader to [11, 59]. Section 6.3 hosts a few considerations concerning the point of view of temporal distributional limit theorems (see [17]) on the problem of analyzing the statistical behaviour of circle averages. Finally, in Section 7, we provide a quantitative treatment of the approach of Duke-Rudnick-Sarnak [18] and Eskin-McMullen [24], which allows to prove both Proposition 1.23 and Theorem 1.24 on the hyperbolic lattice point counting problem.

2. PRELIMINARIES ON HARMONIC ANALYSIS ON $\mathrm{SL}_2(\mathbb{R})$

It is the aim of this section to carefully describe the setting of our main results as well as to review the background on the representation theory of the group $\mathrm{SL}_2(\mathbb{R})$ which will play a central role throughout the article.

2.1. Hyperbolic surfaces and their unit tangent bundles. The subject of this subsection is classical: detailed treatments can be found, for instance, in [4, 6, 8, 13, 23, 34, 36].

The special linear group $\mathrm{SL}_2(\mathbb{R})$ is the group of 2×2 real matrices with determinant 1. It is a three-dimensional Lie group, whose Lie algebra we denote by $\mathfrak{sl}_2(\mathbb{R})$ and identify canonically

¹⁵After the completion of a first draft of the present article, the authors were made aware of the unpublished manuscript [20] by S. Edwards, in which a weaker formulation of the quantitative equidistribution result in Theorem 1.8 is provided. The strategy of proof is entirely analogous to the one pursued here, and is there applied, more generally, to the quantitative investigation of the equidistribution properties of translated orbits of symmetric subgroups on homogeneous spaces of semisimple Lie groups.

with the Lie algebra of traceless 2×2 matrices with real entries. The identity matrix in $\mathrm{SL}_2(\mathbb{R})$ is denoted by I_2 . A basis of $\mathfrak{sl}_2(\mathbb{R})$ as a real vector space is given by the elements

$$X = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

With $\exp: \mathfrak{sl}_2(\mathbb{R}) \rightarrow \mathrm{SL}_2(\mathbb{R})$ we indicate the exponential map, and with $\mathrm{Ad}: \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathrm{GL}(\mathfrak{sl}_2(\mathbb{R}))$, $g \mapsto \mathrm{Ad}_g$ the adjoint representation of $\mathrm{SL}_2(\mathbb{R})$ onto $\mathfrak{sl}_2(\mathbb{R})$.

Let $\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$ be the Poincaré upper-half plane, endowed with the Riemannian metric

$$g_{(x,y)} = \frac{(dx)^2 + (dy)^2}{y^2}, \quad (x, y) \in \mathbb{H}.$$

The Riemannian manifold (\mathbb{H}, g) is a model of the hyperbolic plane, that is, of the unique complete simply connected two-dimensional Riemannian manifold of constant sectional curvature equal to -1 (cf. [10, Part 1 Chap. 6]). The Lie group $\mathrm{SL}_2(\mathbb{R})$ acts smoothly by orientation-preserving¹⁶ isometries of the hyperbolic plane. The action is given by the Möbius transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}, \quad a, b, c, d \in \mathbb{R}, \quad ad - bc = 1, \quad z \in \mathbb{H};$$

it is transitive, and thus gives rise to an $\mathrm{SL}_2(\mathbb{R})$ -equivariant diffeomorphism between \mathbb{H} and the quotient manifold $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$, where the special orthogonal group $\mathrm{SO}_2(\mathbb{R})$ is the stabilizer of the point $i \in \mathbb{H}$.

Let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be a cocompact lattice¹⁷. If the projection of Γ to the projective special linear group $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I_2\}$ has no non-trivial torsion elements, then the quotient space $S = \Gamma \backslash \mathbb{H}$, homeomorphic to the double coset space $\Gamma \backslash \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$, is a compact connected orientable smooth surface inheriting from \mathbb{H} a Riemannian metric g_S of constant sectional curvature equal to -1 , that is, a hyperbolic Riemannian metric. Conversely, it is a well-known consequence of Poincaré-Koebe's Uniformization Theorem (cf. [26, Chap. IV]) that any compact connected orientable hyperbolic surface is isomorphic, as a Riemannian manifold, to a quotient $\Gamma \backslash \mathbb{H}$ where $\Gamma < \mathrm{SL}_2(\mathbb{R})$ is a cocompact lattice with torsion-free projection on $\mathrm{PSL}_2(\mathbb{R})$. In case the projection of Γ to $\mathrm{PSL}_2(\mathbb{R})$ has non-trivial torsion elements, the quotient $S = \Gamma \backslash \mathbb{H}$ is, more generally, a hyperbolic *orbifold* (see [54, Chap. 13] and [77, Chap. 13]). Unifying, though mildly abusing terminology, we shall throughout refer to S as a surface in both the previous cases.

The simply transitive action of $\mathrm{PSL}_2(\mathbb{R})$ on the unit tangent bundle of \mathbb{H} allows to identify, as smooth manifolds, the unit tangent bundle $T^1S = \{(p, v) \in TS : \|v\|_{g_S} = 1\}$ of S with the homogeneous space $M = \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$. Throughout this manuscript, we shall solely appeal to the algebraic structure of M as a homogeneous space of $\mathrm{SL}_2(\mathbb{R})$, and not to its geometric nature of principal circle bundle over the surface S .

For any $r \in \mathbb{N} \cup \{\infty\}$, we denote by $\mathcal{C}^r(M)$ the vector space of complex-valued functions of class \mathcal{C}^r defined on M . It is endowed with the norm $\|\cdot\|_{\mathcal{C}^r}$ defined in (1.5). The \mathcal{C}^0 -norm is abbreviated with $\|\cdot\|_{\infty}$. If $f: M \rightarrow \mathbb{C}$ is a differentiable function and p is a point in M , we indicate with df_p the differential of f at p ; furthermore, if $(\partial_{x_i})_{i=1,2,3}$ is a smooth local frame of the tangent bundle TM around p , we indicate with $\partial_{x_i}|_p f$ the derivative of f at p along the tangent vector $\partial_{x_i}|_p$ for any $i = 1, 2, 3$.

A vector $W \in \mathfrak{sl}_2(\mathbb{R})$ gives rise to the one-parameter subgroup $(\exp(tW))_{t \in \mathbb{R}}$ of $\mathrm{SL}_2(\mathbb{R})$, which in turn acts as a smooth flow on M by right translations. Whenever convenient, we identify W with the infinitesimal generator ([42, Chap. 9]) of such a flow, and denote its action as a derivation on the function space $\mathcal{C}^1(M)$ by $f \mapsto Wf$. This assignment extends to an

¹⁶What is more, the action is by analytic transformations of the Riemann surface \mathbb{H} .

¹⁷We recall that a lattice in a locally compact Hausdorff topological group G is a discrete subgroup $\Lambda < G$ such that the quotient space $\Lambda \backslash G$ admits a non-zero G -invariant Radon probability measure. The lattice Λ is cocompact if $\Lambda \backslash G$ is a compact topological space.

isomorphism of associative \mathbb{R} -algebras between the universal enveloping algebra $U(\mathfrak{sl}_2(\mathbb{R}))$ of $\mathfrak{sl}_2(\mathbb{R})$ and the algebra of $\mathrm{SL}_2(\mathbb{R})$ -invariant differential operators on $\mathcal{C}^\infty(M)$. More generally, any element $Y \in U(\mathfrak{sl}_2(\mathbb{R}))$ of order $k \in \mathbb{N}$ acts naturally on $\mathcal{C}^k(M)$, the action being also denoted by $f \mapsto Yf$. Observe that, for any $k \in \mathbb{N}$, the definition of the \mathcal{C}^k -norm on $\mathcal{C}^k(M)$ implies that $\|Yf\|_\infty \leq \|f\|_{\mathcal{C}^k}$ for any $Y \in U(\mathfrak{sl}_2(\mathbb{R}))$ of order at most k and any $f \in \mathcal{C}^k(M)$.

2.2. Unitary representations, the Casimir and Laplace operators, Weyl's law. Convenient sources for the material presented in this subsection, in addition to those cited in Section 2.1, are [40, 46, 60, 76].

Let us denote by H be the complex Hilbert space $L^2(M)$ of functions which are square-integrable with respect to the unique $\mathrm{SL}_2(\mathbb{R})$ -invariant Borel probability measure¹⁸ vol on M ; the inner product defining the Hilbert space structure on H shall be denoted by $\langle \cdot, \cdot \rangle$. With $\mathcal{U}(H)$ we indicate the group of unitary operators $T: H \rightarrow H$. The standard smooth (left) action of $\mathrm{SL}_2(\mathbb{R})$ on the homogeneous space M , given by $g_0 \cdot \Gamma g = \Gamma g g_0^{-1}$ for every $g_0, g \in \mathrm{SL}_2(\mathbb{R})$, preserves the measure vol , and hence gives rise to a unitary representation $\rho: \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathcal{U}(H)$, $g \mapsto \rho_g$, called the quasi-regular representation of $\mathrm{SL}_2(\mathbb{R})$ on H .

For any $r \in \mathbb{N} \cup \{\infty\}$, denote by $\mathcal{C}^r(H)$ the linear subspace of \mathcal{C}^r -vectors on H , that is,

$$\mathcal{C}^r(H) = \{v \in H : \text{the map } \mathrm{SL}_2(\mathbb{R}) \ni g \mapsto \rho_g(v) \in H \text{ is of class } \mathcal{C}^r\}.$$

An element $W \in \mathfrak{sl}_2(\mathbb{R})$ induces a linear operator $L_W: \mathcal{C}^1(H) \rightarrow H$, called the Lie derivative with respect to W , defined by

$$L_W(v) = \lim_{t \rightarrow 0} \frac{\rho_{\exp(tW)}(v) - v}{t}, \quad v \in \mathcal{C}^1(H). \quad (2.1)$$

Observe that, for a \mathcal{C}^1 -function f on M , we have $L_W(f) = Wf$, regarding $\mathcal{C}^1(M)$ as canonically embedded¹⁹ in H .

The center of the universal enveloping algebra $U(\mathfrak{sl}_2(\mathbb{R}))$ is generated, as an algebra, by any of its non-zero elements, called Casimir elements. For reasons to be clarified shortly, we choose the normalization of the Casimir element to be $\square = -X^2 + X + UV$, where $\{X, U, V\}$ is identified with a set of generators of the \mathbb{R} -algebra $U(\mathfrak{sl}_2(\mathbb{R}))$. By the universal property of the universal enveloping algebra, the Casimir element \square acts as a second-order linear differential operator $\mathcal{C}^2(M) \rightarrow \mathcal{C}^0(M)$, and (in a compatible manner) as an unbounded operator on H , densely defined on the subspace $\mathcal{C}^2(H)$. Slightly abusing notation, we shall again indicate with \square the Casimir operator arising in this fashion. It is a consequence of the Casimir element lying in the center of $U(\mathfrak{sl}_2(\mathbb{R}))$ that the operator \square commutes with all unitary operators ρ_g , $g \in \mathrm{SL}_2(\mathbb{R})$ and all Lie derivatives L_W , $W \in \mathfrak{sl}_2(\mathbb{R})$, wherever they are simultaneously defined.

The Casimir operator \square is an essentially self-adjoint operator on H ; as M is compact, its spectrum²⁰ $\mathrm{Spec}(\square)$ consists solely of eigenvalues and is a discrete subset of \mathbb{R} . Our choice of the normalization of the Casimir element implies that, on the subspace $\mathcal{C}^2(S)$ of continuously twice-differentiable functions defined on the surface $S = \Gamma \backslash \mathbb{H}$, the Casimir operator acts as the Laplace-Beltrami operator Δ_S associated to the hyperbolic Riemannian metric on S . As a consequence, the spectrum of Δ_S is contained in the spectrum of \square . As a matter of fact, it holds that $\mathrm{Spec}(\Delta_S) = \mathrm{Spec}(\square) \cap \mathbb{R}_{\geq 0}$.

A great deal of research has revolved around the properties of the spectrum of Δ_S ; for the purposes of this article, we shall only need a weaker version (cf. the proof of Lemma 5.1) of the

¹⁸The measure vol is given, on each local trivialization of the unit tangent bundle M , as a product of the normalized hyperbolic area measure on S and of the unique isometrically-invariant Borel probability measure on the unit circle \mathbb{S}^1 .

¹⁹The measure vol is fully supported on M , therefore $\mathcal{C}^0(M)$ embeds linearly in $L^2(M)$.

²⁰Bargmann's well-known classification of unitary representations of $\mathrm{SL}_2(\mathbb{R})$ (see [3]) yields that the spectrum of the Casimir operator classifies the irreducible representations strongly contained in ρ as belonging to the principal series, complementary series or discrete series according to whether the corresponding Casimir eigenvalue μ satisfies $\mu \geq 1/4$, $0 < \mu < 1/4$ or $\mu \leq 0$, respectively. We do not appeal to Bargmann's classification in this manuscript.

celebrated Weyl's law concerning the asymptotics of the eigenvalues, which we shall now recall for completeness. For a finite set A , the notation $|A|$ stands for the cardinality of A .

Theorem 2.1 (Weyl's law). *Let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be as above, $S = \Gamma \backslash \mathbb{H}$, and*

$$\lambda_0 < \lambda_1 \leq \dots \leq \lambda_n \leq \dots \rightarrow \infty$$

the eigenvalues, counted with multiplicity, of the Laplace-Beltrami operator Δ_S associated to the hyperbolic Riemannian metric on S . Then

$$\lim_{R \rightarrow \infty} \frac{|\{j \in \mathbb{N} : \lambda_j \leq R\}|}{R} = \frac{\mathrm{Area}(S)}{4\pi},$$

where $\mathrm{Area}(S)$ is the total mass of S with respect to the hyperbolic area measure.

A proof of Theorem 2.1 relying on Selberg's trace formula is given in [49, Prop. 10].

On the other hand, the negative eigenvalues of \square are completely understood, being of the form $-m(m+2)/4$ with m ranging across the set \mathbb{N}^* of strictly positive integers.

In general, any unitary representation of $\mathrm{SL}_2(\mathbb{R})$ is unitarily equivalent to a direct integral of irreducible representations. Owing to compactness of M , the representation space H of ρ actually decomposes as a Hilbert direct sum of (non-zero) ρ -invariant irreducible subspaces:

$$H = \bigoplus_{i \in I} H_i, \quad \rho(H_i) \subset H_i, \quad \rho|_{H_i} \text{ irreducible, } I \text{ countable index set.}$$

As the Casimir operator commutes with all the ρ_g for $g \in \mathrm{SL}_2(\mathbb{R})$, it restricts to an operator $\mathcal{C}^2(H_i) \rightarrow H_i$ for each $i \in I$. By virtue of Schur's lemma for unbounded linear operators, there exists $\mu \in \mathrm{Spec}(\square)$, depending on i , such that $\square v = \mu v$ for any $v \in \mathcal{C}^2(H_i)$. Regrouping all irreducible subspaces according to the corresponding Casimir eigenvalue, we obtain an orthogonal splitting into ρ -invariant subspaces

$$H = \bigoplus_{\mu \in \mathrm{Spec}(\square)} H_\mu, \quad H_\mu = \{v \in \mathcal{C}^2(H) : \square v = \mu v\}. \quad (2.2)$$

Let $\Theta \in \mathfrak{sl}_2(\mathbb{R})$ be the element defined in (1.3), inducing the unbounded linear operator L_Θ on H . As it is the case for any representation space of a unitary representation of $\mathrm{SL}_2(\mathbb{R})$, each subspace H_μ is the internal Hilbert direct sum²¹

$$H_\mu = \bigoplus_{n \in \mathbb{Z}} H_{\mu,n}, \quad H_{\mu,n} = \left\{ v \in \mathcal{C}^1(H_\mu) : L_\Theta v = \frac{i}{2} n v \right\}. \quad (2.3)$$

Remark 2.2. Observe that, in the previous decomposition, it might occur (as it manifestly emerges in the equality (2.6) below) that $H_{\mu,n}$ is the trivial subspace for some $\mu \in \mathrm{Spec}(\square)$ and $n \in \mathbb{Z}$. For this reason, we let the index set of the direct sum in (2.3) be rather $I(\mu) = \{n \in \mathbb{Z} : H_{\mu,n} \neq \{0\}\} \subset \mathbb{Z}$ in the sequel, bearing in mind this caveat.

2.3. Sobolev spaces and the Sobolev Embedding Theorem. The classical theory of Sobolev spaces is thoroughly discussed in [1]. In the context of arbitrary Riemannian manifolds, it is presented, for instance, in [2, 32]; the extension to fractional Sobolev orders is treated e.g. in [72, 78]. Here we shall confine ourselves to homogeneous spaces of $\mathrm{SL}_2(\mathbb{R})$, placing emphasis on how this theory is brought to bear on the study of unitary representations of the special linear group.

Define the element

$$Y = \begin{pmatrix} 0 & -1/2 \\ -1/2 & 0 \end{pmatrix} \in \mathfrak{sl}_2(\mathbb{R}),$$

so that $\{X, \Theta, Y\}$ forms a basis of the real vector space $\mathfrak{sl}_2(\mathbb{R})$. Define the second-order linear operator $\Delta = -(X^2 + Y^2 + \Theta^2) = \square - 2\Theta^2$ on the subspace $\mathcal{C}^2(H)$. As each Lie derivative

²¹Complete reducibility of their unitary representations is a general property of compact abelian groups, in this case $\mathrm{SO}_2(\mathbb{R})$.

L_W ($W \in \mathfrak{sl}_2(\mathbb{R})$) satisfies²² $\langle L_W u, v \rangle = -\langle u, L_W v \rangle$ for any $u, v \in \mathcal{C}^1(H)$, it is clear that $\langle \Delta u, v \rangle = \langle u, \Delta v \rangle$ for any $u, v \in \mathcal{C}^2(H)$, that is, Δ is self-adjoint on its domain of definition. We shall refer to it as the Laplace operator on H . Akin to the Casimir operator, Δ acts as the Laplace-Beltrami operator Δ_S on the subspace $\mathcal{C}^2(S)$. Indeed, any vector $u \in \mathcal{C}^2(S)$ is invariant under the action of $\mathrm{SO}_2(\mathbb{R})$ by right translations, and hence lies in the kernel of L_Θ ; as a consequence,

$$\Delta u = (\square - 2\Theta^2)u = \square u - 2L_\Theta^2 u = \square u = \Delta_S u .$$

For any $s \in \mathbb{R}_{>0}$, we define the Sobolev space of order s on H , denoted by $W^s(H)$, as the maximal linear subspace of H on which the unbounded linear operator $\Delta^{s/2}$ can be defined, and endow it with the inner product given by

$$\langle u, v \rangle_{W^s} = \langle (I + \Delta)^s u, v \rangle , \quad u, v \in W^s(H), \quad (2.4)$$

where I denotes the identity operator on H . This assignment turns $W^s(H)$ into a Hilbert space, whose associated norm is denoted by $\|\cdot\|_{W^s}$. Similarly, we define the Sobolev spaces $W^s(H_\mu)$ and $W^s(H_{\mu,n})$ for any $\mu \in \mathrm{Spec}(\mathbb{Z})$ and $n \in I(\mu)$. It is a fact that the decompositions in (2.2) and (2.3) induce analogous decompositions on the level of Sobolev spaces, namely there are orthogonal²³ splittings

$$W^s(H) = \bigoplus_{\mu \in \mathrm{Spec}(\square)} W^s(H_\mu) = \bigoplus_{\mu \in \mathrm{Spec}(\square)} \bigoplus_{n \in I(\mu)} W^s(H_{\mu,n}) . \quad (2.5)$$

The argument in Section 5 crucially hinges upon the following elementary relationship between Sobolev norms of different order:

Lemma 2.3. *Let $s \in \mathbb{R}_{>0}$, $k \in \mathbb{N}$, and assume $u \in W^{s+k}(H_{\mu,n})$ for some $\mu \in \mathrm{Spec}(\square)$ and $n \in I(\mu)$. Then*

$$\|u\|_{W^{s+k}}^2 = \left(1 + \mu + \frac{n^2}{2}\right)^k \|u\|_{W^s}^2 . \quad (2.6)$$

Proof. Suppose $k = 1$. We may write, using self-adjointness of Δ with respect to the L^2 -inner product,

$$\|u\|_{W^{s+1}}^2 = \langle u, u \rangle_{W^{s+1}} = \langle (I + \Delta)^{s+1} u, u \rangle = \langle (I + \Delta)^s u, (I + \Delta)u \rangle = \langle (1 + \Delta)^s u, u \rangle + \langle (1 + \Delta)^s u, \Delta u \rangle .$$

By the assumption on u , it holds that $\Delta u = (\square - 2\Theta^2)u = \mu - 2(\frac{i}{2}n)^2 u = (\mu + \frac{n^2}{2})u$. Therefore, we infer

$$\|u\|_{W^{s+1}}^2 = \|u\|_{W^s}^2 + \left(\mu + \frac{n^2}{2}\right) \langle (1 + \Delta)^s u, u \rangle = \left(1 + \mu + \frac{n^2}{2}\right) \|u\|_{W^s}^2 ,$$

as desired.

The statement for an arbitrary $k \in \mathbb{N}$ follows arguing by induction. \square

Observe that (2.6) readily implies the following: if $u \in W^{s+k}(H)$ for some $s \in \mathbb{R}_{>0}$ and $k \in \mathbb{N}$, and

$$u = \sum_{\mu \in \mathrm{Spec}(\square)} \sum_{n \in I(\mu)} u_{\mu,n}$$

is its decomposition into joint eigenvectors for \square and Θ provided by (2.5), then the larger the integer k is, the faster the decay of the Sobolev norms $\|u_{\mu,n}\|_{W^s}$ as $|n|$ and $|\mu|$ tend to infinity. This phenomenon²⁴ is going to be essential in our estimates over the course of the proof of Theorem 1.8.

²²This follows at once from the definition of L_W in (2.1) and the fact that ρ_g is a unitary operator for any $g \in \mathrm{SL}_2(\mathbb{R})$.

²³Clearly, we intend that the closed subspaces $W^s(H_{\mu,n})$ are orthogonal with respect to the W^s -inner product defined in (2.4).

²⁴The counterpart of this relationship in classical Fourier analysis is well-known; the regularity of a function is closely interwoven with the decay rate at infinity of its Fourier coefficients.

We conclude this section recalling a version for compact three-manifolds of the celebrated Sobolev Embedding Theorem, which will be sufficient for our purposes.

Theorem 2.4 (Sobolev Embedding Theorem). *For any $r \in \mathbb{N}$ and $s \in \mathbb{R}_{>0}$ fulfilling the inequality $s - r > 3/2$, there is a continuous embedding of $W^s(M)$ into the Banach space $\mathcal{C}^r(M)$: in particular, there exists a constant $C_{r,s} > 0$ such that*

$$\|f\|_{\mathcal{C}^r} \leq C_{r,s} \|f\|_{W^s}$$

for every $f \in W^s(M)$.

3. REDUCTION TO AN ORDINARY DIFFERENTIAL EQUATION

This section presents the gist of the approach we pursue in order to prove Theorem 1.6, which concerns the asymptotic behaviour of circle-arc averages of joint eigenfunctions of the operators \square and Θ defined in Section 2.2; the partial differential equations expressing the eigenfunction condition, which are classically referred to in the literature as eigenvalue equations for the relevant operator, are here shown to give rise to ordinary differential equations for the corresponding circle-arc averages, when the latter are seen as functions of the time parameter.

We fix a function $f: M \rightarrow \mathbb{C}$ of class \mathcal{C}^2 and a parameter $\theta \in (0, 4\pi]$. Recall from (1.6) the definition of the averages $k_{f,\theta}(p, t)$, for $p \in M$ and $t \in \mathbb{R}$. As the observable f and the base point p are fixed throughout the considerations that follow, we choose to abbreviate $k_{f,\theta}(p, t)$ with $k_\theta(t)$ for notational simplicity.

Our goal is to show that the function $k_\theta(t)$ satisfies a second-order linear ordinary differential equation (ODE). In the upcoming computations, the following lemma will be of use. For any left-invariant vector field $W \in \mathfrak{sl}_2(\mathbb{R})$, we indicate with $(\phi_t^W)_{t \in \mathbb{R}}$ be the one-parameter flow on M defined by $\phi_t^W(\Gamma g) = \Gamma g \exp tW$ for any $t \in \mathbb{R}$ and $g \in \mathrm{SL}_2(\mathbb{R})$. For any pair $Y, W \in \mathfrak{sl}_2(\mathbb{R})$ and any point $q \in M$, the derivative of the smooth curve $s \mapsto \phi_t^Y \circ \phi_s^W(q)$ (seen as a function from \mathbb{R} to the tangent bundle of M), where $t \in \mathbb{R}$ is fixed, is denoted by $\frac{d}{ds} \phi_t^Y \circ \phi_s^W(q)$. Lastly, if $W \in \mathfrak{sl}_2(\mathbb{R})$ and $q \in M$, we denote by W_q the value at q of the infinitesimal generator of the smooth flow $(\phi_t^W)_{t \in \mathbb{R}}$ on M .

Lemma 3.1. *For every $Y, W \in \mathfrak{sl}_2(\mathbb{R}) \setminus \{0\}$ and every $q \in M$, it holds*

$$\frac{d}{ds} \phi_t^Y \circ \phi_s^W(q) = \mathrm{Ad}_{\exp(-tY)}(W)_{\phi_t^Y \circ \phi_s^W(q)}.$$

Proof. It follows from elementary algebraic manipulations, see [58, Lem. 4]. \square

We may now state the crucial result of this section.

Proposition 3.2. *Let $\mu \in \mathrm{Spec}(\square)$, $n \in \mathbb{Z}$ and $f \in \mathcal{C}^2(M)$ be a function satisfying $\square f = \mu f$, $\Theta f = \frac{i}{2}nf$. For every $p \in M$ and $\theta \in (0, 4\pi]$, there is a bounded continuous function $G_{\theta,n}f(p, \cdot): \mathbb{R}_{>0} \rightarrow \mathbb{C}$ such that the function $k_\theta(t) = \frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) ds$ satisfies the linear ordinary differential equation*

$$k_\theta''(t) + k_\theta'(t) + \mu k_\theta(t) = e^{-t} G_{\theta,n}(p, t) \quad (3.1)$$

for any $t > 0$.

Proof. Fix μ, n, f, p and θ as in the assumptions. Since f is of class \mathcal{C}^2 on M , differentiation under the integral sign gives

$$k_\theta'(t) = \frac{1}{\theta} \int_0^\theta Xf \circ \phi_t^X \circ r_s(p) ds, \quad k_\theta''(t) = \frac{1}{\theta} \int_0^\theta X^2f \circ \phi_t^X \circ r_s(p) ds$$

for any $t \in \mathbb{R}$, as the geodesic flow $(\phi_t^X)_{t \in \mathbb{R}}$ on M is generated by the vector field X . Therefore, the assumption $\square f = \mu f$, which when spelled out amounts to

$$-X^2f + Xf - UVf = \mu f,$$

translates readily into

$$k''_{\theta}(t) - k'_{\theta}(t) + \mu k_{\theta}(t) = -\frac{1}{\theta} \int_0^{\theta} UV f \circ \phi_t^X \circ r_s(p) \, ds .$$

As $V = U - 2\Theta$, we have

$$k''_{\theta}(t) - k'_{\theta}(t) + \mu k_{\theta}(t) = -\frac{1}{\theta} \int_0^{\theta} U^2 f \circ \phi_t^X \circ r_s(p) \, ds + \frac{in}{\theta} \int_0^{\theta} U f \circ \phi_t^X \circ r_s(p) \, ds , \quad (3.2)$$

taking into account the assumption $\Theta f = \frac{i}{2} n f$. Now, by Stokes' theorem²⁵, we get that

$$f \circ \phi_t^X \circ r_{\theta}(p) - f \circ \phi_t^X(p) = \int_0^{\theta} \frac{d}{ds} (f \circ \phi_t^X \circ r_s(p)) \, ds = \int_0^{\theta} df_{\phi_t^X \circ r_s(p)} \left(\frac{d}{ds} (\phi_t^X \circ r_s(p)) \right) \, ds ,$$

where $df_{\phi_t^X \circ r_s(p)}$ indicates the differential of f at the point $\phi_t^X \circ r_s(p)$ and the last equality follows from the chain rule for differentials. Recalling that the flow $(r_s)_{s \in \mathbb{R}}$ is generated by the vector field Θ , Lemma 3.1 delivers

$$\frac{d}{ds} \phi_t^X \circ r_s(p) = \text{Ad}_{\exp(-tX)}(\Theta)_{\phi_t^X \circ r_s(p)} ,$$

so that

$$\begin{aligned} f \circ \phi_t^X \circ r_{\theta}(p) - f \circ \phi_t^X(p) &= \int_0^{\theta} df_{\phi_t^X \circ r_s(p)} \left((\text{Ad}_{\exp(-tX)}(\Theta))_{\phi_t^X \circ r_s(p)} \right) \, ds \\ &= \int_0^{\theta} df_{\phi_t^X \circ r_s(p)} \left(((-\sinh t)U + e^t \Theta)_{\phi_t^X \circ r_s(p)} \right) \, ds \\ &= -\sinh t \int_0^{\theta} U f \circ \phi_t^X \circ r_s(p) \, ds + \frac{i}{2} n \theta e^t k_{\theta}(t) , \end{aligned}$$

the second equality being obtained by straightforward matrix multiplications.

From now we choose t strictly positive. We may thus write

$$\int_0^{\theta} U f \circ \phi_t^X \circ r_s(p) \, ds = \frac{1}{\sinh t} \left(\frac{i}{2} n \theta e^t k_{\theta}(t) - A_{\theta}(t) \right) , \quad (3.3)$$

where

$$A_{\theta}(t) = f \circ \phi_t^X \circ r_{\theta}(p) - f \circ \phi_t^X(p) , \quad t > 0 .$$

Arguing as before, we also deduce that

$$\begin{aligned} U f \circ \phi_t^X \circ r_{\theta}(p) - U f \circ \phi_t^X(p) &= \int_0^{\theta} \frac{d}{ds} (U f \circ \phi_t^X \circ r_s(p)) \, ds \\ &= \frac{1}{2} e^{-t} \int_0^{\theta} U^2 f \circ \phi_t^X \circ r_s(p) \, ds - \frac{1}{2} e^t \int_0^{\theta} V U f \circ \phi_t^X \circ r_s(p) \, ds . \end{aligned} \quad (3.4)$$

From $UV - VU = 2X$ we get

$$VU = UV - 2X = U(U - 2\Theta) - 2X = U^2 - 2U\Theta - 2X ,$$

so that

$$\int_0^{\theta} V U f \circ \phi_t^X \circ r_s(p) \, ds = \int_0^{\theta} U^2 f \circ \phi_t^X \circ r_s(p) \, ds - in \int_0^{\theta} U f \circ \phi_t^X \circ r_s(p) \, ds - 2\theta k'_{\theta}(t) . \quad (3.5)$$

²⁵Here, it really boils down to the fundamental theorem of calculus.

Combining (3.3), (3.4) and (3.5) yields

$$\begin{aligned} \int_0^\theta U^2 f \circ \phi_t^X \circ r_s(p) \, ds &= \frac{1}{\sinh t} \left(\frac{i}{2} n e^t \int_0^\theta U f \circ \phi_t^X \circ r_s(p) \, ds + \theta e^t k'_\theta(t) - B_\theta(t) \right) \\ &= \frac{1}{\sinh t} \left(\frac{i n e^t}{4 \sinh t} (i n \theta e^t k_\theta(t) - A_\theta(t)) + \theta e^t k'_\theta(t) - B_\theta(t) \right), \end{aligned} \quad (3.6)$$

where

$$B_\theta(t) = U f \circ \phi_t^X \circ r_\theta(p) - U f \circ \phi_t^X(p), \quad t > 0.$$

From (3.2), (3.3) and (3.6) we thus infer that

$$k''_\theta(t) - k'_\theta(t) + \mu k_\theta(t) = \frac{n^2 e^{2t} k_\theta(t)}{4 \sinh^2 t} + \frac{i n e^t A_\theta(t)}{4 \theta \sinh^2 t} - \frac{e^t k'_\theta(t)}{\sinh t} - \frac{n^2 e^t k_\theta(t)}{2 \sinh t} + \frac{-i n A_\theta(t) + B_\theta(t)}{\theta \sinh t}.$$

Finally, adding $2k'_\theta(t)$ on both sides gives

$$\begin{aligned} k''_\theta(t) + k'_\theta(t) + \mu k_\theta(t) &= \left(\frac{n^2}{(1 - e^{-2t})^2} - \frac{n^2}{1 - e^{-2t}} \right) k_\theta(t) + \left(2 - \frac{2}{1 - e^{-2t}} \right) k'_\theta(t) \\ &\quad + \left(\frac{i n}{2 \theta \sinh t (1 - e^{-2t})} - \frac{i n}{\theta \sinh t} \right) A_\theta(t) + \frac{B_\theta(t)}{\theta \sinh t}, \end{aligned}$$

so that the ODE (3.1) in the statement is satisfied for

$$G_{\theta,n} f(p, t) = \frac{n^2 e^{-t}}{(1 - e^{-2t})^2} k_\theta(t) - \frac{2 e^{-t}}{1 - e^{-2t}} k'_\theta(t) + \frac{2 i n e^{-2t}}{\theta (1 - e^{-2t})^2} A_\theta(t) + \frac{2}{\theta (1 - e^{-2t})} B_\theta(t), \quad t > 0. \quad (3.7)$$

The function f being of class \mathcal{C}^2 , it is clear that the functions $k_\theta, k'_\theta, A_\theta$ and B_θ are continuous, hence so is the function $t \mapsto G_{\theta,n} f(p, t)$. Furthermore, the trivial upper bounds

$$|k_\theta(t)| \leq \|f\|_\infty, \quad |k'_\theta(t)| \leq \|Xf\|_\infty, \quad |A_\theta(t)| \leq 2 \|f\|_\infty, \quad |B_\theta(t)| \leq 2 \|Uf\|_\infty$$

imply that the latter function is uniformly bounded on $\mathbb{R}_{>0}$. \square

For later purposes, we estimate explicitly the supremum norm of $G_{\theta,n} f$. Using the bounds $e^{-t} \leq 1$ and $1 - e^{-2t} \geq 1 - e^{-1}$, valid for all $t \geq 1/2$, together with the fact that the three quantities $\|f\|_\infty, \|Xf\|_\infty, \|Uf\|_\infty$ are bounded from above by $\|f\|_{\mathcal{C}^1}$ (see Section 2.1), we obtain that

$$\sup_{t \geq 1/2} |G_{\theta,n} f(p, t)| \leq C_{\theta,n} \|f\|_{\mathcal{C}^1}$$

where $C_{\theta,n}$ may be chosen to be equal to

$$C_{\theta,n} = \left(\frac{e}{e-1} \right)^2 \frac{n(\theta n + 2)}{\theta} + \frac{e}{e-1} \frac{2\theta + 2}{\theta}.$$

Setting

$$\kappa_0 = \frac{2e^2(1 + 4\pi)}{(e-1)^2},$$

we may estimate

$$C_{\theta,n} \leq \left(\frac{e}{e-1} \right)^2 \frac{2(\theta + 1)}{\theta} (n^2 + 1) \leq \frac{\kappa_0}{\theta} (n^2 + 1),$$

and hence conclude that

$$\sup_{t \geq 1/2} |G_{\theta,n} f(p, t)| \leq \frac{\kappa_0}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} \quad (3.8)$$

for every choice of $n \in \mathbb{Z}$, $f \in \mathcal{C}^2(M)$ satisfying $\Theta f = \frac{i}{2} n f$, $p \in M$, and $\theta \in (0, 4\pi]$.

Remark 3.3. The quadratic growth in n of the estimate (3.8) shall play a significant role in the generalization of Theorem 1.6 to arbitrary observables, carried out in Section 5; specifically, it will impact the minimal Sobolev regularity we need to impose on the test functions.

4. ASYMPTOTICS FOR JOINT EIGENFUNCTIONS

The purpose of this section is to prove Theorem 1.6 by explicitly solving the ordinary differential equation established in Proposition 3.2. For definiteness, we choose to impose initial conditions at time $t = 1$ for the ensuing Cauchy problem.

We thus start with the following:

Lemma 4.1. *Let μ be an eigenvalue of the Casimir operator. If $G: \mathbb{R}_{>0} \rightarrow \mathbb{C}$ is a continuous function, then for any complex numbers y_1 and y'_1 the solution to the Cauchy problem*

$$\begin{cases} y''(t) + y'(t) + \mu y(t) = e^{-t}G(t) \\ y(1) = y_1 \\ y'(1) = y'_1 \end{cases} \quad (4.1)$$

is given by

$$\begin{aligned} y(t) = & e^{-\frac{1-\nu}{2}t} \left(\frac{(1+\nu)y_1 + 2y'_1}{2\nu e^{-\frac{1-\nu}{2}}} + \frac{1}{\nu} \int_1^t e^{-\frac{1+\nu}{2}\xi} G(\xi) \, d\xi \right) \\ & - e^{-\frac{1+\nu}{2}t} \left(\frac{(1-\nu)y_1 + 2y'_1}{2\nu e^{-\frac{1+\nu}{2}}} + \frac{1}{\nu} \int_1^t e^{-\frac{1-\nu}{2}\xi} G(\xi) \, d\xi \right) \end{aligned} \quad (4.2)$$

if $\mu \neq 1/4$, and by

$$\begin{aligned} y(t) = & e^{-\frac{t}{2}} \left(\frac{\sqrt{e}(y_1 - 2y'_1)}{2} - \int_1^t \xi e^{-\frac{\xi}{2}} G(\xi) \, d\xi \right) \\ & + t e^{-\frac{t}{2}} \left(\frac{\sqrt{e}(y_1 + 2y'_1)}{2} + \int_1^t e^{-\frac{\xi}{2}} G(\xi) \, d\xi \right) \end{aligned} \quad (4.3)$$

if $\mu = 1/4$.

Proof. Let ν be the unique complex number in $\mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ such that $1 - \nu^2 = 4\mu$. The characteristic polynomial of the associated homogeneous equation

$$y''(t) + y'(t) + \mu y(t) = 0$$

is $P(Z) = Z^2 + Z + \mu$, having two distinct roots $-\frac{1-\nu}{2}$, $-\frac{1+\nu}{2}$ if $\mu \neq 1/4$, and a double root $-\frac{1}{2}$ if $\mu = 1/4$. We examine the case $\mu \neq 1/4$; the case $\mu = 1/4$ requires only minor modifications. A particular solution of the inhomogeneous equation is given by

$$e^{-\frac{1-\nu}{2}t} + \int_1^t \frac{1}{\nu} e^{-\frac{1+\nu}{2}\xi} G(\xi) \, d\xi - e^{-\frac{1+\nu}{2}t} \int_1^t \frac{1}{\nu} e^{-\frac{1-\nu}{2}\xi} G(\xi) \, d\xi$$

as direct computations allow to verify. Hence, the general solution of

$$y''(t) + y'(t) + \mu y(t) = e^{-t}G(t)$$

is given by

$$y(t) = e^{-\frac{1-\nu}{2}t} \left(c_1 + \frac{1}{\nu} \int_1^t e^{-\frac{1+\nu}{2}\xi} G(\xi) \, d\xi \right) + e^{-\frac{1+\nu}{2}t} \left(c_2 - \frac{1}{\nu} \int_1^t e^{-\frac{1-\nu}{2}\xi} G(\xi) \, d\xi \right), \quad c_1, c_2 \in \mathbb{C}.$$

Imposing the initial conditions $y(1) = y_1, y'(1) = y'_1$ enables to determine the coefficients

$$c_1 = \frac{(1+\nu)y_1 + 2y'_1}{2\nu e^{-\frac{1-\nu}{2}}}, \quad c_2 = -\frac{(1-\nu)y_1 + 2y'_1}{2\nu e^{-\frac{1+\nu}{2}}}.$$

□

We may now apply Proposition 3.2 in conjunction with Lemma 4.1 to determine the explicit analytic expression of the circle-arc average $k_{f,\theta}(p, t)$, defined in (1.6), in terms of the coefficient $G_{\theta,n}f(p, t)$ (cf. (3.7)). Before proceeding with this, it will be convenient to introduce some useful notation first.

Let μ be a Casimir eigenvalue, f a function on M of class \mathcal{C}^2 and $\theta \in (0, 4\pi]$. Define the functions $a_{\theta,\mu}^+, a_{\theta,\mu}^- f: M \rightarrow \mathbb{C}$ by

$$a_{\theta,\mu}^\pm f(p) = \mp \frac{(1 \mp \nu)\theta^{-1} \int_0^\theta f \circ \phi_1^X \circ r_s(p) \, ds + 2\theta^{-1} \int_0^\theta Xf \circ \phi_1^X \circ r_s(p) \, ds}{2\nu e^{-\frac{1\pm\nu}{2}}} \quad (4.4)$$

if $\mu \neq 1/4$, and by

$$a_{\theta,1/4}^\pm f(p) = \frac{\sqrt{e}(\theta^{-1} \int_0^\theta f \circ \phi_1^X \circ r_s(p) \, ds \mp 2\theta^{-1} \int_0^\theta Xf \circ \phi_1^X \circ r_s(p) \, ds)}{2} \quad (4.5)$$

if $\mu = 1/4$. When $\mu \neq 1/4$, it holds

$$\|a_{\theta,\mu}^\pm f\|_\infty \leq \frac{(1 + |\nu|)\theta^{-1} \int_0^\theta \|f\|_\infty \, ds + 2\theta^{-1} \int_0^\theta \|Xf\|_\infty \, ds}{2e^{-1}|\nu|} \leq \frac{e(3 + |\nu|)}{2|\nu|} \|f\|_{\mathcal{C}^1}, \quad (4.6)$$

for $\|f\|_\infty \leq \|f\|_{\mathcal{C}^1}$ and $\|Xf\|_\infty \leq \|f\|_{\mathcal{C}^1}$. If $\mu = 1/4$, similar estimates lead readily to

$$\|a_{\theta,1/4}^\pm f\|_\infty \leq \frac{3\sqrt{e}}{2} \|f\|_{\mathcal{C}^1}. \quad (4.7)$$

We are now in a position to start the proof of Theorem 1.6, which will occupy the remainder of this section. We thus fix $\theta \in (0, 4\pi]$ and a function $f \in \mathcal{C}^2(M)$ satisfying $\Theta f = \mu f$ and $\Theta f = \frac{i}{2} n f$ for some $\mu \in \text{Spec}(\square)$ and $n \in \mathbb{Z}$. For any $p \in M$, the function $k_{f,\theta}(p, \cdot): \mathbb{R}_{>0} \rightarrow \mathbb{C}$ we are interested in satisfies the Cauchy problem (4.1) with initial conditions

$$y_1 = \frac{1}{\theta} \int_0^\theta f \circ \phi_1^X \circ r_s(p) \, ds, \quad y_1' = \frac{1}{\theta} \int_0^\theta Xf \circ \phi_1^X \circ r_s(p) \, ds.$$

We distinguish five cases as in the statement of Theorem 1.6, that is, according to the value of the Casimir eigenvalue μ . Recall that ν is the unique complex number in $\mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ verifying $1 - \nu^2 = 4\mu$.

4.1. The case $\mu > 1/4$. Suppose $\mu > 1/4$, so that $\nu = i\Im\nu \in i\mathbb{R}_{>0}$. As indicated in (4.2), the solution to the ODE (3.1) with the prescribed initial conditions is given by

$$\begin{aligned} k_{f,\theta}(p, t) = & e^{-\frac{t}{2}} \cos\left(\frac{\Im\nu}{2}t\right) \left(a_{\theta,\mu}^+ f(p) + a_{\theta,\mu}^- f(p) - \frac{2}{\Im\nu} \int_1^t e^{-\frac{\xi}{2}} \sin\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi \right) \\ & + e^{-\frac{t}{2}} \sin\left(\frac{\Im\nu}{2}t\right) \left(a_{\theta,\mu}^- f(p) - a_{\theta,\mu}^+ f(p) - \frac{2i}{\Im\nu} \int_1^t e^{-\frac{\xi}{2}} \cos\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi \right) \end{aligned}$$

for every $t > 0$. The functions

$$e^{-\xi/2} \cos\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi), \quad e^{-\xi/2} \sin\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi)$$

are integrable over the closed half-line $[1, +\infty)$, as $G_{\theta,n} f(p, \cdot)$ is bounded thereon (see (3.8)); we may therefore define functions $D_{\theta,\mu,n}^+, D_{\theta,\mu,n}^- f: M \rightarrow \mathbb{C}$ by setting

$$D_{\theta,\mu,n}^+ f(p) = a_{\theta,\mu}^+ f(p) + a_{\theta,\mu}^- f(p) - \frac{2}{\Im\nu} \int_1^\infty e^{-\frac{\xi}{2}} \sin\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi \quad (4.8)$$

and

$$D_{\theta,\mu,n}^- f(p) = a_{\theta,\mu}^- f(p) - a_{\theta,\mu}^+ f(p) - \frac{2i}{\Im\nu} \int_1^\infty e^{-\frac{\xi}{2}} \cos\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi \quad (4.9)$$

for every $p \in M$. Then the expansion (1.7) is valid with

$$\begin{aligned} \mathcal{R}_{\theta,\mu,n} f(p, t) = & e^{-\frac{t}{2}} \cos\left(\frac{\Im\nu}{2}t\right) \int_t^\infty \frac{2}{\Im\nu} e^{-\frac{\xi}{2}} \sin\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi \\ & + e^{-\frac{t}{2}} \sin\left(\frac{\Im\nu}{2}t\right) \int_t^\infty \frac{2i}{\Im\nu} e^{-\frac{\xi}{2}} \cos\left(\frac{\Im\nu}{2}\xi\right) G_{\theta,n} f(p, \xi) \, d\xi, \quad t \geq 1, p \in M. \end{aligned} \quad (4.10)$$

Let us now estimate the uniform norms of $D_{\theta,\mu,n}^\pm f$ and of $\mathcal{R}_{\theta,\mu,n}f(\cdot, t)$ for any $t \geq 1$. From the explicit expressions in (4.8) and (4.9), it follows at once that

$$\begin{aligned} \|D_{\theta,\mu,n}^\pm f\|_\infty &\leq \|a_{\theta,\mu}^+ f\|_\infty + \|a_{\theta,\mu}^- f\|_\infty + \frac{2}{\Im\nu} \int_1^\infty e^{-\frac{\xi}{2}} \sup_{p \in M, \xi \geq 1} |G_{\theta,n}f(p, \xi)| \, d\xi \\ &\leq \frac{1}{\Im\nu} \left(e(3 + \Im\nu) + \frac{4\kappa_0(n^2 + 1)}{\theta\sqrt{e}} \right) \|f\|_{\mathcal{C}^1} \\ &\leq \frac{\kappa(\mu)}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} \end{aligned} \quad (4.11)$$

applying (3.8) and (4.6) in the second inequality, with $\kappa(\mu) = \frac{1}{\Im\nu} (4e\pi(3 + \Im\nu) + \frac{4\kappa_0}{\sqrt{e}})$.

The remainder term defined in (4.10) is bounded from above by

$$|\mathcal{R}_{\theta,\mu,n}f(p, t)| \leq 2e^{-\frac{t}{2}} \sup_{p \in M, \xi \geq 1} |G_{\theta,n}f(p, \xi)| \int_t^\infty \frac{2}{\Im\nu} e^{-\frac{\xi}{2}} \, d\xi \leq \frac{8\kappa_0(n^2 + 1)}{\theta \Im\nu} \|f\|_{\mathcal{C}^1} e^{-t}$$

for every $p \in M$ and $t \geq 1$, again relying on the upper bound in (3.8).

Up to the regularity claims on the coefficients $D_{\theta,\mu,n}^\pm f$, which are the subject of Section 4.6, the proof of Theorem 1.6(1) is complete.

4.2. The case $\mu = 1/4$. Suppose $\mu = 1/4$, whence $\nu = 0$. This time the solution to the ODE (3.1) with the given initial conditions has the expression (see (4.3))

$$k_{f,\theta}(p, t) = e^{-\frac{t}{2}} \left(a_{\theta,1/4}^+ f(p) - \int_1^t \xi e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi \right) + te^{-\frac{t}{2}} \left(a_{\theta,1/4}^- f(p) + \int_1^t e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi \right)$$

for every $t > 0$. Following the steps carried out in Section 4.1 almost verbatim, define functions $D_{\theta,1/4,n}^+, D_{\theta,1/4,n}^- : M \rightarrow \mathbb{C}$ via

$$D_{\theta,1/4,n}^+ f(p) = a_{\theta,1/4}^+ f(p) - \int_1^\infty \xi e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi, \quad D_{\theta,1/4,n}^- f(p) = a_{\theta,1/4}^- f(p) + \int_1^\infty e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi \quad (4.12)$$

for every $p \in M$. Then the expansion (1.9) holds with

$$\mathcal{R}_{\theta,1/4,n}f(p, t) = e^{-\frac{t}{2}} \int_t^\infty \xi e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi - te^{-\frac{t}{2}} \int_t^\infty e^{-\frac{\xi}{2}} G_{\theta,n}f(p, \xi) \, d\xi, \quad t \geq 1, p \in M. \quad (4.13)$$

From (4.12) we estimate, by virtue of (3.8) and (4.7),

$$\begin{aligned} \|D_{\theta,1/4,n}^\pm f\|_\infty &\leq \frac{3\sqrt{e}}{2} \|f\|_{\mathcal{C}^1} + \frac{\kappa_0}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} \int_1^\infty \xi e^{-\frac{\xi}{2}} \, d\xi \\ &= \left(\frac{3\sqrt{e}}{2} + \frac{6\kappa_0}{\theta\sqrt{e}} (n^2 + 1) \right) \|f\|_{\mathcal{C}^1} \leq \frac{\kappa(1/4)}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1}, \end{aligned}$$

where we may choose $\kappa(1/4) = 36\pi\sqrt{e}\kappa_0$. Moreover, we deduce from (4.13) that, for every $p \in M$ and $t \geq 1$,

$$\begin{aligned} |\mathcal{R}_{\theta,1/4,n}f(p, t)| &\leq \frac{\kappa_0}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} \left(e^{-\frac{t}{2}} \int_t^\infty \xi e^{-\frac{\xi}{2}} \, d\xi + te^{-\frac{t}{2}} \int_t^\infty e^{-\frac{\xi}{2}} \, d\xi \right) \\ &= \frac{4\kappa_0}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} (t + 1) e^{-t}. \end{aligned}$$

In conjunction with the forthcoming Section 4.6, this concludes the proof of Theorem 1.6(2).

4.3. The case $0 < \mu < 1/4$. When $0 < \mu < 1/4$, we have $\nu \in (0, 1)$. The general solution to the ODE (3.1) given in (4.2) specifies to

$$\begin{aligned} k_{f,\theta}(p, t) = & e^{-\frac{1+\nu}{2}t} \left(a_{\theta,\mu}^+ f(p) - \frac{1}{\nu} \int_1^t e^{-\frac{1-\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi \right) \\ & + e^{-\frac{1-\nu}{2}t} \left(a_{\theta,\mu}^- f(p) + \frac{1}{\nu} \int_1^t e^{-\frac{1+\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi \right) \end{aligned}$$

for every $t > 0$. Setting

$$\begin{aligned} D_{\theta,\mu,n}^+ f(p) &= a_{\theta,\mu}^+ f(p) - \frac{1}{\nu} \int_1^\infty e^{-\frac{1-\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi, \\ D_{\theta,\mu,n}^- f(p) &= a_{\theta,\mu}^- f(p) + \frac{1}{\nu} \int_1^\infty e^{-\frac{1+\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi \end{aligned} \quad (4.14)$$

and

$$\mathcal{R}_{\theta,\mu,n} f(p, t) = \frac{1}{\nu} \left(e^{-\frac{1+\nu}{2}t} \int_t^\infty e^{-\frac{1-\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi - e^{-\frac{1-\nu}{2}t} \int_t^\infty e^{-\frac{1+\nu}{2}\xi} G_{\theta,n} f(p, \xi) \, d\xi \right)$$

for every $p \in M$ and $t \geq 1$, it is clear that the expansion (1.10) holds. As far as estimates on the supremum norm are concerned, we have

$$\begin{aligned} \|D_{\theta,\mu,n}^\pm f\|_\infty &\leq \frac{e(3+\nu)}{2\nu} \|f\|_{\mathcal{C}^1} + \frac{\kappa_0}{\theta\nu} (n^2 + 1) \|f\|_{\mathcal{C}^1} \int_1^\infty e^{-\frac{1\mp\nu}{2}\xi} \, d\xi \\ &= \frac{1}{\nu} \left(\frac{e(3+\nu)}{2} + \frac{\kappa_0}{\theta} \frac{2}{1\mp\nu} e^{-\frac{1\mp\nu}{2}} (n^2 + 1) \right) \|f\|_{\mathcal{C}^1} \\ &\leq \frac{\kappa(\mu)}{\theta} (n^2 + 1) \|f\|_{\mathcal{C}^1} \end{aligned}$$

with $\kappa(\mu) = \frac{2\kappa_0 e^{-\frac{1-\nu}{2}}}{\nu(1-\nu)} + \frac{2e\pi(3+\nu)}{\nu}$, and

$$\begin{aligned} |\mathcal{R}_{\theta,\mu,n} f(p, t)| &\leq \frac{\kappa_0}{\theta\nu} (n^2 + 1) \|f\|_{\mathcal{C}^1} \left(e^{-\frac{1+\nu}{2}t} \int_t^\infty e^{-\frac{1-\nu}{2}\xi} \, d\xi + e^{-\frac{1-\nu}{2}t} \int_t^\infty e^{-\frac{1+\nu}{2}\xi} \, d\xi \right) \\ &= \frac{4\kappa_0}{\theta\nu(1-\nu)(1+\nu)} (n^2 + 1) \|f\|_{\mathcal{C}^1} e^{-t}, \end{aligned}$$

which achieves the proof of Theorem 1.6(3) except for the regularity of the coefficients which is addressed separately in Section 4.6.

4.4. The case $\mu = 0$. As $\nu = 1$ when $\mu = 0$, equation (4.2) delivers the following expression for the solution to the ODE (3.1):

$$\begin{aligned} k_{f,\theta}(p, t) &= a_{\theta,0}^- f(p) + \int_1^t e^{-\xi} G_{\theta,n} f(p, \xi) \, d\xi + e^{-t} \left(a_{\theta,0}^+ f(p) - \int_1^t G_{\theta,n} f(p, \xi) \, d\xi \right) \\ &= a_{\theta,0}^- f(p) + \int_1^\infty e^{-\xi} G_{\theta,n} f(p, \xi) \, d\xi + \left(- \int_t^\infty e^{-\xi} G_{\theta,n} f(p, \xi) \, d\xi \right. \\ &\quad \left. + e^{-t} a_{\theta,0}^+ f(p) - e^{-t} \int_1^t G_{\theta,n} f(p, \xi) \, d\xi \right) \end{aligned} \quad (4.15)$$

for every $t > 0$. Observe that the term between parentheses in the last expression is infinitesimal as t tends to infinity, so that $k_{f,\theta}(p, t)$ has a well-defined limit, as t tends to infinity, for every

$p \in M$. We claim that²⁶

$$\lim_{t \rightarrow \infty} k_{f,\theta}(p, t) = \int_M f \, d \text{vol} \quad (4.16)$$

for any $p \in M$. For a start, we show the equality in (4.16) holds on average with respect to the measure vol . Indeed, Fubini's theorem gives, for any $t > 0$,

$$\begin{aligned} \int_M k_{f,\theta}(p, t) \, d \text{vol}(p) &= \frac{1}{\theta} \int_M \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds \, d \text{vol}(p) = \frac{1}{\theta} \int_0^\theta \int_M f \circ \phi_t^X \circ r_s(p) \, d \text{vol}(p) \, ds \\ &= \frac{1}{\theta} \int_0^\theta \int_M f \, d \text{vol} \, ds = \int_M f \, d \text{vol} , \end{aligned}$$

the second-to-last equality following from the fact that the transformation $\phi_t^X \circ r_s: M \rightarrow M$ preserves the measure vol for any $s, t \in \mathbb{R}$. By dominated convergence,

$$\int_M \lim_{t \rightarrow \infty} k_{f,\theta}(p, t) \, d \text{vol}(p) = \lim_{t \rightarrow \infty} \int_M k_{f,\theta}(p, t) \, d \text{vol}(p) = \lim_{t \rightarrow \infty} \int_M f \, d \text{vol} = \int_M f \, d \text{vol} . \quad (4.17)$$

In order to finish the proof of the claim, it remains to show that the limit $\lim_{t \rightarrow \infty} k_{f,\theta}(p, t)$ does not depend on p . Choose a countable orthonormal basis $(u_k)_{k \in I}$ of $L^2(M)$ consisting of smooth eigenfunctions²⁷ of the operator Θ and containing a constant function u_{k_0} . If $\Theta u_k = \frac{i}{2} n_k u_k$ for $n_k \in \mathbb{Z}$, then $u_k \circ r_s = e^{2\pi i n_k s} u_k$ for every $s \in \mathbb{R}$ (see Section 2.2). Therefore, we can compute for every $k \in I$ the L^2 -inner product

$$\begin{aligned} \int_M \lim_{t \rightarrow \infty} k_{f,\theta}(p, t) \overline{u_k(p)} \, d \text{vol}(p) &= \lim_{t \rightarrow \infty} \int_M \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds \right) \overline{u_k(p)} \, dp \\ &= \lim_{t \rightarrow \infty} \frac{1}{\theta} \int_0^\theta \int_M f \circ \phi_t^X \circ r_s(p) \overline{u_k(p)} \, dp \, ds \\ &= \lim_{t \rightarrow \infty} \frac{1}{\theta} \int_0^\theta \int_M f \circ \phi_t^X(p) \overline{u_k \circ r_{-s}(p)} \, dp \, ds \\ &= \lim_{t \rightarrow \infty} \frac{1}{\theta} \int_0^\theta e^{-2\pi i n_k s} \int_M f \circ \phi_t^X(p) \overline{u_k(p)} \, dp \, ds \\ &= \frac{1}{\theta} \int_0^\theta e^{-2\pi i n_k s} \, ds \cdot \lim_{t \rightarrow \infty} \langle f \circ \phi_t^X, u_k \rangle , \end{aligned}$$

where we used, in successive order, the dominated convergence theorem, Fubini's theorem, invariance of the measure vol under the transformation r_{-s} and the defining property of u_k . Mixing of the geodesic flow $(\phi_t^X)_{t \in \mathbb{R}}$ on M ([4, Cor. 2.3]) delivers

$$\lim_{t \rightarrow \infty} \langle f \circ \phi_t^X, u_k \rangle = \int_M f \, d \text{vol} \int_M u_k \, d \text{vol} .$$

As u_k is orthogonal in $L^2(M)$ to the constant u_{k_0} for any $k \neq k_0$, the last expression vanishes for any such k . Therefore, we have just shown that the function

$$p \mapsto \lim_{t \rightarrow \infty} k_{f,\theta}(p, t)$$

is orthogonal to u_k for every $k \neq k_0$; necessarily, it must be constant, which achieves the proof of our claim.

²⁶The claim amounts to the qualitative equidistribution statement that circle-arc averages of f converge to its spatial average with respect to the uniform measure vol . This has been shown by Margulis (for complete circles) in [48] via a thickening argument resting on the mixing properties of the geodesic flow. We prefer not to invoke Margulis' result here, and instead prove directly this version of equidistribution using spectral considerations coupled with mixing.

²⁷It is actually a fact that $L^2(M)$ admits an orthonormal basis consisting of smooth joint eigenfunctions of the operators \square and Θ .

Define now

$$\mathcal{R}_{\theta,0,n}f(p,t) = a_{\theta,0}^+f(p)e^{-t} - \int_t^\infty e^{-\xi}G_{\theta,n}f(p,\xi) d\xi$$

for every $p \in M$ and $t \geq 1$; then (3.8) and (4.6) give

$$|\mathcal{R}_{\theta,0,n}f(p,t)| \leq 2e\|f\|_{\mathcal{C}^1}e^{-t} + \frac{\kappa_0}{\theta}(n^2+1)\|f\|_{\mathcal{C}^1} \int_t^\infty e^{-\xi} d\xi = \frac{1}{\theta}(8e\pi + \kappa_0(n^2+1))\|f\|_{\mathcal{C}^1}e^{-t};$$

moreover, combining (4.15) with (4.16), we obtain in (1.12), which establishes Theorem 1.6(4).

4.5. The case $\mu < 0$. We finally turn to the case $\mu < 0$ or, equivalently, $\nu > 1$. The solution to the ODE (3.1) given in (4.2) becomes

$$\begin{aligned} k_{f,\theta}(p,t) = & e^{\frac{\nu-1}{2}t} \left(a_{\theta,\mu}^-f(p) + \frac{1}{\nu} \int_1^t e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right) \\ & + e^{-\frac{\nu+1}{2}t} \left(a_{\theta,\mu}^+f(p) - \frac{1}{\nu} \int_1^t e^{-\frac{\nu-1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right) \end{aligned} \quad (4.18)$$

for every $t > 0$. It follows that the quantity

$$\begin{aligned} e^{\frac{\nu-1}{2}t} \left(a_{\theta,\mu}^-f(p) + \frac{1}{\nu} \int_1^\infty e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right) = & -e^{-\frac{\nu+1}{2}t} \left(a_{\theta,\mu}^+f(p) - \frac{1}{\nu} \int_1^t e^{-\frac{\nu-1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right) \\ & - k_{f,\theta}(p,t) + \frac{e^{\frac{\nu-1}{2}t}}{\nu} \int_t^\infty e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \end{aligned}$$

is uniformly bounded in t , which forces

$$a_{\theta,\mu}^-f(p) + \frac{1}{\nu} \int_1^\infty e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi = 0$$

for every $p \in M$. Therefore (4.18) results in

$$k_{f,\theta}(p,t) = e^{-t} \left(-\frac{e^{\frac{\nu+1}{2}t}}{\nu} \int_t^\infty e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi + e^{-\frac{\nu-1}{2}t} a_{\theta,\mu}^+f(p) - \frac{e^{-\frac{\nu-1}{2}t}}{\nu} \int_1^t e^{\frac{\nu-1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right).$$

With the help of (3.8) and (4.6), we may estimate the three summands inside the parentheses. For the first, we have

$$\left| \frac{e^{\frac{\nu+1}{2}t}}{\nu} \int_t^\infty e^{-\frac{\nu+1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right| \leq \frac{2\kappa_0}{\theta\nu(\nu+1)}(n^2+1)\|f\|_{\mathcal{C}^1},$$

while the second can be bounded as

$$\left| e^{-\frac{\nu-1}{2}t} a_{\theta,\mu}^+f(p) \right| \leq \frac{e(3+\nu)}{2\nu} \|f\|_{\mathcal{C}^1};$$

lastly,

$$\begin{aligned} \left| \frac{e^{-\frac{\nu-1}{2}t}}{\nu} \int_1^t e^{\frac{\nu-1}{2}\xi} G_{\theta,n}f(p,\xi) d\xi \right| & \leq \frac{2\kappa_0}{\theta\nu(\nu-1)}(n^2+1)\|f\|_{\mathcal{C}^1} e^{-\frac{\nu-1}{2}t} (e^{\frac{\nu-1}{2}t} - e^{-\frac{\nu-1}{2}t}) \\ & \leq \frac{2\kappa_0}{\theta\nu(\nu-1)}(n^2+1)\|f\|_{\mathcal{C}^1}. \end{aligned}$$

We conclude that

$$|k_{f,\theta}(p,t)| \leq \frac{1}{\theta} \left(\frac{4\kappa_0}{(\nu-1)(\nu+1)} + \frac{2e\pi(3+\nu)}{\nu} \right) (n^2+1)\|f\|_{\mathcal{C}^1} e^{-t},$$

as desired.

This settles Theorem 1.6(5).

4.6. Regularity of the coefficients in the asymptotic expansion. We now turn to the examination of the regularity properties of the coefficients $D_{\theta, \mu, n}^{\pm} f$ appearing in the asymptotic expansion of $k_{f, \theta}(p, t)$, as in Theorem 1.6, for $f \in \mathcal{C}^2(M)$ satisfying $\square f = \mu f$ and $\Theta f = \frac{i}{2} n f$ for some $\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}$ and $n \in \mathbb{Z}$. In so doing, we shall complete the proof of Theorem 1.6.

Let us fix $\theta \in (0, 4\pi]$ throughout this subsection. As follows readily from the definitions of the coefficients (4.4), (4.5), (4.8), (4.9), (4.12) and (4.14), it suffices to analyze the regularity of

$$k_{f, \theta}(p, 1) = \int_0^{\theta} f \circ \phi_1^X \circ r_s(p) \, ds, \quad k'_{f, \theta}(p, 1) = \int_0^{\theta} X f \circ \phi_1^X \circ r_s(p) \, ds, \quad \int_1^{\infty} g(\xi) G_{\theta, n} f(p, \xi) \, d\xi$$

as functions of $p \in M$, with $g(\xi)$ being a function of the following forms:

$$e^{-\xi/2}, \quad \xi e^{-\xi/2}, \quad e^{-\xi/2} \cos\left(\frac{\Im \nu}{2} \xi\right), \quad e^{-\xi/2} \sin\left(\frac{\Im \nu}{2} \xi\right), \quad e^{-\frac{1 \pm \nu}{2} \xi} \quad (0 < \nu < 1). \quad (4.19)$$

We start with the following elementary lemma.

Lemma 4.2. *Let θ be a positive real number. If $h: M \rightarrow \mathbb{C}$ is of class \mathcal{C}^1 , then the function*

$$p \mapsto \int_0^{\theta} h \circ r_s(p) \, ds, \quad p \in M$$

is of class \mathcal{C}^1 on M .

Proof. Fix a point $p_0 \in M$, and let $(\partial_{x_i})_{i=1,2,3}$ be a local frame of the tangent bundle TM around p_0 . It suffices to prove that, for each $i = 1, 2, 3$, the partial derivative

$$p \mapsto \partial_{x_i}|_p \left(\int_0^{\theta} h \circ r_s \, ds \right) \quad (4.20)$$

exists and is continuous in an open neighborhood of p . Upon passing to local smooth charts for M , the classical theorem of differentiation under the integral sign ensures the validity of the formal passage

$$p \mapsto \partial_{x_i}|_p \left(\int_0^{\theta} h \circ r_s \, ds \right) = \int_0^{\theta} \partial_{x_i}|_p (h \circ r_s) \, ds,$$

provided that there exists a positive real-valued function φ on $[0, \theta]$, integrable with respect to the Lebesgue measure, such that $|\partial_{x_i}|_q (h \circ r_s)| \leq \varphi(s)$ for any q in an open neighborhood of p . Notice that, by the dominated convergence theorem, this would yield continuity of the partial derivative in (4.20) at the same time. The chain rule for the differential gives $d(h \circ r_s)_q = (dh)_{r_s(q)} \circ (dr_s)_q$ for any $q \in M$ and $s \in [0, \theta]$. As follows readily from the explicit expression for r_s in (1.3) and direct computations, the operator norm²⁸ of the linear operator $(dr_s)_q$ is uniformly bounded in q and s , as the entries of any Jacobian matrix associated to it only involve finite linear combinations of the sine and cosine functions. Therefore, there exists a constant $C > 0$ such that

$$|\partial_{x_i}|_q (h \circ r_s)| \leq C \|h\|_{\mathcal{C}^1}$$

for any $s \in [0, \theta]$ and any q in the domain of definition of the local frame $(\partial_{x_i})_{i=1,2,3}$. The conclusion is thus achieved by setting φ to be constantly equal to $C \|h\|_{\mathcal{C}^1}$. \square

As $(\phi_t^X)_{t \in \mathbb{R}}$ is a smooth flow on M and f is of class \mathcal{C}^2 , the functions $f \circ \phi_1^X$, $X f \circ \phi_1^X$ are of class \mathcal{C}^1 . By virtue of Lemma 4.2, the functions

$$p \mapsto \int_0^{\theta} f \circ \phi_1^X \circ r_s(p) \, ds, \quad p \mapsto \int_0^{\theta} X f \circ \phi_1^X \circ r_s(p) \, ds$$

are of class \mathcal{C}^1 on M .

²⁸Formally, we would need to specify a Riemannian metric on the compact manifold M . For the purposes of the proof however, only boundedness of the relevant quantities matters, so that any such metric would serve our goal (cf. Remark 1.11(1)).

Therefore, it remains to deal with $\int_1^\infty g(\xi)G_{\theta,n}f(p,\xi) d\xi$ as a function of $p \in M$, g being as in (4.19). Expanding out the expression from (3.7), we obtain that it equals

$$\begin{aligned} & \int_1^\infty g(\xi) \left(\frac{n^2 e^{-\xi}}{\theta(1-e^{-2\xi})^2} \int_0^\theta f \circ \phi_\xi^X \circ r_s(p) ds - \frac{2e^{-\xi}}{\theta(1-e^{-2\xi})} \int_0^\theta Xf \circ \phi_\xi^X \circ r_s(p) ds \right. \\ & \quad \left. + \frac{2ine^{-2\xi}}{\theta(1-e^{-2\xi})^2} (f \circ \phi_\xi^X \circ r_\theta(p) - f \circ \phi_\xi^X(p)) + \frac{2}{\theta(1-e^{-2\xi})} (Uf \circ \phi_\xi^X \circ r_\theta(p) - Uf \circ \phi_\xi^X(p)) \right) d\xi. \end{aligned}$$

We shall now treat the four summands separately.

Lemma 4.3. *If g takes one of the forms in (4.19), the functions*

$$p \mapsto \int_1^\infty \frac{g(\xi)e^{-\xi}}{(1-e^{-2\xi})^2} \left(\int_0^\theta f \circ \phi_\xi^X \circ r_s(p) ds \right) d\xi, \quad (4.21)$$

$$p \mapsto \int_1^\infty \frac{g(\xi)e^{-\xi}}{1-e^{-2\xi}} \left(\int_0^\theta Xf \circ \phi_\xi^X \circ r_s(p) ds \right) d\xi, \quad (4.22)$$

$$p \mapsto \int_1^\infty \frac{g(\xi)e^{-\xi}}{(1-e^{-2\xi})^2} \left(f \circ \phi_\xi^X \circ r_\theta(p) - f \circ \phi_\xi^X(p) \right) d\xi \quad (4.23)$$

are of class \mathcal{C}^1 on M .

Proof. The proof proceeds along the same lines as the proof of Lemma 4.2. The crucial point is that, for any point $q \in M$ and any $\xi \geq 1$, the operator norm of the differential $(d\phi_\xi^X)_q$ doesn't exceed (up to a constant factor depending only on the choice of a Riemannian metric on the tangent bundle TM) the quantity $e^{\xi/2}$, as direct computations allow to verify starting from the explicit expression of ϕ_ξ^X in (1.4). As a consequence, there exists a constant $C > 0$ such that

$$\left| \frac{g(\xi)e^{-\xi}}{(1-e^{-2\xi})^2} \left(\int_0^\theta \partial_{x_i}|_q (f \circ \phi_\xi^X \circ r_s) ds \right) \right| \leq C \frac{|g(\xi)|e^{-\xi}}{(1-e^{-2\xi})^2} \int_0^\theta \|f\|_{\mathcal{C}^1} e^{\frac{\xi}{2}} ds = C \|f\|_{\mathcal{C}^1} \frac{|g(\xi)|e^{-\frac{\xi}{2}}}{(1-e^{-2\xi})^2}$$

for every $\xi \geq 1$ and $i = 1, 2, 3$, where $(\partial_{x_i})_{i=1,2,3}$ is a local frame of TM around a given fixed point $p_0 \in M$. Since the function

$$\frac{|g(\xi)|e^{-\xi/2}}{(1-e^{-2\xi})^2}$$

is integrable on the half-line $[1, \infty)$, we deduce as in the proof of Lemma 4.2 that the function in (4.21) is of class \mathcal{C}^1 on M .

The same assertion for the remaining two functions in (4.22) and (4.23) follows by a similar argument. \square

What is left to investigate, up to multiplicative constants, is thus the regularity of the function

$$p \mapsto \int_1^\infty \frac{g(\xi)}{1-e^{-2\xi}} (Uf \circ \phi_\xi^X \circ r_\theta(p) - Uf \circ \phi_\xi^X(p)) d\xi \quad (4.24)$$

on the manifold M . As we shall presently see, it depends on the function g .

Lemma 4.4. *If $g(\xi) = e^{-\frac{1+\nu}{2}\xi}$, then the function in (4.24) is of class \mathcal{C}^1 on M .*

Proof. It suffices to argue as in the proof of Lemma 4.3 observing that, when $0 < \nu < 1$, the function

$$\frac{e^{-\frac{1+\nu}{2}\xi}}{1-e^{-2\xi}} e^{\frac{\xi}{2}} = \frac{e^{-\frac{\nu}{2}\xi}}{1-e^{-2\xi}}$$

is integrable on the half-line $[1, \infty)$. \square

It is straightforward to realize that the same argument does not carry over to the other possible forms of $g(\xi)$ listed in (4.19). For those remaining cases, we instead establish Hölder-continuity of the function in (4.24) by a different argument.

Fix a Riemannian metric g on the connected manifold M , inducing a Riemannian distance function d . The choice is immaterial for our purposes, as pointed out in Remark 1.11(1). We start with the following well-known properties of the flows $(\phi_t^X)_{t \in \mathbb{R}}$, $(r_s)_{s \in \mathbb{R}}$.

Lemma 4.5. *There exist real constants $C_{X,d}, C_{\Theta,d}$, depending only on d , such that, for any pair of points $p, q \in M$, it holds*

$$d(\phi_t^X(p), \phi_t^X(q)) \leq C_{X,d} e^{|t|} d(p, q) \quad (4.25)$$

for every $t \in \mathbb{R}$ and

$$d(r_s(p), r_s(q)) \leq C_{\Theta,d} d(p, q) \quad (4.26)$$

for every $s \in \mathbb{R}$.

Proof. By compactness of M , we have the freedom to prove the lemma for a judicious choice of d . To profit most from the algebraic description of the flows $(\phi_t)_{t \in \mathbb{R}}$ and $(r_s)_{s \in \mathbb{R}}$, we fix a left-invariant Riemannian metric $g_{\mathrm{SL}_2(\mathbb{R})}$ on the Lie group $\mathrm{SL}_2(\mathbb{R})$ and let d be the Riemannian distance function on M determined by the unique Riemannian metric g for which the projection map $(\mathrm{SL}_2(\mathbb{R}), g_{\mathrm{SL}_2(\mathbb{R})}) \rightarrow (M, g)$ is a Riemannian submersion (cf. [43, Cor. 2.29]) or, equivalently for a covering map, a local isometry. As M is compact, we can choose a finite open cover $\tilde{\mathcal{U}} = (\tilde{U}_i)_{i \in I}$ of M and a collection $\mathcal{U} = (U_i)_{i \in I}$ of open subsets of $\mathrm{SL}_2(\mathbb{R})$ such that, for any $i \in I$, the restriction of the projection to U_i is an isometry from U_i onto \tilde{U}_i .

The distance $d_{\mathrm{SL}_2(\mathbb{R})}$ induced by $g_{\mathrm{SL}_2(\mathbb{R})}$ is locally equivalent to the distance induced by the operator norm $\|\cdot\|_{\mathrm{op}}$ on the vector space of 2×2 real matrices corresponding to the Euclidean norm on \mathbb{R}^2 (cf. [23, Lem. 9.12]): for every $g \in \mathrm{SL}_2(\mathbb{R})$, there exists an open neighborhood W_g of g and a constant $C_{d,g}$ such that

$$C_{d,g}^{-1} \|x - y\|_{\mathrm{op}} \leq d_{\mathrm{SL}_2(\mathbb{R})}(x, y) \leq C_{d,g} \|x - y\|_{\mathrm{op}}$$

for any $x, y \in W_g$. Upon restricting the U_i 's (and the \tilde{U}_i 's) if necessary, we may assume that each \tilde{U}_i is contained in W_{g_i} for some $g_i \in \mathrm{SL}_2(\mathbb{R})$. Define C_d to be the supremum of the C_{d,g_i} , $i \in I$. We also select a second finite open cover $\mathcal{V} = (V_j)_{j \in J}$ in such a way that the closure of each V_j is compact and contained in some $U_{i(j)}$. Observe that, for every $j \in J$, the function $\tau_j: \overline{V_j} \rightarrow (0, \infty]$ defined as the first exit time

$$\tau_j(p) = \inf\{t > 0 : \phi_t^X(p) \notin U_{i(j)}\}, \quad p \in V_j$$

is continuous, and as such attains a strictly positive minimal value t_j . Set $t_0 = \inf_{j \in J} t_j$ and let $\delta_{\mathcal{V}}$ be a Lebesgue number for the covering \mathcal{V} ([50, Lem. 27.5]).

Consider now two points $p, q \in M$, and suppose first that $d(p, q) \geq \delta_{\mathcal{V}}$; then

$$d(\phi_t^X(p), \phi_t^X(q)) \leq \mathrm{diam}_d(M) \leq \delta_{\mathcal{V}}^{-1} \mathrm{diam}_d(M) d(p, q) \leq \delta_{\mathcal{V}}^{-1} \mathrm{diam}_d(M) e^{|t|} d(p, q) \quad (4.27)$$

for every $t \in \mathbb{R}$, where $\mathrm{diam}_d(M) = \sup_{p', q' \in M} d(p', q')$ is the diameter of M with respect to d .

Now assume that $d(p, q) < \delta_{\mathcal{V}}$, so that p and q both lie in some V_j . Choose representatives x and y of p and q , respectively, inside $\tilde{U}_{i(j)}$; then, for every $0 \leq t < t_0$, we have

$$\begin{aligned} d(\phi_t^X(p), \phi_t^X(q)) &\leq d_{\mathrm{SL}_2(\mathbb{R})}(x \exp tX, y \exp tX) \leq C_d \|x \exp tX - y \exp tX\|_{\mathrm{op}} \\ &\leq C_d \|x - y\|_{\mathrm{op}} \|\exp tX\|_{\mathrm{op}} \leq C_d^2 d_{\mathrm{SL}_2(\mathbb{R})}(x, y) e^{|t|/2} \\ &\leq C_d^2 e^t d_{\mathrm{SL}_2(\mathbb{R})}(x, y) = C_d^2 e^t d(p, q), \end{aligned} \quad (4.28)$$

where we used the fact that $p, q, \phi_t^X(p), \phi_t^X(q)$ all belong to $U_i^{(j)}$. If now $t_0 \leq t < 2t_0$, then we distinguish two cases.

- If $d(\phi_{t_0}^X(p), \phi_{t_0}^X(q)) \geq \delta_{\mathcal{V}}$, then (4.27) applies giving

$$d(\phi_t^X(p), \phi_t^X(q)) \leq \delta_{\mathcal{V}}^{-1} \mathrm{diam}_d(M) e^{t-t_0} d(\phi_{t_0}^X(p), \phi_{t_0}^X(q)) \leq C_d^2 \delta_{\mathcal{V}}^{-1} \mathrm{diam}_d(M) e^t d(p, q).$$

This estimate is actually valid for any $t \geq t_0$.

- If $d(\phi_{t_0}^X(p), \phi_{t_0}^X(q)) < \delta_{\mathcal{V}}$, then the computations in (4.28) are valid for the given t and yield

$$d(\phi_t^X(p), \phi_t^X(q)) \leq C_d^2 e^t d(p, q) .$$

Subdividing the half-line $\mathbb{R}_{\geq 0}$ into the intervals $[kt_0, (k+1)t_0)$, $k \in \mathbb{N}$, and arguing as above on each of them, we conclude that

$$d(\phi_t^X(p), \phi_t^X(q)) \leq \sup\{C_d^2, C_d^2 \delta_{\mathcal{V}}^{-1} \text{diam}_d(M)\} e^t d(p, q)$$

for any $t > 0$.

The same analysis can be performed, with the appropriate modifications, for times $t < 0$. This shows (4.25).

The inequality in (4.26) is taken care of in an entirely analogous fashion, observing that $\|\exp(s\Theta)\|_{\text{op}} = 1$ for every $s \in \mathbb{R}$. \square

Let us now fix two points $p, q \in M$. As in the previous proof, we denote by $\text{diam}_d(M)$ the diameter of M with respect to the distance d ; with

$$\text{Lip}_d(Uf) = \sup_{p', q' \in M, p' \neq q'} \frac{|Uf(p') - Uf(q')|}{d(p, q)}$$

we indicate the Lipschitz constant of the function Uf with respect to d .

We may then estimate

$$\begin{aligned} & \left| \int_1^\infty \frac{g(\xi)}{1 - e^{-2\xi}} (Uf \circ \phi_\xi^X \circ r_\theta(p) - Uf \circ \phi_\xi^X(p)) d\xi - \int_1^\infty \frac{g(\xi)}{1 - e^{-2\xi}} (Uf \circ \phi_\xi^X \circ r_\theta(q) - Uf \circ \phi_\xi^X(q)) d\xi \right| \\ & \leq \int_1^\infty \frac{|g(\xi)|}{1 - e^{-2\xi}} (|Uf \circ \phi_\xi^X \circ r_\theta(p) - Uf \circ \phi_\xi^X \circ r_\theta(q)| + |Uf \circ \phi_\xi^X(p) - Uf \circ \phi_\xi^X(q)|) d\xi \\ & \leq (1 - e^{-2})^{-1} \text{Lip}(Uf) \int_1^\infty |g(\xi)| (\min\{\text{diam}_d(M), C_{X,d} e^\xi d(r_\theta(p), r_\theta(q))\} + \\ & \quad + \inf\{\text{diam}_d(M), C_{X,d} e^\xi d(p, q)\}) d\xi \\ & \leq 2(1 - e^{-2})^{-1} \text{Lip}(Uf) \int_1^\infty |g(\xi)| \inf\{\text{diam}_d(M), C d(p, q) e^\xi\} d\xi , \end{aligned} \tag{4.29}$$

where $C = C_{X,d} \sup\{1, C_{\Theta,d}\}$.

The following elementary estimates allow to finalize the argument.

Lemma 4.6. *Let $r, K \in \mathbb{R}_{>0}$, $a \in (0, 1)$. Then*

$$\int_1^\infty e^{-a\xi} \inf\{K, r e^\xi\} d\xi \leq \frac{1}{a(1-a)K^{a-1}} r^a . \tag{4.30}$$

Furthermore,

$$\int_1^\infty \xi e^{-\frac{\xi}{2}} \inf\{K, r e^\xi\} d\xi = 4\sqrt{K} r^{\frac{1}{2}} (\log k + \log r^{-1}) . \tag{4.31}$$

Proof. It suffices to split the integral as

$$\begin{aligned} \int_1^\infty e^{-a\xi} \inf\{K, r e^\xi\} d\xi &= \int_1^{\log \frac{K}{r}} e^{-a\xi} \cdot r e^\xi d\xi + \int_{\log \frac{K}{r}}^\infty e^{-a\xi} \cdot K d\xi \\ &= \frac{r}{1-a} \left(\left(\frac{r}{K} \right)^{a-1} - e^{1-a} \right) + \frac{K}{a} \left(\frac{r}{K} \right)^a \\ &\leq \frac{r^a}{(1-a)K^{a-1}} + \frac{r^a}{aK^{a-1}} = \frac{1}{a(1-a)K^{a-1}} r^a , \end{aligned}$$

which delivers the inequality (4.30). Analogous computations allow to establish (4.31). \square

Combining Lemmata 4.3, 4.4, 4.6 together with the estimate in (4.29) and the explicit expressions for the coefficients $D_{\theta,\mu,n}^\pm f$ in (4.8), (4.9), (4.12), (4.14), we deduce the following:

- when $\mu > 1/4$, $D_{\theta,\mu,n}^\pm f$ are Hölder continuous with Hölder exponent $1/2$;
- when $\mu = 1/4$, $D_{\theta,1/4,n}^+ f$ and $D_{\theta,1/4,n}^-$ are Hölder continuous, the latter with Hölder exponent $1/2$, while the former with Hölder exponent $1/2 - \varepsilon$ for every $\varepsilon > 0$;
- when $0 < \mu < 1/4$, $D_{\theta,\mu,n}^+ f$ is Hölder continuous with Hölder exponent $\frac{1-\nu}{2}$, while $D_{\theta,\mu,n}^- f$ is of class \mathcal{C}^1 on M .

The proof of Theorem 1.6 is concluded.

5. ASYMPTOTICS FOR ARBITRARY FUNCTIONS

The bulk of this section is devoted to the deduction of Theorem 1.8, which addresses the equidistribution rate of sufficiently regular observables on M not subject to any eigenfunction condition, from the special case of joint eigenfunctions of \square and Θ phrased in Theorem 1.6. The argument is crucially based upon the orthogonal decompositions of Sobolev spaces into joint eigenspaces of \square and Θ , which is recalled in detail in Section 2.2. We then proceed by proving Theorem 1.21, concerning the asymptotic behaviour of arbitrary translates of compact orbits inside M ; in light of the classical Cartan decomposition of the Lie group $\mathrm{SL}_2(\mathbb{R})$, the result follows from Theorem 1.8 in a fairly straightforward manner. Along the way, we shall also clarify the steps needed to derive, from those two main results, Corollaries 1.9 and 1.14.

5.1. Sum estimates on Sobolev norms of eigenfunctions. Before embarking on the proof of Theorem 1.8, we collect in this subsection a few estimates relating sums of norms of Sobolev eigenfunctions with a higher-order Sobolev norm of their sum, which will prove to be instrumental in the sequel. The need for those estimates stems from the fact that the Hilbert-sum decompositions in Section 2.2 only provide, by Bessel's inequality ([62, 2, XXIII, 6; 14]), estimates on the sum of squares of the components' norms, while our approach necessitates ℓ^1 -bounds (see Section 5.2).

Notation is as in Section 2.2.

Lemma 5.1. *Let k be a natural number.*

(1) *The infinite series*

$$\sum_{\mu \in \mathrm{Spec}(\square)} \sum_{n \in I(\mu)} \frac{1}{(1 + \mu + \frac{n^2}{2})^k} \quad (5.1)$$

is summable if and only if $k \geq 2$.

(2) *The infinite series*

$$\sum_{\mu \in \mathrm{Spec}(\square)} \sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^k} \quad (5.2)$$

is summable if and only if $k \geq 3$.

Remark 5.2. Observe that the series in (5.1) and (5.2) consist of nonnegative real numbers: see Lemma 2.3.

Proof of Lemma 5.1. It is convenient to examine separately the convergence properties of

$$\sum_{\mu \in \mathrm{Spec}(\square) \cap \mathbb{R}_{\geq 0}} \sum_{n \in I(\mu)} \frac{1}{(1 + \mu + \frac{n^2}{2})^k} \quad (5.3)$$

and of

$$\sum_{\mu \in \mathrm{Spec}(\square) \cap \mathbb{R}_{< 0}} \sum_{n \in I(\mu)} \frac{1}{(1 + \mu + \frac{n^2}{2})^k}. \quad (5.4)$$

We know (see Section 2.2) that negative eigenvalues of the Casimir operator are of the form $\mu_m = -m(m+2)/4$ for m ranging over the set \mathbb{N}^* of positive natural numbers; therefore

$$\sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{<0}} \sum_{n \in I(\mu)} \frac{1}{(1 + \mu + \frac{n^2}{2})^k} = \sum_{m \in \mathbb{N}^*} \sum_{n \in I(\mu_m)} \frac{1}{(1 - \frac{m(m+2)}{4} + \frac{n^2}{2})^k}, \quad (5.5)$$

which has the same convergence properties of the series

$$\sum_{(m,n) \in \mathbb{Z}^2} \frac{1}{(1 + m^2 + n^2)^k}.$$

By comparison with the integral

$$\int_{\mathbb{R}^2} \frac{1}{(1 + x^2 + y^2)^k} dx dy,$$

which is convergent if and only if $k \geq 2$, as is well-known, we infer that:

- (1) the series in (5.1) cannot converge if $k = 1$;
- (2) the series in (5.5) converges for any $k \geq 2$.

As to the first summation in (5.3), we may now suppose $k \geq 2$ and appeal to the Weyl law for the positive eigenvalues of the Casimir operator (see Theorem 2.1), which we list in increasing order as $\mu_0^{(p)} = 0 < \mu_1^{(p)} < \dots < \mu_m^{(p)} < \dots$, without multiplicity. Recall that $\text{area}(S)$ is the volume of the surface $S = \Gamma \backslash \mathbb{H}$ with respect to the hyperbolic area measure. Choose a real number $c > \text{area}(S)/4\pi$; then, there exists $R_0 \in \mathbb{R}_{>0}$ such that $\mu_m^{(p)} > m/c$ for any integer $m > cR_0$. On the one hand, the quantity

$$\sum_{0 \leq m \leq cR_0} \sum_{n \in I(\mu_m^{(p)})} \frac{1}{(1 + \mu_m^{(p)} + \frac{n^2}{2})^k}$$

is a finite sum of converging series. On the other hand,

$$\begin{aligned} \sum_{m > cR_0} \sum_{n \in I(\mu_m^{(p)})} \frac{1}{(1 + \mu_m^{(p)} + \frac{n^2}{2})^k} &< \sum_{m > cR_0} \sum_{n \in I(\mu_m^{(p)})} \frac{1}{(1 + \frac{m}{c} + \frac{n^2}{2})^k} \\ &\leq \sum_{m > cR_0} \frac{1}{(1 + \frac{m}{c})^k} + \sum_{m > cR_0} \sum_{n \in I(\mu_m^{(p)}) \setminus \{0\}} \frac{1}{(1 + \frac{m}{c} + \frac{n^2}{2})^k}. \end{aligned}$$

Plainly, the first summand in the last expression converges whenever $k \geq 2$, and so does the second by comparison with the integral

$$\begin{aligned} \int_{cR_0}^{\infty} \int_0^{\infty} \frac{1}{(1 + \frac{x}{c} + \frac{y^2}{2})^k} dy dx &= \int_{cR_0}^{\infty} \frac{1}{(1 + \frac{x}{c})^k} \int_0^{\infty} \frac{\sqrt{2}(1 + \frac{x}{c})^{1/2}}{(1 + u^2)^k} du dx \\ &= \left(\int_{cR_0}^{\infty} \frac{\sqrt{2}}{(1 + \frac{x}{c})^{k-1/2}} dx \right) \left(\int_0^{\infty} \frac{1}{(1 + u^2)^k} du \right) < \infty. \end{aligned}$$

As far as the series in (5.2) is concerned, its summability properties follows readily by running the argument above with the appropriate modifications. \square

Leveraging the estimates in Lemma 5.1, we are now in a position to prove the following.

Proposition 5.3. *Let $k \geq 2$ be an integer. There exists a constant $C_{\text{Spec},k} > 0$, depending only on k and on the spectrum of the Laplace-Beltrami operator on the hyperbolic surface S , such that the following holds. Let s be a positive real number, f a function in the Sobolev space $W^{s+k}(H)$; for any $\mu \in \text{Spec}(\square)$ and $n \in I(\mu)$, let $f_{\mu,n}$ be the orthogonal projection of f , with respect to the inner product in $W^{s+k}(H)$, onto the closed subspace $W^{s+k}(H_{\mu,n})$. Then*

$$\sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^s} \leq C_{\text{Spec},k} \|f\|_{W^{s+k}}. \quad (5.6)$$

Proof. Recall from Lemma 2.3 that, for any $\mu \in \text{Spec}(\square)$ and $n \in I(\mu)$,

$$\|f_{\mu,n}\|_{W^s}^2 = \left(1 + \mu + \frac{n^2}{2}\right)^{-k} \|f_{\mu,n}\|_{W^{s+k}}^2 .$$

Using the Cauchy-Schwartz inequality, we get

$$\begin{aligned} \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^s} &= \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \left(1 + \mu + \frac{n^2}{2}\right)^{-k/2} \|f_{\mu,n}\|_{W^{s+k}} \\ &\leq \left(\sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \frac{1}{\left(1 + \mu + \frac{n^2}{2}\right)^k} \right)^{1/2} \left(\sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^{s+k}}^2 \right)^{1/2} . \end{aligned}$$

The inequality in (5.6) is thus a consequence of Parseval's identity ([62, 2, XXIII, 6; 17])

$$\|f\|_{W^{s+k}}^2 = \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^{s+k}}^2 ,$$

where we define the constant $C_{\text{Spec},k}$ as

$$C_{\text{Spec},k} = \left(\sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \frac{1}{\left(1 + \mu + \frac{n^2}{2}\right)^k} \right)^{1/2} ,$$

which is finite by Lemma 5.1 and satisfies the dependence properties claimed in the statement (cf. Sections 2.2, 2.3). \square

5.2. Equidistribution of expanding translates of circle arcs. We are now ready to prove Theorem 1.8. Let θ be a real parameter in the interval $(0, 4\pi]$, s a real number satisfying $s > 11/2$, f a function in the Sobolev space $W^s(M)$. Keeping with the notation introduced in the foregoing subsection, we denote by $f_{\mu,n} \in W^s(H_{\mu,n})$ the orthogonal projection of f onto the closed subspace $W^s(H_{\mu,n})$, for any Casimir eigenvalue μ and any $n \in I(\mu)$. In what follows, the equivalence classes f and $f_{\mu,n}$ are identified with their unique²⁹ continuous representatives. The asymptotic expansion in (1.17) will result from the sum of the contributions of each component $f_{\mu,n}$, which are provided by Theorem 1.6. We now expose the details.

Choose a real parameter s' satisfying $3/2 < s' \leq s - 2$; the Sobolev Embedding Theorem (Theorem 2.4) gives the bound $\|f_{\mu,n}\|_{\infty} \leq C_{0,s'} \|f_{\mu,n}\|_{W^{s'}}$ for any μ and n as before, for some constant $C_{0,s'} > 0$ depending only on s' and on the manifold M . For any point $q \in M$, we estimate

$$\begin{aligned} \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} |f_{\mu,n}(q)| &\leq \sum_{n \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{\infty} \leq C_{0,s'} \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^{s'}} \\ &\leq C_{0,s'} C_{\text{Spec},2} \|f\|_{W^{s'+2}} , \end{aligned} \tag{5.7}$$

the last inequality being given by Proposition 5.3. Select now a base point $p \in M$, which will remain fixed until the end of this subsection. By virtue of (5.7), the dominated convergence theorem for infinite series yields

$$\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds = \sum_{\mu \in \text{Spec}(\square)} \sum_{n \in I(\mu)} \frac{1}{\theta} \int_0^\theta f_{\mu,n} \circ \phi_t^X \circ r_s(p) \, ds . \tag{5.8}$$

Observe that, since $s > 11/2$, the components $f_{\mu,n}$ are of class \mathcal{C}^2 by the Sobolev Embedding Theorem (Theorem 2.4); to each summand on the right-hand side of (5.8), we may thus apply

²⁹Recall that uniqueness is a result of the fact that the uniform measure vol on M is fully supported.

Theorem 1.6, which delivers on a formal level the equality³⁰

$$\begin{aligned}
& \frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds = \int_M f \, d \text{vol} \\
& + \sum_{\mu \in \text{Spec}(\square), \mu \geq 1/4} e^{-\frac{t}{2}} \left(\cos \left(\frac{\Im \nu}{2} t \right) \left(\sum_{n \in I(\mu)} D_{\theta, \mu, n}^+ f_{\mu, n}(p) \right) + \sin \left(\frac{\Im \nu}{2} t \right) \left(\sum_{n \in I(\mu)} D_{\theta, \mu, n}^- f_{\mu, n}(p) \right) \right) \\
& + \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2} t} \left(\sum_{n \in I(\mu)} D_{\theta, \mu, n}^+ f_{\mu, n}(p) \right) + e^{-\frac{1-\nu}{2} t} \left(\sum_{n \in I(\mu)} D_{\theta, \mu, n}^- f_{\mu, n}(p) \right) \\
& + \varepsilon_0 \left(e^{-\frac{t}{2}} \left(\sum_{n \in I(1/4)} D_{\theta, 1/4, n}^+ f_{1/4, n}(p) \right) + t e^{-\frac{t}{2}} \left(\sum_{n \in I(1/4)} D_{\theta, 1/4, n}^- f_{1/4, n}(p) \right) \right) \\
& + \mathcal{R}_{\theta, \text{pos}} f(p, t) + e^{-t} \mathcal{M}_{\theta, 0} f(p, t) + \sum_{n \in I(0)} \mathcal{R}_{\theta, 0, n} f_{0, n}(p, t) + \mathcal{R}_{\theta, d} f(p, t)
\end{aligned} \tag{5.9}$$

for every $t \geq 1$, where ε_0 is defined in (1.15), $G_{\theta, n} f_{0, n}$ is as in (3.7) and the quantities

$$\mathcal{R}_{\theta, \text{pos}} f(p, t), \quad \mathcal{M}_{\theta, 0} f(p, t), \quad \mathcal{R}_{\theta, d} f(p, t), \quad p \in M, \quad t \geq 1$$

are defined as follows:

$$\mathcal{R}_{\theta, \text{pos}} f(p, t) = \sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}} \sum_{n \in I(\mu)} \mathcal{R}_{\theta, \mu, n} f(p, t), \tag{5.10}$$

$$\mathcal{M}_{\theta, 0} f(p, t) = \sum_{n \in I(0)} \int_1^t -G_{\theta, n} f_{0, n}(p, \xi) \, d\xi, \tag{5.11}$$

$$\mathcal{R}_{\theta, d} f(p, t) = \sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{<0}} \sum_{n \in I(\mu)} \frac{1}{\theta} \int_0^\theta f_{\mu, n} \circ \phi_t^X \circ r_s(p) \, ds. \tag{5.12}$$

The equality in (1.17) would follow directly from (5.9) by defining

$$D_{\theta, \mu}^\pm f(p) = \sum_{n \in I(\mu)} D_{\theta, \mu, n}^\pm f_{\mu, n}(p), \quad p \in M, \quad \mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0} \tag{5.13}$$

and

$$\mathcal{R}_\theta f(p, t) = \mathcal{R}_{\theta, \text{pos}} f(p, t) + e^{-t} \mathcal{M}_{\theta, 0} f(p, t) + \sum_{n \in I(0)} \mathcal{R}_{\theta, 0, n} f_{0, n}(p, t) + \mathcal{R}_{\theta, d} f(p, t), \quad p \in M, \quad t \geq 1; \tag{5.14}$$

It is left to show that all the infinite sums we are considering with a formal meaning are actually convergent.

Let us begin by examining the sums in (5.13). Fix $\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}$; for any $p \in M$ and $n \in I(\mu)$, we have from Theorem 1.6 that

$$|D_{\theta, \mu, n}^\pm f_{\mu, n}(p)| \leq \|D_{\theta, \mu, n}^\pm f_{\mu, n}\|_\infty \leq \frac{\kappa(\mu)}{\theta} (n^2 + 1) \|f_{\mu, n}\|_{\mathcal{C}^1}.$$

Choose a real parameter s'' so that $5/2 < s'' \leq s - 3$; then the Sobolev Embedding Theorem (Theorem 2.4) allows to deduce

$$\|D_{\theta, \mu, n}^\pm f_{\mu, n}\|_\infty \leq C_{1, s''} \frac{\kappa(\mu)}{\theta} (n^2 + 1) \|f_{\mu, n}\|_{W^{s''}}. \tag{5.15}$$

³⁰Notice that $\int_M f_{\mu, n} \, d \text{vol} = 0$ for every Casimir eigenvalue $\mu \neq 0$ and every $n \in I(\mu)$, as $f_{\mu, n}$ is orthogonal to the joint eigenspace $H_{0,0}$ which contains the constant functions. For the same reason, $\int_M f_{0, n} \, d \text{vol} = 0$ for every $n \in I(0) \setminus \{0\}$. Therefore, dominated convergence gives $\int_M f_{0,0} \, d \text{vol} = \int_M f \, d \text{vol}$.

Now, in view of Lemma 2.3 and applying Cauchy-Schwartz's inequality and Parseval's identity ([62, 2, XXIII, 6; 17]), we get

$$\begin{aligned} \sum_{n \in I(\mu)} (n^2 + 1) \|f_{\mu,n}\|_{W^{s''}} &= \sum_{n \in I(\mu)} \frac{n^2 + 1}{(1 + \mu + \frac{n^2}{2})^{3/2}} \|f_{\mu,n}\|_{W^{s''+3}} \\ &\leq \left(\sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} \left(\sum_{n \in I(\mu)} \|f_{\mu,n}\|_{W^{s''+3}}^2 \right)^{1/2} \\ &= \left(\sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} \|f_{\mu}\|_{W^{s''+3}} , \end{aligned} \quad (5.16)$$

where f_{μ} is the orthogonal projection of f onto the closed subspace $W^s(H_{\mu})$ and the infinite sum in the last expression converges (see Lemma 5.1).

Setting

$$C_{\mu} = \left(\sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} , \quad (5.17)$$

our argument thus leads, combining (5.15) and (5.16), to the estimate

$$\sum_{n \in I(\mu)} \|D_{\theta,\mu,n}^{\pm} f_{\mu,n}\|_{\infty} \leq C_{1,s''} C_{\mu} \frac{\kappa(\mu)}{\theta} \|f_{\mu}\|_{W^{s''+3}} , \quad (5.18)$$

which implies that the sums $\sum_{n \in I(\mu)} D_{\theta,\mu,n}^{\pm} f_{\mu,n}$ converge normally in the Banach space $\mathcal{C}^0(M)$, hence absolutely and uniformly for all $p \in M$. In particular, the functions $D_{\theta,\mu}^{\pm} f$ are well-defined, continuous on M and fulfill the upper bound

$$\|D_{\theta,\mu}^{\pm} f\|_{\infty} \leq C_{1,s-3} C_{\mu} \frac{\kappa(\mu)}{\theta} \|f_{\mu}\|_{W^s} , \quad (5.19)$$

obtained by picking $s'' = s - 3$ in (5.18).

We now consider sums over all positive Casimir eigenvalues. We have

$$\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} |D_{\theta,\mu}^{\pm} f(p)| \leq \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \|D_{\theta,\mu}^{\pm} f\|_{\infty} \leq \frac{C_{1,s-3}}{\theta} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} C_{\mu} \kappa(\mu) \|f_{\mu}\|_{W^s} .$$

Observe that $\kappa(\mu)$ (whose value is specified after (4.11)) is uniformly bounded by a constant $C_{\text{Spec, pos}}$ depending only on the infimum of the set $\text{Spec}(\square) \cap (1/4, \infty)$; recalling the definition of C_{μ} in (5.17), we apply once again Cauchy-Schwartz's inequality and Parseval's identity to infer

$$\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \|D_{\theta,\mu}^{\pm} f\|_{\infty} \leq C_{\text{Spec, pos}} C_{1,s-3} \|f\|_{W^s} \left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} , \quad (5.20)$$

where the term between parentheses on the right-hand side is finite because of Lemma 5.1.

Since the spectrum of the Casimir operator is discrete (see Section 2.2), there are only finitely many distinct eigenvalues in the interval $(0, 1/4)$, so that the series

$$\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \|D_{\theta,\mu}^{\pm} f\|_{\infty}$$

involves only finitely many additional terms with respect to (5.20); each of those terms can be bounded with the help of (5.19). The claimed bound in (1.16) follows, by defining the constant C'_{Spec} appropriately in terms of $C_{\text{Spec, pos}}$ and of the C_{μ} for $0 < \mu < 1/4$.

In order to finalize the proof of Theorem 1.8, we address now the remainder terms defined in

(5.11), (5.12) and (5.14). (5.10),

We start with the term in (5.10) stemming from positive Casimir eigenvalues. Define μ_{princ} to be the infimum of $\text{Spec}(\square) \cap (1/4, \infty)$, and let ν_{princ} be the corresponding parameter in $i\mathbb{R}_{>0}$ fulfilling $1 - \nu_{\text{princ}}^2 = 4\mu_{\text{princ}}$. Using the bounds for the remainder terms $\mathcal{R}_{\theta, \mu, n} f_{\mu, n}$ corresponding to the single components $f_{\mu, n}$, provided by Theorem 1.6, we estimate

$$\begin{aligned}
& \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \sum_{n \in I(\mu)} |\mathcal{R}_{\theta, \mu, n} f_{\mu, n}(p, t)| \\
& \leq \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \frac{8\kappa_0}{\theta \Im \nu} e^{-t} \sum_{n \in I(\mu)} (n^2 + 1) \|f_{\mu, n}\|_{\mathcal{C}^1} \\
& \leq \frac{8\kappa_0 C_{1, s-3}}{\theta} e^{-t} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \frac{1}{\Im \nu} \sum_{n \in I(\mu)} (n^2 + 1) \|f_{\mu, n}\|_{W^{s-3}} \\
& \leq \frac{8\kappa_0 C_{1, s-3}}{\theta \Im \nu_{\text{princ}}} e^{-t} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \left(\sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} \left(\sum_{n \in I(\mu)} \|f_{\mu, n}\|_{W^s}^2 \right)^{1/2} \\
& \leq \frac{8\kappa_0 C_{1, s-3}}{\theta \Im \nu_{\text{princ}}} e^{-t} \left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \sum_{n \in I(\mu)} \frac{(n^2 + 1)^2}{(1 + \mu + \frac{n^2}{2})^3} \right)^{1/2} \left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \sum_{n \in I(\mu)} \|f_{\mu, n}\|_{W^s}^2 \right)^{1/2} \\
& \leq \frac{8\kappa_0 C_{1, s-3} C_{\text{Spec}, 3}}{\theta} \frac{1}{\Im \nu_{\text{princ}}} \|f\|_{W^s} e^{-t}
\end{aligned} \tag{5.21}$$

for any $t \geq 1$, applying in successive order the bound in (1.8), Theorem 2.4, the Cauchy-Schwartz's inequality (twice) and Bessel's inequality ([62, 2, XXIII, 6; 14]) to the Hilbert space $W^s(M)$. Similarly, the bounds in (1.9) and (1.11) yield, respectively,

$$\sum_{n \in I(1/4)} |\mathcal{R}_{\theta, 1/4, n} f_{1/4, n}(p, t)| \leq \frac{4\kappa_0 C_{1, s-3} C_{\text{Spec}, 3}}{\theta} \|f\|_{W^s} (t+1) e^{-t}$$

and

$$\sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} \sum_{n \in I(\mu)} |\mathcal{R}_{\theta, \mu, n} f_{\mu, n}(p, t)| \leq \frac{4\kappa_0 C_{1, s-3} C_{\text{Spec}, 3} C_{\text{comp}}}{\theta} \|f\|_{W^s} e^{-t}$$

for any $t \geq 1$, where we set

$$C_{\text{comp}} = \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} \frac{1}{\nu(1-\nu)(1+\nu)}.$$

Defining thus

$$C_{\text{pos}} = \frac{2}{\Im \nu_{\text{princ}}} + C_{\text{comp}} + 1 \tag{5.22}$$

and applying the triangular inequality for infinite sums, we get from (5.10) that

$$|\mathcal{R}_{\theta, \text{pos}} f(p, t)| \leq \frac{4\kappa_0 C_{1, s-3} C_{\text{Spec}, 3} C_{\text{pos}}}{\theta} \|f\|_{W^s} (t+1) e^{-t}, \quad t \geq 1. \tag{5.23}$$

An entirely analogous argument, using the bound in (1.13), shows that

$$\left| \sum_{n \in I(0)} \mathcal{R}_{\theta, 0, n} f_{0, n}(p, t) \right| \leq \frac{(8e\pi + \kappa_0) C_{1, s-3} C_{\text{Spec}, 3}}{\theta} \|f\|_{W^s} e^{-t}, \quad t \geq 1. \tag{5.24}$$

Define now

$$C_{\text{disc}} = \left(\inf_{\mu \in \text{Spec}(\square), \mu < 0} |\mu| \right)^{-1} + \frac{2e\pi}{\kappa_0} \sup_{\mu \in \text{Spec}(\square), \mu < 0} \frac{3 + \nu}{\nu}.$$

Then, the bound in (1.14) leads to

$$\begin{aligned} & \sum_{\mu \in \text{Spec}(\square), \mu < 0} \sum_{n \in I(\mu)} \frac{1}{\theta} \left| \int_0^\theta f_{\mu,n} \circ \phi_t^X \circ r_s(p) \, ds \right| \\ & \leq \sum_{\mu \in \text{Spec}(\square), \mu < 0} \frac{C_{1,s-3}}{\theta} \left(\frac{4\kappa_0}{(\nu-1)(\nu+1)} + \frac{2e\pi(3+\nu)}{\nu} \right) e^{-t} \sum_{n \in I(\mu)} (n^2+1) \|f_{\mu,n}\|_{W^{s-3}} \\ & \leq \frac{\kappa_0 C_{1,s-3} C_{\text{Spec},3} C_{\text{disc}}}{\theta} \|f\|_{W^s} e^{-t} \end{aligned}$$

for any $t \geq 1$, arguing as in (5.21). It follows at once from (5.12) that

$$|\mathcal{R}_{\theta,d}f(p,t)| \leq \frac{\kappa_0 C_{1,s-3} C_{\text{Spec},3} C_{\text{disc}}}{\theta} \|f\|_{W^s} e^{-t} \quad (5.25)$$

for any $t \geq 1$. We finally come to the estimation of the term $\mathcal{M}_{\theta,0}f(p,t)$, defined in (5.11). The inequality (3.8) gives

$$\begin{aligned} \sum_{n \in I(0)} \left| \int_1^t -G_{\theta,n} f_{0,n}(p,\xi) \, d\xi \right| & \leq \frac{\kappa_0}{\theta} \sum_{n \in I(0)} (n^2+1) \int_1^t \|f_{0,n}\|_{\mathcal{G}^1} \, d\xi \\ & = \frac{\kappa_0}{\theta} (t-1) \left(\sum_{n \in I(0)} (n^2+1) \|f_{0,n}\|_{\mathcal{G}^1} \right) \\ & \leq \frac{\kappa_0 C_{1,s-3}}{\theta} (t-1) \left(\sum_{n \in I(0)} (n^2+1) \|f_{0,n}\|_{W^s} \right), \end{aligned}$$

so that

$$|\mathcal{M}_{\theta,0}f(p,t)| \leq \frac{\kappa_0 C_{1,s-3} C_{\text{Spec},3}}{\theta} \|f\|_{W^s} (t-1) \quad (5.26)$$

for any $t \geq 1$. Recalling (5.14) and combining the estimates in (5.23), (5.24), (5.25) and (5.26) we conclude that

$$\begin{aligned} |\mathcal{R}_\theta f(p,t)| & \leq |\mathcal{R}_{\theta,\text{pos}}f(p,t)| + e^{-t} |\mathcal{M}_{\theta,0}f(p,t)| + \sum_{n \in I(0)} |\mathcal{R}_{0,\theta,n}f_{0,n}(p,t)| + |\mathcal{R}_{\theta,d}f(p,t)| \\ & \leq \frac{C_{1,s-3} C_{\text{Spec}}}{\theta} \|f\|_{W^s} (t+1) e^{-t}, \end{aligned}$$

for any $t \geq 1$, where we set

$$C_{\text{Spec}} = C_{\text{Spec},3} (8e\pi + \kappa_0(2 + C_{\text{pos}} + C_{\text{disc}})),$$

which ostensibly depends only on the spectrum of the Casimir operator.

The proof of Theorem 1.8 is complete.

5.3. Effective equidistribution and shrinking circle arcs. In this subsection we briefly comment on the proof of Corollaries 1.9 and 1.14. Recall that μ_* indicates the spectral gap of the surface $\Gamma \backslash \mathbb{H}$, defined in (1.19).

As to Corollary 1.9, which concerns the optimal effective equidistribution statement that can be drawn from the asymptotic expansion in Theorem 1.8, it suffices to define the function $D_\theta^{\text{main}}f: M \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{C}$ as follows:

- $D_\theta^{\text{main}}f = D_{\theta,\mu_*}^- f$ if $\mu_* \leq 1/4$;
- $D_\theta^{\text{main}}f = D_{\theta,\mu_*}^+ f + D_{\theta,\mu_*}^- f$ if $\mu_* > 1/4$.

The error-term estimate in (1.21) then follows directly from the asymptotics in (1.17).

Now suppose that we let the boundaries of the parametrization of the circle arcs depend on the time parameter t , so as to deal with a collection of time-varying subarcs

$$\gamma_t = \{\phi_t^X \circ r_s(p) : \theta_1(t) \leq s \leq \theta_2(t)\}$$

as in the statement of Corollary 1.14. If, as in the assumptions to the latter, we suppose that, for a given function $\eta: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ satisfying $\eta(t) \rightarrow \infty$ as $t \rightarrow \infty$, we have

$$\theta_2(t) - \theta_1(t) \geq \eta(t)e^{-\frac{1-\Re\nu_*}{2}t}$$

for every sufficiently large t , ν_* corresponding to the spectral gap μ_* , then we obtain

$$\begin{aligned} \frac{1}{\theta_2(t) - \theta_1(t)} \int_{\theta_1(t)}^{\theta_2(t)} f \circ \phi_t^X \circ r_s(p) \, ds &= \frac{1}{\theta_2(t) - \theta_1(t)} \int_0^{\theta_2(t) - \theta_1(t)} f \circ \phi_t^X \circ r_{s+\theta_1(t)}(p) \, ds \\ &= \int_M f \, d\text{vol} + D_{\theta_2(t) - \theta_1(t)}^{\text{main}} f(r_{\theta_1(t)}(p)) t^{\varepsilon_0} e^{-\frac{1-\Re\nu_*}{2}t} + o(e^{-\frac{1-\Re\nu_*}{2}t}) . \end{aligned} \quad (5.27)$$

The bound (1.20) results into

$$|D_{\theta_2(t) - \theta_1(t)}^{\text{main}} f(r_{\theta_1(t)}(p))| \leq C_{1,s-3} C'_{\text{Spec}} \|f\|_{W^s} \left(\frac{1}{\theta_2(t) - \theta_1(t)} \right) \leq C_{1,s-3} C'_{\text{Spec}} \|f\|_{W^s} \eta(t)^{-1} e^{\frac{1-\Re\nu_*}{2}t}$$

for any $t \geq t_0$. We deduce that the right-hand side of (5.27) is equal to

$$\int_M f \, d\text{vol} + o(t)$$

as t tends to infinity. An elementary application of the Stone-Weierstrass' theorem ([28, Thm. 4.51]) gives that smooth functions are dense in the space of continuous functions on the compact manifold M ; it follows that the convergence

$$\frac{1}{\theta_2(t) - \theta_1(t)} \int_{\theta_1(t)}^{\theta_2(t)} f \circ \phi_t^X \circ r_s(p) \, ds \xrightarrow{t \rightarrow \infty} \int_M f \, d\text{vol}$$

can be upgraded to hold for every $f \in \mathcal{C}^0(M)$, whereby the desired equidistribution follows.

5.4. Equidistribution of circle arcs on the underlying surface. This short subsection is occupied by a few comments concerning the statement of Theorem 1.12, which is nothing but a specialization of Theorem 1.8 to the case of observables defined on the underlying surface $S = \Gamma \backslash \mathbb{H}$, except for the lower regularity assumed on the test function f . First, we remark that $\text{SO}_2(\mathbb{R})$ -invariance of the coefficients $D_{4\pi, \mu}^\pm f$ and of the remainder term $\mathcal{R}_{4\pi} f$ follows at once from their definition (see (5.13) and (5.14)) and the fact that f is assumed to be $\text{SO}_2(\mathbb{R})$ -invariant. We are only left to show that we might take $s > 9/2$, less restrictively in comparison to an arbitrary f defined on M . The relevant observation here is that, for any $\text{SO}_2(\mathbb{R})$ -invariant function $f \in L^2(M)$, the components f_μ appearing in the decomposition³¹

$$f = \sum_{\mu \in \text{Spec}(\Delta_S)} f_\mu, \quad f_\mu \in H_\mu$$

inherit $\text{SO}_2(\mathbb{R})$ -invariance, that is, they satisfy $\Theta f_\mu = 0$. The estimate in (5.20) thus only requires $s > 9/2 = 11/2 - 1$, as the sum

$$\sum_{\mu \in \text{Spec}(\Delta_S)} \frac{1}{(1 + \mu)^k}$$

converges already for $k = 2$, and not only for $k = 3$ as is the case in (5.20).

³¹Recall from Section 2.2 that the Casimir operator \square acts as the Laplace-Beltrami operator Δ_S on $\text{SO}_2(\mathbb{R})$ -invariant functions.

5.5. Equidistribution of arbitrary translates. In light of the asymptotic expansion of averages along expanding circles given in Theorem 1.8, the generalization to the behaviour of arbitrary $\mathrm{SL}_2(\mathbb{R})$ -translates stated in Theorem 1.21 is a rather straightforward consequence of the classical Cartan decomposition for the semisimple Lie group $\mathrm{SL}_2(\mathbb{R})$, for which the reader is referred to [38, Chap. VI]. We present the details of the argument in this subsection.

Let $A = \{\exp tX : t \in \mathbb{R}\}$ be the subgroup of $\mathrm{SL}_2(\mathbb{R})$ consisting of diagonal matrices with positive entries (recall that X is defined as in (1.4)). The product map

$$\mathrm{SO}_2(\mathbb{R}) \times A \times \mathrm{SO}_2(\mathbb{R}) \rightarrow \mathrm{SL}_2(\mathbb{R}), \quad (k_1, a, k_2) \mapsto k_1 a k_2$$

is surjective. For any $g \in \mathrm{SL}_2(\mathbb{R})$, choose a decomposition $g = k_1(g)a(g)k_2(g)$, where $a(g)$ is the diagonal matrix having as entries the singular values of the matrix g , in decreasing order. In particular, if $t(g) \in \mathbb{R}_{\geq 0}$ is (uniquely) determined by the condition

$$a(g) = \begin{pmatrix} e^{t(g)/2} & 0 \\ 0 & e^{-t(g)/2} \end{pmatrix}, \quad (5.28)$$

then $t(g)$ and the operator norm $\|g\|_{\mathrm{op}}$ of g associated to the Euclidean norm on \mathbb{R}^2 are related by

$$\|g\|_{\mathrm{op}} = e^{t(g)/2}, \quad t(g) = 2 \log \|g\|_{\mathrm{op}}. \quad (5.29)$$

Fix now a real number $s > 11/2$ and a function f in the Sobolev space $W^s(M)$. Recall that, for any $p \in M$, we indicate with $m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ the unique $\mathrm{SO}_2(\mathbb{R})$ -invariant measure supported on the compact orbit $\mathrm{SO}_2(\mathbb{R}) \cdot p$; furthermore, $g_* \mathrm{SO}_2(\mathbb{R})$ denotes the push-forward of the latter measure under the right translation map $R_g(\Gamma g') = \Gamma g'g$ on M . For any $p \in M$ and $g \in \mathrm{SL}_2(\mathbb{R})$, we resort to the Cartan decomposition of g and write

$$\begin{aligned} \int_M f \, dg_* m_{\mathrm{SO}_2(\mathbb{R}) \cdot p} &= \int_M f \circ R_g \, dm_{\mathrm{SO}_2(\mathbb{R}) \cdot p} = \int_M f \circ R_{k_2(g)} \circ R_{a(g)} \circ R_{k_1(g)} \, dm_{\mathrm{SO}_2(\mathbb{R}) \cdot p} \\ &= \int_M f \circ R_{k_2(g)} \circ R_{a(g)} \, dm_{\mathrm{SO}_2(\mathbb{R}) \cdot p} = \frac{1}{4\pi} \int_0^{4\pi} (f \circ R_{k_2(g)}) \circ \phi_{t(g)}^X \circ r_s(p) \, ds, \end{aligned} \quad (5.30)$$

using the $R_{k_1(g)}$ -invariance of $m_{\mathrm{SO}_2(\mathbb{R}) \cdot p}$ and the fact that $R_{a(g)} = \phi_{t(g)}^X$ in view of (5.28).

We may now make use of the asymptotic expansion provided by Theorem 1.8 for the function $f \circ R_{k_2(g)}$, which lies in the same Sobolev space $W^s(M)$ of f since $R_{k_2(g)}$ is a smooth diffeomorphism of M . We thereby obtain, for a fixed base point $p \in M$,

$$\begin{aligned} &\frac{1}{4\pi} \int_0^{4\pi} (f \circ R_{k_2(g)}) \circ \phi_{t(g)}^X \circ r_s(p) \, ds = \int_M f \circ R_{k_2(g)} \, d \mathrm{vol} \\ &+ e^{-\frac{t(g)}{2}} \left(\sum_{\mu \in \mathrm{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} t(g)\right) D_{4\pi, \mu}^+(f \circ R_{k_2(g)})(p) + \sin\left(\frac{\Im \nu}{2} t(g)\right) D_{4\pi, \mu}^-(f \circ R_{k_2(g)})(p) \right) \\ &+ \sum_{\mu \in \mathrm{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2} t(g)} D_{4\pi, \mu}^+(f \circ R_{k_2(g)})(p) + e^{-\frac{1-\nu}{2} t(g)} D_{4\pi, \mu}^-(f \circ R_{k_2(g)})(p) \\ &+ \varepsilon_0 \left(e^{-\frac{t(g)}{2}} D_{4\pi, 1/4}^+(f \circ R_{k_2(g)})(p) + t(g) e^{-\frac{t(g)}{2}} D_{4\pi, 1/4}^-(f \circ R_{k_2(g)})(p) \right) + \mathcal{R}_{4\pi}(f \circ R_{k_2(g)})(p, t(g)) \end{aligned} \quad (5.31)$$

for any $g \in \mathrm{SL}_2(\mathbb{R})$ with $\|g\|_{\mathrm{op}} \geq \sqrt{e}$. Define now, for any Casimir eigenvalue $\mu \in \mathbb{R}_{>0}$, the functions $D_\mu^\pm: M \times \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ by

$$D_\mu^\pm f(p, g) = D_{4\pi, \mu}^\pm(f \circ R_{k_2(g)})(p), \quad p \in M, g \in \mathrm{SL}_2(\mathbb{R}), \quad (5.32)$$

and set also

$$\mathcal{R}f(p, g) = \mathcal{R}_{4\pi}(f \circ R_{k_2(g)})(p, t(g)), \quad p \in M, g \in \mathrm{SL}_2(\mathbb{R}).$$

Then, combining (5.30) with (5.31) and recalling (5.29) together with the fact that $\int_M f \circ R_{k_2} \, d \text{vol} = \int_M f \, d \text{vol}$, we deduce

$$\begin{aligned} \int_M f \, dg_* m_{\text{SO}_2(\mathbb{R}) \cdot p} &= \int_M f \, d \text{vol} \\ &+ \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \|g\|_{\text{op}}^{-1} (\cos(\Im \nu \log \|g\|_{\text{op}}) D_\mu^+ f(p, g) + \sin(\Im \nu \log \|g\|_{\text{op}}) D_\mu^- f(p, g)) \\ &+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} \|g\|_{\text{op}}^{-(1+\nu)} D_\mu^+ f(p, g) + \|g\|_{\text{op}}^{-(1-\nu)} D_\mu^- f(p, g) \\ &+ \varepsilon_0 (\|g\|_{\text{op}}^{-1} D_{1/4}^+ f(p, g) + 2 \|g\|_{\text{op}}^{-1} \log \|g\|_{\text{op}} D_{1/4}^- f(p, g)) + \mathcal{R}f(p, g) \end{aligned}$$

for any $g \in \text{SL}_2(\mathbb{R})$ with $\|g\|_{\text{op}} \geq \sqrt{e}$, which is precisely the asymptotic expansion appearing in the statement of Theorem 1.21.

As stated in Theorem 1.8, the functions $D_{4\pi, \mu}^\pm(f \circ R_{k_2(g)})$ are continuous on M for any fixed $\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}$ and $g \in \text{SL}_2(\mathbb{R})$; equivalently, by (5.32), $D_\mu^\pm f(\cdot, g)$ is continuous on M for any fixed $g \in \text{SL}_2(\mathbb{R})$.

Furthermore, we have that, for any $p \in M$, $g \in \text{SL}_2(\mathbb{R})$ and $\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}$,

$$|D_\mu^\pm f(p, g)| = |D_{4\pi, \mu}^\pm(f \circ R_{k_2(g)})(p)| \leq \|D_{4\pi, \mu}^\pm(f \circ R_{k_2(g)})\|_\infty \leq \frac{C_{1, s-3} C_{\mu \kappa}(\mu)}{4\pi} \|f \circ R_{k_2(g)}\|_{W^s},$$

where the last inequality is given by (1.16). It remains to observe that compactness of $\text{SO}_2(\mathbb{R})$ implies that there exists a constant $C_{s, \text{rot}} > 0$ such that

$$\|f \circ R_k\|_{W^s} \leq C_{s, \text{rot}} \|f\|_{W^s}$$

for any $k \in \text{SO}_2(\mathbb{R})$. The proof of this assertion runs along the same lines of the proof of Lemma 4.2, with the appropriate modifications. Therefore, we get

$$\sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}} \sup_{p \in M, g \in \text{SL}_2(\mathbb{R})} |D_\mu^\pm f(p, g)| \leq \frac{C_{1, s-3} C_{s, \text{rot}} C'_{\text{Spec}}}{4\pi} \|f\|_{W^s},$$

where C'_{Spec} is as in Theorem 1.8.

To conclude the proof of Theorem 1.21, it is left to take care of the remainder term $\mathcal{R}f$. We easily estimate, from the bound (1.18),

$$\begin{aligned} |\mathcal{R}f(p, g)| &= |\mathcal{R}_{4\pi}(f \circ R_{k_2(g)})(p, t(g))| \leq \frac{C_{\text{Spec}} C_{1, s-3}}{4\pi} \|f \circ R_{k_2(g)}\|_{W^s} (t(g) + 1) e^{-t(g)} \\ &\leq \frac{C_{\text{Spec}} C_{1, s-3} C_{s, \text{rot}}}{4\pi} \|f\|_{W^s} (2 \log \|g\|_{\text{op}} + 1) \|g\|_{\text{op}}^{-2} \end{aligned}$$

for any $p \in M$ and $g \in \text{SL}_2(\mathbb{R})$ with $\|g\|_{\text{op}} \geq \sqrt{e}$.

This achieves the proof of Theorem 1.21.

6. DISTRIBUTIONAL LIMIT THEOREMS FOR DEVIATIONS FROM THE AVERAGE

The purpose of this section is threefold, articulated in three subsections. First, we establish the quantitative distributional convergence claimed in Proposition 1.16, from which the qualitative statements in Theorem 1.15 follow directly; secondly, we prove absence of a central limit theorem under the circumstances of Theorem 1.18, and finally we explore further ways of examining the statistical behaviour of averages along circle arcs.

6.1. Quantitative distributional convergence. Let us fix the length parameter $\theta \in (0, 4\pi]$, and consider a real-valued function f lying in the Sobolev space $W^s(M)$ for some real $s > 11/2$. We are interested in the statistical behaviour of the deviations from the mean

$$d_f(T, p) = \frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d \text{vol} , \quad T \geq 1, p \in M$$

appropriately renormalized, as the time parameter T tends to infinity and when the base point p is sampled according to the uniform probability measure vol on M (which we abbreviate with $p \sim \text{vol}$). Define

$$\mu_f = \inf \{ \mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0} : D_{\theta, \mu}^- f \text{ does not vanish identically on } M \} .$$

As in the hypotheses of Proposition 1.16, we assume that μ_f is finite, that is, the set of Casimir eigenvalues over which the previous infimum is taken is non-empty. Let ν_f be the corresponding parameter, namely $\nu_f \in \mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ satisfies $1 - \nu_f^2 = 4\mu_f$.

In order to quantify the rate of distributional convergence of the random variables under consideration, we make use of the Lévy-Prokhorov metric d_{LP} on the set $\mathcal{P}(\mathbb{R})$ of Borel probability measures on \mathbb{R} . We recall that this is defined as

$$d_{LP}(\lambda, \rho) = \inf \{ \varepsilon > 0 : \lambda(Y) \leq \rho(Y_\varepsilon) + \varepsilon \text{ and } \rho(Y) \leq \lambda(Y_\varepsilon) + \varepsilon \text{ for every Borel set } Y \subset \mathbb{R} \}$$

for any $\lambda, \rho \in \mathcal{P}(\mathbb{R})$, where Y_ε denotes the open ε -neighborhood of a subset Y with respect to the Euclidean metric on \mathbb{R} . The distance d_{LP} induces the topology of weak convergence of probability measures on $\mathcal{P}(\mathbb{R})$, namely the coarsest topology for which the evaluation maps

$$\mathcal{P}(\mathbb{R}) \ni \lambda \mapsto \int_{\mathbb{R}} \varphi \, d\lambda \in \mathbb{R} , \quad \varphi: \mathbb{R} \rightarrow \mathbb{R} \text{ continuous and bounded}$$

are continuous.

In forthcoming estimates we shall make use of the following trivial upper bound for the Lévy-Prokhorov distance between the laws of two random variables defined on the same probability space and taking on nearby values almost surely.

Lemma 6.1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\varepsilon > 0$. Suppose $X, X': \Omega \rightarrow \mathbb{R}$ are random variables satisfying $|X(\omega) - X'(\omega)| < \varepsilon$ for \mathbf{P} -almost every $\omega \in \Omega$. If λ_X and $\lambda_{X'}$ denote the laws of X and X' , respectively, then $d_{LP}(\lambda_X, \lambda_{X'}) \leq \varepsilon$.*

Proof. Let $A \subset \mathbb{R}$ be a Borel subset. The event $\{X \in A\}$ is contained in the event $\{X' \in A_\varepsilon\}$, up to a \mathbf{P} -negligible subset, by the assumption on the distance between X and X' . Therefore,

$$\lambda_X(A) = \mathbf{P}(X \in A) \leq \mathbf{P}(X' \in A_\varepsilon) = \lambda_{X'}(A_\varepsilon) < \lambda_{X'}(A_\varepsilon) + \varepsilon ;$$

a similar inequality holds reversing the role of X and X' , whence $d_{LP}(\lambda_X, \lambda_{X'}) \leq \varepsilon$. \square

We now proceed with the proof of Proposition 1.16 by distinguishing the three different cases $0 < \mu_f < 1/4$, $\mu_f = 1/4$ and $\mu_f > 1/4$.

Suppose first $0 < \mu_f < 1/4$. We would then like to show that the random variables

$$e^{\frac{1-\nu_f}{2}T} d_f(T, p) , \quad p \sim \text{vol}$$

converge in distribution, as T tends to infinity, to the random variable $D_{\theta, \mu_f}^- f(p)$, $p \sim \text{vol}$. Observe that, because of the asymptotic expansion (1.17) and the assumption on μ_f , we have

$$\begin{aligned} e^{\frac{1-\nu_f}{2}T} d_f(T, p) - D_{\theta, \mu_f}^- f(p) &= \sum_{\mu \in \text{Spec}(\square), \mu_f < \mu < 1/4} e^{-\frac{\nu_f - \nu}{2}T} D_{\theta, \mu}^- f(p) \\ &+ e^{-\frac{\nu_f}{2}T} \left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^- f(p) \right) \\ &+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{\nu_f + \nu}{2}T} D_{\theta, \mu}^+ f(p) \\ &+ \varepsilon_0 \left(e^{-\frac{\nu_f}{2}T} D_{\theta, 1/4}^+ f(p) + T e^{-\frac{\nu_f}{2}T} D_{\theta, 1/4}^- f(p) \right) + e^{\frac{1-\nu_f}{2}T} \mathcal{R}_\theta f(p, T), \end{aligned}$$

so that, owing to the uniform bound (1.16), we may estimate

$$\left| e^{\frac{1-\nu_f}{2}T} d_f(T, p) - D_{\theta, \mu_f}^- f(p) \right| \leq \frac{C_{1,s-3} C'_{\text{Spec}}}{\theta} \|f\|_{W^s} T e^{-\frac{\nu_f - \Re \nu_f^{\text{next}}}{2}T} \quad (6.1)$$

for any $p \in M$ and $T \geq 1$, where ν_f^{next} is the parameter corresponding to the smallest eigenvalue μ_f^{next} of the Casimir operator exceeding³² μ_f .

By Lemma 6.1, and recalling the definitions of $\mathbf{P}_{\theta, f}^{\text{circ}}(T)$ and $\mathbf{P}_{\theta, f}$ introduced in Section 1.4, we get

$$d_{LP}(\mathbf{P}_{\theta, f}^{\text{circ}}(T), \mathbf{P}_{\theta, f}) \leq \frac{C_{1,s-3} C'_{\text{Spec}}}{\theta} \|f\|_{W^s} T e^{-\eta_f T}$$

for every $T \geq 1$, with $\eta_f = \frac{\nu_f - \Re \nu_f^{\text{next}}}{2}$.

Similarly, if $\mu_f = 1/4$, we readily obtain from the expansion (1.17) that

$$\begin{aligned} T^{-1} e^{\frac{T}{2}} d_f(T, p) - D_{\theta, 1/4}^- f(p) &= T^{-1} \left(D_{\theta, 1/4}^+ f(p) + \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \left(\cos\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^+ f(p) \right. \right. \\ &\left. \left. + \sin\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^- f(p) \right) + \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{\nu}{2}T} D_{\theta, \mu}^+ f(p) + e^{\frac{T}{2}} \mathcal{R}_\theta f(p, t) \right) \end{aligned}$$

for any $p \in M$ and $T \geq 1$; recalling the definition of the constant C_{pos} in (5.22), we deduce the bound

$$\left| T^{-1} e^{\frac{T}{2}} d_f(T, p) - D_{\theta, 1/4}^- f(p) \right| \leq \frac{C_{1,s-3} C_{\text{pos}}}{\theta} \|f\|_{W^s} T^{-1},$$

so that, again by Lemma 6.1,

$$d_{LP}(\mathbf{P}_{\theta, f}^{\text{circ}}(T), \mathbf{P}_{\theta, f}) \leq \frac{C_{1,s-3} C_{\text{pos}}}{\theta} \|f\|_{W^s} T^{-1}$$

for every $T \geq 1$, as desired.

Finally, for $\mu_f > 1/4$, we have from the asymptotics (1.17) that

$$\begin{aligned} e^{\frac{T}{2}} d_f(T, p) - \varepsilon_0 D_{\theta, 1/4}^+ f(p) - \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2}T\right) D_{\theta, \mu}^- f(p) \\ = \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{\nu}{2}T} D_{\theta, \mu}^+ f(p) + e^{\frac{T}{2}} \mathcal{R}_\theta f(p, T) \end{aligned} \quad (6.2)$$

for any $p \in M$ and $T \geq 1$, from which we deduce what follows. Let μ_* be the spectral gap of $S = \Gamma \backslash \mathbb{H}$, defined in (1.19), and ν_* the corresponding parameter.

³²Observe that we may dispense with the additional factor T in the upper bound (6.1) whenever $\nu_f^{\text{next}} \in \mathbb{R}$.

- If $\mu_* < 1/4$, then (6.2) and (1.16) give

$$\begin{aligned} & \left| e^{\frac{T}{2}} d_f(T, p) - \varepsilon_0 D_{\theta, 1/4}^+ f(p) - \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^- f(p) \right| \\ & \leq \frac{C_{1, s-3} C'_{\text{Spec}}}{\theta} \|f\|_{W^s} e^{-\frac{\nu_*}{2} T}, \end{aligned}$$

whence

$$d_{LP}(\mathbf{P}_{\theta, f}^{\text{circ}}(T), \mathbf{P}_{\theta, f}(T)) \leq \frac{C_{1, s-3} C'_{\text{Spec}}}{\theta} \|f\|_{W^s} e^{-\frac{\nu_*}{2} T}$$

for every $T \geq 1$.

- If $\mu_* \geq 1/4$, then (6.2) and (1.18) give

$$\begin{aligned} & \left| e^{\frac{T}{2}} d_f(T, p) - \varepsilon_0 D_{\theta, 1/4}^+ f(p) - \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \cos\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^+ f(p) + \sin\left(\frac{\Im \nu}{2} T\right) D_{\theta, \mu}^- f(p) \right| \\ & \leq \frac{C_{1, s-3} C_{\text{Spec}}}{\theta} \|f\|_{W^s} (T+1) e^{-\frac{T}{2}}; \end{aligned}$$

we deduce that

$$d_{LP}(\mathbf{P}_{\theta, f}^{\text{circ}}(T), \mathbf{P}_{\theta, f}(T)) \leq \frac{C_{1, s-3} C_{\text{Spec}}}{\theta} \|f\|_{W^s} (T+1) e^{-\frac{T}{2}}$$

for every $T \geq 1$.

This completes the proof of Proposition 1.16.

6.2. Failure of a distributional limit theorem. We now turn to the proof of Theorem 1.18. Once again, we consider a fixed length parameter $\theta \in (0, 4\pi]$ and a function $f \in W^s(M)$ for some real $s > 11/2$. This time, we suppose that the coefficients $D_{\theta, \mu}^\pm f$ vanish identically on M for any Casimir eigenvalue $\mu > 0$. As a result, the asymptotic expansion provided in Theorem 1.8 reduces to

$$\begin{aligned} \frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds &= \int_M f \, d\text{vol} + e^{-T} \int_1^T \sum_{n \in I(0)} -G_{\theta, n} f_{0, n}(p, \xi) \, d\xi \\ &+ \sum_{n \in I(0)} \mathcal{R}_{\theta, 0, n} f_{0, n}(p, T) + \mathcal{R}_{\theta, d} f(p, T) \end{aligned} \quad (6.3)$$

for any $p \in M$ and $T \geq 1$, where $\mathcal{R}_{\theta, d}$ is defined in (5.12).

The estimates carried out in Section 5.2 lead to the bound

$$\left| \sum_{n \in I(0)} \mathcal{R}_{\theta, 0, n} f_{0, n}(p, T) + \mathcal{R}_{\theta, d} f(p, T) \right| \leq \frac{(8e\pi + \kappa_0) C_{1, s-3} C_{\text{Spec}, 3} \sup\{1, C_{\text{disc}}\}}{\theta} \|f\|_{W^s} e^{-T}. \quad (6.4)$$

On the other hand, by means of the explicit expression for $G_{\theta, n} f_{0, n}$ given in (3.7), we expand

$$\begin{aligned} \sum_{n \in I(0)} -G_{\theta, n} f_{0, n}(p, \xi) &= \frac{2}{\theta(1 - e^{-2\xi})} \sum_{n \in I(0)} (U f_{0, n} \circ \phi_\xi^X(p) - U f_{0, n} \circ \phi_\xi^X \circ r_\theta(p)) \\ &+ \frac{e^{-2\xi}}{\theta(1 - e^{-2\xi})^2} \sum_{n \in I(0)} 2in(f_{0, n} \circ \phi_\xi^X(p) - f_{0, n} \circ \phi_\xi^X \circ r_\theta(p)) \\ &- \frac{e^{-\xi}}{\theta(1 - e^{-2\xi})^2} \sum_{n \in I(0)} \int_0^\theta f_{0, n} \circ \phi_\xi^X \circ r_s(p) \, ds \\ &+ \frac{2e^{-\xi}}{\theta(1 - e^{-2\xi})} \sum_{n \in I(0)} \int_0^\theta X f_{0, n} \circ \phi_\xi^X \circ r_s(p) \, ds, \end{aligned}$$

from which

$$\sum_{n \in I(0)} -G_{\theta,n} f_{0,n}(p, \xi) = \frac{2}{\theta(1 - e^{-2\xi})} (Uf_0 \circ \phi_\xi^X(p) - Uf_0 \circ \phi_\xi^X \circ r_\theta(p)) + \mathcal{R}_G f(p, \xi),$$

where

$$\begin{aligned} |\mathcal{R}_G f(p, \xi)| &\leq \frac{1}{(1 - e^{-2})^2} \left(\|f\|_{W^s} e^{-2\xi} + \|f\|_\infty e^{-\xi} + 2\|f\|_{\mathcal{C}^1} e^{-\xi} (1 - e^{-2\xi}) \right) \\ &\leq \frac{1 + C_{0,s} + 2C_{1,s}}{(1 - e^{-2})^2} \|f\|_{W^s} e^{-\xi}, \end{aligned} \quad (6.5)$$

using the bound $1 - e^{-2\xi} \geq 1 - e^{-2}$ valid for any $\xi \geq 1$.

Let now $(B_T)_{T>0}$ be a collection of positive real numbers such that $B_T \rightarrow \infty$ as $T \rightarrow \infty$. In light of (6.3), (6.4) and (6.5), and because of the assumption on $(B_T)_{T>0}$, the distributional limits of the random variables

$$\frac{e^T \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds - \int_M f \, d\text{vol} \right)}{B_T}, \quad p \sim \text{vol}$$

as T tends to infinity coincide with the distributional limits of the random variables

$$\frac{\frac{2}{\theta} \int_1^T \frac{1}{1 - e^{-2\xi}} (Uf_0 \circ \phi_\xi^X(p) - Uf_0 \circ \phi_\xi^X \circ r_\theta(p)) \, d\xi}{B_T}, \quad p \sim \text{vol}. \quad (6.6)$$

When $\theta = 4\pi$, we have $r_{4\pi}(p) = p$, so that the integrand in the numerator of the above expression vanishes. Therefore, the distributional limit we are seeking after equals to zero almost surely, which proves Theorem 1.18.

As to Remark 1.19, suppose now $\theta \in (0, 4\pi]$ is arbitrary, and that Uf_0 is a coboundary for $(\phi_t^X)_{t \in \mathbb{R}}$: there exists a measurable function $g: M \rightarrow \mathbb{C}$ with³³ finite norm

$$\|g\|_{L^\infty(M, \text{vol})} = \inf \{ \lambda \in \mathbb{R}_{>0} : |g(p)| \leq \lambda \text{ for vol-almost every } p \in M \}$$

such that, for all $T > 0$, it holds

$$\int_0^T Uf_0 \circ \phi_\xi^X(p) \, d\xi = g \circ \phi_T^X(p) - g(p)$$

for vol-almost every $p \in M$.

It follows trivially that, for every $T > 0$,

$$\left| \int_1^T \frac{1}{1 - e^{-2\xi}} (Uf_0 \circ \phi_\xi^X(p) - Uf_0 \circ \phi_\xi^X \circ r_\theta(p)) \, d\xi \right| \leq \frac{4}{1 - e^{-2}} \|g\|_{L^\infty(M, \text{vol})}$$

for vol-almost every $p \in M$. As a result, the distributional limit as $T \rightarrow \infty$ of the random variables in (6.6) vanishes almost surely, since $B_T \rightarrow \infty$.

Remark 6.2. Slightly more generally, when Uf_0 is cohomologous to a constant function, namely it differs from a constant function by a coboundary, any distributional limit of the random variables in (6.6) is almost surely constant.

Assume now Uf_0 is not cohomologous to a constant function (and $\theta \neq 4\pi$). The classical central limit theorem for geodesic ergodic integrals (see [69] for the constant curvature case, and [57] for variable negative curvature) gives that both

$$\frac{\int_1^T Uf_0 \circ \phi_\xi^X(p) \, d\xi - \int_M Uf_0 \, d\text{vol}}{\sqrt{T}}, \quad p \sim \text{vol}$$

³³More accurately, this is the notion of a measurable coboundary; by the celebrated work of Livsic [45] on the cohomological equation for Anosov flows, the condition is actually equivalent to the seemingly more restrictive one of Uf_0 being a continuous coboundary, namely of requiring the transfer function g to be continuous.

and

$$\frac{\int_1^T U f_0 \circ \phi_\xi^X \circ r_\theta(p) \, d\xi - \int_M U f_0 \, d \text{vol}}{\sqrt{T}}, \quad p \sim \text{vol}$$

converge in distribution to a non-trivial centered Gaussian random variable as T tends to infinity. *A priori*, the combination of these two distributional convergences doesn't provide any information on the distributional limits of the difference, which is what appears in (6.6) up to the constant factor $2/\theta$; it would be desirable to reach a full understanding of this limiting distributional behaviour by carefully inspecting the dependence properties of the random variables $\int_1^T U f_0 \circ \phi_\xi^X(p) \, d\xi$ and $\int_1^T U f_0 \circ \phi_\xi^X \circ r_\theta(p) \, d\xi$ as p is sampled according to the uniform measure on M .

6.3. Some reflections on temporal distributional limit theorems. An upshot of the two foregoing subsections is the following consideration: examining the statistical behaviour, for large times T , of the (appropriately renormalized) averages

$$\frac{1}{\theta} \int_0^\theta f \circ \phi_T^X \circ r_s(p) \, ds$$

by randomly sampling the base point p according to the uniform measure on M leads to meaningful asymptotic results if and only if³⁴ at least one of the coefficients $D_{\theta,\mu}^\pm f$ does not vanish identically on M . Irrespective of whether this is the case or not, it is natural to look for different sources of randomness, which might capture oscillatory behaviours more accurately. In accordance with the perspective of temporal distributional limit theorems, pioneered by Dolgopyat and Sarig [17] in the context of ergodic sums and integrals, we enquire about the existence of non-trivial distributional limits for the random variables

$$\frac{e^t \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds \right) - A_T}{B_T},$$

where p is a fixed base point in M , $(A_T)_{T>0}$ and $(B_T)_{T>0}$ are collections of real numbers, possibly depending on p , with $B_T > 0$ and $B_T \rightarrow \infty$ as $T \rightarrow \infty$, and the time t is chosen uniformly at random in the interval $[0, T]$.

Remark 6.3. It is informative to compare this to the quest for temporal limit theorems for ergodic integrals along the orbits of a flow: see, in particular, [17, Def. 1.3]. Observe notably that our rescaling of the circle-arc average $\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds$ by a factor of e^t (the latter being asymptotically of the same order of the length of the expanding circle arc along which the average is taken) parallels the renormalization of ergodic averages by the linear factor t .

Let us denote by $\mathcal{U}_{[0,T]}$ the uniform probability measure on the compact interval $[0, T]$, for any $T > 0$. If there is a non-identically vanishing coefficient $D_{\theta,\mu}^\pm f$ for some Casimir eigenvalue $\mu > 0$, then a rather straightforward adaptation of the proof of [17, Cor. 5.7] shows that, for vol-almost every $p \in M$, any limiting distribution of

$$\frac{e^t \left(\frac{1}{\theta} \int_0^\theta f \circ \phi_t^X \circ r_s(p) \, ds \right) - A_T}{B_T}, \quad t \sim \mathcal{U}_{[0,T]} \tag{6.7}$$

is necessarily constant almost surely, no matter the choice of the constants A_T and B_T .

Suppose now that the coefficients $D_{\theta,\mu}^\pm f$ vanish identically on M for any positive Casimir eigenvalue μ . The deduction in Section 6.2 applies almost verbatim, showing that the distributional limits of the random variables in (6.7) are the same as the limits of

$$\frac{\frac{2}{\theta} \int_1^t \frac{1}{1-e^{-2\xi}} (U f_0 \circ \phi_\xi^X(p) - U f_0 \circ \phi_\xi^X \circ r_\theta(p)) \, d\xi - A_T}{B_T}, \quad t \sim \mathcal{U}_{[0,T]} \tag{6.8}$$

³⁴Possibly with the exception of the case examined at the end of Section 6.2.

as T tends to infinity. In the first place, this allows to rule out the existence of any non-trivial (namely not almost surely constant) distributional limit whenever one of the following conditions is met:

- (a) $\theta = 4\pi$;
- (b) Uf_0 is cohomologous to a constant function for the geodesic flow.

On the other hand, when Uf_0 is not cohomologous to a constant function, then the geodesic ergodic integrals $\int_1^t Uf_0 \circ \phi_\xi^X(p) \, d\xi$ are well-approximated by Brownian trajectories. More precisely, the Almost Sure Invariance Principle (see [70, 71], [51, Chap. 1] and [16]) for geodesic ergodic integrals asserts that there exist an auxiliary probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and two continuous-time stochastic processes $(X_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbf{P})$ such that the following hold:

- the law of the process $(X_t)_{t \geq 0}$ under the probability measure \mathbf{P} coincides with the law of the process $(\int_0^t Uf_0 \circ \phi_\xi^X(p) \, d\xi)_{t \geq 0}$ when p is sampled according to the probability measure vol ;
- the process $(B_t)_{t \geq 0}$ is a standard one-dimensional Brownian motion (cf. [44, Chap. 2]);
- there exists $\sigma \in \mathbb{R}^\times$ such that, for \mathbf{P} -almost every $\omega \in \Omega$,

$$|X_t(\omega) - B_{\sigma^2 t}(\omega)| = o(\sqrt{t}) \quad \text{as } t \rightarrow \infty. \quad (6.9)$$

As typical Brownian trajectories are of size $\sim \sqrt{t}$ at time t , the approximation in (6.9) enables to transfer classical results about the statistical behaviour of Brownian paths to analogous properties for geodesic ergodic integrals. In particular, there is no distributional limit³⁵ for

$$\frac{\int_0^t Uf_0 \circ \phi_\xi^X(p) \, d\xi - A_T}{B_T}, \quad t \sim \mathcal{U}_{[0, T]} \quad (6.10)$$

as T tends to infinity (cf. [17, Sec. 3.1]). Since the process $(\int_0^t Uf_0 \circ \phi_\xi^X \circ r_\theta(p) \, d\xi)_{t \geq 0}$ has the same law, for $p \sim \text{vol}$, as $(X_t(\omega))_{t \geq 0}$ for $\omega \sim \mathbf{P}$, the same applies to the random variables

$$\frac{\int_0^t Uf_0 \circ \phi_\xi^X \circ r_\theta(p) \, d\xi - A_T}{B_T}, \quad t \sim \mathcal{U}_{[0, T]}. \quad (6.11)$$

As already argued in Section 6.2 in the situation where the point p is selected randomly and the time T is fixed, here again the absence of distributional limits for each of the summands does not rule out, in principle, the possibility of non-trivial limits for the difference, hence for the random variables in (6.8). Once more, a painstaking analysis of the dependence features of the two processes in (6.10) and (6.11) might clarify the seemingly elusive pathwise behaviour of their difference.

7. THE HYPERBOLIC LATTICE POINT COUNTING PROBLEM

This final section is consecrated to the applications of our equidistribution results to lattice point counting problems in the hyperbolic plane; specifically, we shall first prove the precise asymptotics for the averaged counting function stated in Proposition 1.23 and subsequently deduce Theorem 1.24 on the error estimate for the pointwise counting.

Let Γ be a cocompact lattice in $\text{SL}_2(\mathbb{R})$, and denote by $d_{\mathbb{H}}$ the hyperbolic distance function on the hyperbolic upper-half plane \mathbb{H} (see Section 2.1). For each real number $R > 0$, let B_R be the closed $d_{\mathbb{H}}$ -ball of radius R centered at the point $i \in \mathbb{H}$, and define $N(R) = |\Gamma \cdot i \cap B_R|$, the cardinality of intersection of the Γ -orbit of i with B_R .

Recall also from Section 2.1 that $\text{SL}_2(\mathbb{R})$ acts on \mathbb{H} by Möbius transformations. In what follows, we identify the quotient manifold $\text{SL}_2(\mathbb{R})/\text{SO}_2(\mathbb{R})$ with \mathbb{H} whenever convenient, by means of the diffeomorphism $g\text{SO}_2(\mathbb{R}) \mapsto g \cdot i$, $g \in \text{SL}_2(\mathbb{R})$. The hyperbolic area measure $m_{\mathbb{H}}$ (namely the volume measure arising from the hyperbolic structure on \mathbb{H}) is the Radon measure

³⁵Actually, when Uf_0 has zero average over M , $A_T = 0$ and $B_T = \sqrt{T}$, any random variable may appear as distributional limit along an appropriate subsequence $(T_n)_{n \in \mathbb{N}}$ of times: see [17, Thm. 3.2].

on \mathbb{H} with density $dm_{\mathbb{H}}(x, y) = y^{-2} dx dy$ with respect to the induced Lebesgue measure on $\mathbb{H} \subset \mathbb{C}$.

Notation. In order not to overburden notation in the sequel, we shall denote $\mathrm{SL}_2(\mathbb{R})$ by G and $\mathrm{SO}_2(\mathbb{R})$ by K .

7.1. Asymptotics for the averaged counting function. For any subset $A \subset G/K$, we denote by $\mathbf{1}_A$ the indicator function of the set A . For any $R \in \mathbb{R}_{>0}$, we define a function $F_R: G/\Gamma \rightarrow \mathbb{R}_{\geq 0}$ by

$$F_R(g\Gamma) = \frac{|g\Gamma \cdot i \cap B_R|}{m_{\mathbb{H}}(B_R)} = \frac{1}{m_{\mathbb{H}}(B_R)} \sum_{\gamma\Gamma \cap K \in \Gamma/\Gamma \cap K} \mathbf{1}_{B_R}(g\gamma K), \quad g \in G; \quad (7.1)$$

observe that the function F_R is the subject of the averaged counting result in Proposition 1.23, which we now set out to prove.

Remark 7.1. We choose to deal with spaces of left cosets in the sequel; in particular, we replace the homogeneous spaces $M = \Gamma \backslash G$ we have been considering so far with G/Γ , identifying them via the diffeomorphism $\Gamma g \mapsto g^{-1}\Gamma$.

We follow the classical argument of Eskin and McMullen [24], which relies on the well-known folding-unfolding formula for invariant measures on homogeneous spaces. For the sake of completeness, we recall it in the setting of the group $G = \mathrm{SL}_2(\mathbb{R})$, referring the reader to [28, Sec. 2.6] or to [53, Chap. 1] for the general statements and their proofs.

Proposition 7.2. (1) *Let $H < G$ be a unimodular closed subgroup. Then there exists a non-zero G -invariant positive Radon measure on the quotient space G/H , which is uniquely determined up to positive real scalars. Moreover, if μ_G and μ_H are Haar measures³⁶ on G and H , respectively, there exists a unique normalization $\mu_{G/H}$ of the G -invariant measure on G/H such that, for any continuous compactly supported function $\varphi: G \rightarrow \mathbb{C}$, the following folding-unfolding formula holds:*

$$\int_G \varphi \, d\mu_G = \int_{G/H} \int_H \varphi(gh) \, d\mu_H(h) \, d\mu_{G/H}(gH). \quad (7.2)$$

(2) *Let $H < L$ be closed subgroups of G , and suppose G/L admits a non-zero finite G -invariant measure $\mu_{G/L}$ and L/H admits a non-zero finite L -invariant measure $\mu_{L/H}$. Then G/H admits a non-zero finite G -invariant measure $\mu_{G/H}$. Moreover, if $\mu_{G/L}$, $\mu_{L/H}$ and $\mu_{G/H}$ are compatibly normalized, then, for any continuous compactly supported function $\varphi: G/H \rightarrow \mathbb{C}$, it holds*

$$\int_{G/H} \varphi \, d\mu_{G/H} = \int_{G/L} \int_{L/H} \varphi(glH) \, d\mu_{L/H}(lH) \, d\mu_{G/L}(gL). \quad (7.3)$$

By means of standard approximation arguments in measure theory, formula (7.2) (resp. formula (7.3)) holds for any Borel-measurable function $\varphi: G \rightarrow \mathbb{C}$ (resp. $\varphi: G/H \rightarrow \mathbb{C}$) which either takes positive real values or is integrable with respect to μ_G (resp. $\mu_{G/H}$).

Let now m_K be the unique probability Haar measure on the compact group K , and normalize the Haar measure m_G on G so that, under the identification of G/K with \mathbb{H} , the resulting G -invariant measure on G/K (given by Proposition 7.2) corresponds to the hyperbolic area measure $m_{\mathbb{H}}$. Let $m_{G/\Gamma}$ be the unique G -invariant finite Borel measure on G corresponding, according to Proposition 7.2, to the given choice of Haar measure on G and to the counting measure on the discrete group Γ . Similarly, endowing the finite discrete group $\Gamma \cap K$ with the counting measure, we indicate with $m_{G/\Gamma \cap K}$, $m_{K/\Gamma \cap K}$ and $m_{\Gamma/\Gamma \cap K}$ the induced measures on the respective homogeneous spaces. Recall also that with $m_{K,\Gamma}$ we indicate the unique K -invariant Borel probability measure supported on the compact K -orbit of the identity coset Γ inside G/Γ (see the paragraph preceding Theorem 1.21).

³⁶The group G is perfect, hence unimodular; thus μ_G is also a right Haar measure.

The volumes of the homogeneous spaces G/Γ and $K/\Gamma \cap K$ with respect to the measures $m_{G/\Gamma}$ and $m_{K/\Gamma \cap K}$ are indicated with $\text{covol}_K(\Gamma \cap K)$ and $\text{covol}_G(\Gamma)$, respectively.

Fix now a real parameter $s > 11/2$ and a test function $\psi \in W^s(G/\Gamma)$. We expand, for any $R > 0$,

$$\begin{aligned}
\int_M \psi F_R \, dm_{G/\Gamma} &= \int_M \psi(g\Gamma) \left(\frac{1}{m_{\mathbb{H}}(B_R)} \sum_{\gamma \Gamma \cap K \in \Gamma/\Gamma \cap K} \mathbb{1}_{B_R}(g\gamma K) \right) dm_{G/\Gamma}(g\Gamma) \\
&= \frac{1}{m_{\mathbb{H}}(B_R)} \int_M \int_{\Gamma/\Gamma \cap K} \psi(g\Gamma) \mathbb{1}_{B_R}(g\gamma K) \, dm_{\Gamma/\Gamma \cap K}(\gamma \Gamma \cap K) \, dm_{G/\Gamma}(g\Gamma) \\
&= \frac{1}{m_{\mathbb{H}}(B_R)} \int_{G/\Gamma \cap K} \psi(g\Gamma) \mathbb{1}_{B_R}(gK) \, dm_{G/\Gamma \cap K}(g\Gamma \cap K) \\
&= \frac{1}{m_{\mathbb{H}}(B_R)} \int_{G/K} \int_{K/\Gamma \cap K} \psi(gk\Gamma) \mathbb{1}_{B_R}(gK) \, dm_{K/\Gamma \cap K}(k\Gamma \cap K) \, dm_{G/K}(gK) \\
&= \frac{1}{m_{\mathbb{H}}(B_R)} \int_{B_R} \int_{K/\Gamma \cap K} \psi \, dg_* m_{K/\Gamma \cap K} \, dm_{\mathbb{H}}(gK) \\
&= \frac{\text{covol}_K(\Gamma \cap K)}{m_{\mathbb{H}}(B_R)} \int_{B_R} \int_{G/\Gamma} \psi \, dg_* m_{K \cdot \Gamma} \, dm_{\mathbb{H}}(gK).
\end{aligned} \tag{7.4}$$

In the previous chain of equalities, we applied, in successive order,

- (1) the definition (7.1) of the function F_R ,
- (2) the fact that the invariant measure on the discrete space $\Gamma/\Gamma \cap K$ given by Proposition 7.2 is the counting measure,
- (3) formula (7.3) to the tower of subgroups $\Gamma \cap K < \Gamma < G$,
- (4) formula (7.3) to the tower of subgroups $\Gamma \cap K < K < G$,
- (5) the identification of \mathbb{H} with G/K and of $m_{\mathbb{H}}$ with $m_{G/K}$,
- (6) the relationship $m_{K/\Gamma \cap K} = \text{covol}_K(\Gamma \cap K) m_{K \cdot \Gamma}$, which is derived from the fact that $m_{K \cdot \Gamma}$ is a probability measure, the definition of $\text{covol}_K(\Gamma \cap K)$ and uniqueness up to scalars of the K -invariant measure on $K/\Gamma \cap K$, the latter being isomorphic to the orbit $K \cdot \Gamma$ as a topological K -homogeneous space.

We may now replace the inner integral in the last expression of (7.4) with the asymptotic expansion provided by Theorem 1.21 (with the caveat of Remark 7.1), thereby obtaining

$$\begin{aligned}
\int_{G/\Gamma} \psi F_R \, dm_{G/\Gamma} &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} \int_{G/\Gamma} \psi \, dm_{G/\Gamma} + \frac{\text{covol}_K(\Gamma \cap K)}{m_{\mathbb{H}}(B_R)} \\
&\int_{B_R} \left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \|g\|_{\text{op}}^{-1} (\cos(\mathfrak{S}\nu \log \|g\|_{\text{op}}) D_{\mu}^+ \psi(\Gamma, g) + \sin(\mathfrak{S}\nu \log \|g\|_{\text{op}}) D_{\mu}^- \psi(\Gamma, g)) \right. \\
&+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} \|g\|_{\text{op}}^{-(1+\nu)} D_{\mu}^+ \psi(\Gamma, g) + \|g\|_{\text{op}}^{-(1-\nu)} D_{\mu}^- \psi(\Gamma, g) \\
&\left. + \varepsilon_0 (\|g\|_{\text{op}}^{-1} D_{1/4}^+ \psi(\Gamma, g) + 2 \|g\|_{\text{op}}^{-1} \log \|g\|_{\text{op}} D_{1/4}^- \psi(\Gamma, g)) + \mathcal{R}\psi(\Gamma, g) \right) dm_{\mathbb{H}}(gK).
\end{aligned} \tag{7.5}$$

We shall need the following analogue of the classical integration formula on spheres in Euclidean spaces: for any $r > 0$, let $S_r = \partial B_r = \{z \in \mathbb{H} : d_{\mathbb{H}}(z, i) = r\}$ and σ_r the induced hyperbolic length measure on the circle S_r .

Proposition 7.3. *Let $f : \mathbb{H} \rightarrow \mathbb{C}$ be integrable with respect to $m_{\mathbb{H}}$. Then*

$$\int_{\mathbb{H}} f \, dm_{\mathbb{H}} = \int_0^{\infty} \int_{S_r} f(z) \, d\sigma_r(z) \, dr$$

The proof does not differ from the Euclidean case, for which we refer to [28, Thm. 2.49].

Define now, for any positive Casimir eigenvalue μ and ψ as above,

$$\alpha_{\psi,\mu}^{\pm}(r) = \int_{S_r} D_{\mu}^{\pm} \psi(\Gamma, z) \, d\sigma_r(z), \quad r > 0.$$

From (7.5) we get, thanks to Proposition 7.3,

$$\begin{aligned} \int_{G/\Gamma} \psi F_R \, dm_{G/\Gamma} &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} \int_{G/\Gamma} \psi \, dm_{G/\Gamma} + \frac{\text{covol}_K(\Gamma \cap K)}{m_{\mathbb{H}}(B_R)} \\ &\left(\sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \int_0^R e^{-\frac{r}{2}} \left(\cos\left(\frac{\mathfrak{S}\nu}{2} r\right) \alpha_{\psi,\mu}^+(r) + \sin\left(\frac{\mathfrak{S}\nu}{2} r\right) \alpha_{\psi,\mu}^-(r) \right) dr \right. \\ &+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} \int_0^R e^{-\frac{1+\nu}{2} r} \alpha_{\psi,\mu}^+(r) + e^{-\frac{1-\nu}{2} r} \alpha_{\psi,\mu}^-(r) \, dr \\ &\left. + \varepsilon_0 \left(\int_0^R e^{-\frac{r}{2}} \alpha_{\psi,1/4}^+(r) + r e^{-\frac{r}{2}} \alpha_{\psi,1/4}^-(r) \, dr \right) + \int_0^R \int_{S_r} \mathcal{R} \psi(\Gamma, z) \, d\sigma_r(z) \, dr \right). \end{aligned} \quad (7.6)$$

Let ψ_{μ} be the orthogonal projection of ψ onto the closed subspace $W^s(H_{\mu})$. By means of the uniform bounds (1.16) on the coefficients, we estimate

$$\sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}} |\alpha_{\psi,\mu}^{\pm}(r)| \leq \sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}} \|D_{\mu}^{\pm} \psi(\Gamma, \cdot)\|_{\infty} \int_{S_r} d\sigma_r(z) \leq 2\pi \frac{C_{1,s-3} C'_{\text{Spec}}}{4\pi} \|\psi\|_{W^s} \sinh r \quad (7.7)$$

for any $r > 0$, as the hyperbolic length of S_r equals³⁷ $2\pi \sinh r$. Recalling that $m_{\mathbb{H}}(B_R) = 2\pi(\cosh R - 1)$ for any $R > 0$, we deduce from (7.6) that

$$\begin{aligned} \int_{G/\Gamma} \psi F_R \, dm_{G/\Gamma} &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} \int_{G/\Gamma} \psi \, dm_{G/\Gamma} \\ &+ \text{covol}_K(\Gamma \cap K) \left(e^{-\frac{R}{2}} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \beta_{\psi,\mu}^+(R) + \beta_{\psi,\mu}^-(R) \right. \\ &+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2} R} \beta_{\psi,\mu}^+(R) + e^{-\frac{1-\nu}{2} R} \beta_{\psi,\mu}^-(R) \\ &\left. + \varepsilon_0 (e^{-\frac{R}{2}} \beta_{\psi,1/4}^+(R) + R e^{-\frac{R}{2}} \beta_{\psi,1/4}^-(R)) + \gamma_{\psi}(R) \right) \end{aligned}$$

³⁷This is an elementary verification in hyperbolic geometry, for instance approximating circles with regular n -gons; their hyperbolic perimeter can in turn be easily computed by means of explicit formulas for the hyperbolic distance (see [36, Thm. 1.2.6]) and of the hyperbolic cosine law (see [36, Thm. 1.5.2]).

Similarly, the hyperbolic area of a ball is easily computed by approximation via the Gauss-Bonnet formula for the area of hyperbolic triangles (cf. [36, Thm. 1.4.2]).

for any $R \geq 1$, where we have set

$$\begin{aligned}\beta_{\psi,\mu}^+(R) &= \frac{e^{-\frac{R}{2}}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R e^{-\frac{r}{2}} \cos\left(\frac{\Im\nu}{2}r\right) \alpha_{\psi,\mu}^+(r) dr, \quad \mu > 1/4, \\ \beta_{\psi,\mu}^-(R) &= \frac{e^{-\frac{R}{2}}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R e^{-\frac{r}{2}} \sin\left(\frac{\Im\nu}{2}r\right) \alpha_{\psi,\mu}^-(r) dr, \quad \mu > 1/4, \\ \beta_{\psi,\mu}^\pm(R) &= \frac{e^{-\frac{1\pm\nu}{2}R}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R e^{-\frac{1\pm\nu}{2}r} \alpha_{\psi,\mu}^\pm(r) dr, \quad 0 < \mu < 1/4, \\ \beta_{\psi,1/4}^+(R) &= \frac{e^{-\frac{R}{2}}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R e^{-\frac{r}{2}} \alpha_{\psi,1/4}^+(r) dr, \\ \beta_{\psi,1/4}^-(R) &= \frac{R^{-1}e^{-\frac{R}{2}}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R e^{-\frac{r}{2}} \alpha_{\psi,1/4}^-(r) dr, \\ \gamma_\psi(R) &= \frac{e^{-R}}{\pi(1-2e^{-R}+e^{-2R})} \int_0^R \int_{S_r} \mathcal{R}\psi(\Gamma, z) d\sigma_r(z) dr.\end{aligned}$$

Because of (7.7), we have the following estimates on the previous coefficients: for any $R \geq 1$,

$$\sum_{\mu \in \text{Spec}(\square) \cap \mathbb{R}_{>0}} |\beta_{\psi,\mu}^\pm(R)| \leq \frac{5C_{1,s-3}C'_{\text{Spec}}}{2\pi} \|\psi\|_{W^s}$$

and

$$|\gamma_\psi(R)| \leq \frac{5C_{1,s-3}C_{\text{Spec}}}{4\pi} \|\psi\|_{W^s} (R+1)e^{-R},$$

using the (crude) bound $(1-2e^{-R}+e^{-2R})^{-1} \leq 5$ in each of the previous inequalities.

This establishes Proposition (1.23) in its entirety.

7.2. Error estimate for the pointwise counting problem. This subsection is devoted to the deduction of the error estimate for the counting problem stated in Theorem 1.24, starting from the asymptotic expansion (1.27) for the averaged counting function.

Recall from (7.1) that, for any real number $R > 0$, the ratio $N(R)/m_{\mathbb{H}}(B_R)$ equals the value of the function F_R at the identity coset $\Gamma \in G/\Gamma$. In order to find a convenient approximation for the latter, we shall compare it with the averages

$$\int_{G/\Gamma} \psi F_R dm_{G/\Gamma}$$

where the function ψ ranges over a suitably defined approximate identity³⁸ in G/Γ .

We now present the details. Let us fix a parameter $\delta \in \mathbb{R}_{>0}$, on which we shall subsequently impose conditions according to the needs of the argument; choose

- (a) an open symmetric³⁹ neighborhood U_δ of the identity in G such that, for any $R > 0$,

$$B_{R-\delta} \subset \bigcap_{g \in U_\delta} g \cdot B_R \subset \bigcup_{g \in U_\delta} g \cdot B_R \subset B_{R+\delta} \quad (7.8)$$

- (b) and a smooth function $\psi_\delta: G/\Gamma \rightarrow \mathbb{R}_{\geq 0}$ with compact support contained in the open set $U_\delta\Gamma = \{g\Gamma : g \in U_\delta\}$, and satisfying

$$\int_{G/\Gamma} \psi_\delta dm_{G/\Gamma} = 1. \quad (7.9)$$

³⁸The terminology is common in the context of locally compact groups; see, for instance, [28, Sec. 2.5].

³⁹This means that U_δ coincides with the set of inverses of its elements.

Remark 7.4. The existence, for any $\delta > 0$, of a neighborhood U_δ with the properties claimed above is routinely referred to in the literature (see, for instance, [24]) as the *well-roundedness* property of the collection of balls $(B_R)_{R>0}$. A geometric condition of this sort affords to leverage equidistribution results to study lattice point counting problems.

Observe that we may harmlessly replace U_δ with

$$KU_\delta = \bigcup_{k \in K} kU_\delta,$$

and thus assume that U_δ is saturated with respect to left translations by elements of K . Property (7.8) is unaffected: for any $k \in K$ and $z \in \mathbb{H}$, we have

$$d_{\mathbb{H}}(k \cdot z, i) = d_{\mathbb{H}}(k \cdot z, k \cdot i) = d_{\mathbb{H}}(z, i),$$

as the subgroup K fixes i and acts by hyperbolic isometries; therefore $k \cdot B_r = B_r$ for any $k \in K$ and any $r > 0$. As a consequence of this, we might and shall assume that ψ_δ is K -invariant.

We now express, the ratio $N(R)/m_{\mathbb{H}}(B_R)$, for any $R > 0$, as

$$\begin{aligned} F_R(\Gamma) &= F_R(\Gamma) - \int_{G/\Gamma} \psi_\delta F_R \, dm_{G/\Gamma} + \int_{G/\Gamma} \psi_\delta F_R \, dm_{G/\Gamma} \\ &= \int_{G/\Gamma} \psi_\delta(g\Gamma)(F_R(\Gamma) - F_R(g\Gamma)) \, dm_{G/\Gamma}(g\Gamma) + \int_{G/\Gamma} \psi_\delta F_R \, dm_{G/\Gamma}, \end{aligned}$$

where the second inequality follows from the property (7.9). Let us call $\mathcal{E}_\delta(R)$, for notational simplicity, the quantity

$$\int_{G/\Gamma} \psi_\delta(g\Gamma)(F_R(\Gamma) - F_R(g\Gamma)) \, dm_{G/\Gamma}(g\Gamma);$$

in view of the asymptotics (1.27) applied to $\int_{G/\Gamma} \psi_\delta F_R \, dm_{G/\Gamma}$, we may write

$$\begin{aligned} F_R(\Gamma) &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} + \mathcal{E}_\delta(R) \\ &+ \text{covol}_K(\Gamma \cap K) \left(e^{-\frac{R}{2}} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \beta_{\psi_\delta, \mu}^+(R) + \beta_{\psi_\delta, \mu}^-(R) \right. \\ &+ \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2}R} \beta_{\psi_\delta, \mu}^+(R) + e^{-\frac{1-\nu}{2}R} \beta_{\psi_\delta, \mu}^-(R) \\ &\left. + \varepsilon_0 \left(e^{-\frac{R}{2}} \beta_{\psi_\delta, 1/4}^+(R) + R e^{-\frac{R}{2}} \beta_{\psi_\delta, 1/4}^-(R) \right) + \gamma_{\psi_\delta}(R) \right). \end{aligned} \quad (7.10)$$

We estimate, for any $R > 0$,

$$|\mathcal{E}_\delta(R)| \leq \int_{G/\Gamma} \psi_\delta(g\Gamma) |F_R(\Gamma) - F_R(g\Gamma)| \, dm_{G/\Gamma}(g\Gamma) \leq \sup_{g \in U_\delta} |F_R(\Gamma) - F_R(g\Gamma)|, \quad (7.11)$$

the last inequality being a consequence of (7.9) and the fact that $\text{supp } \psi_\delta \subset U_\delta \Gamma$.

Now, for any $g \in U_\delta$, we have

$$\begin{aligned} |F_R(\Gamma) - F_R(g\Gamma)| &= \frac{|\Gamma \cdot H \cap B_R| - |g\Gamma \cdot H \cap B_R|}{m_{\mathbb{H}}(B_R)} = \frac{|\Gamma \cdot H \cap B_R| - |\Gamma \cdot H \cap g^{-1} \cdot B_R|}{m_{\mathbb{H}}(B_R)} \\ &\leq \frac{|\Gamma \cdot H \cap ((\bigcup_{g \in U_\delta} g \cdot B_R) \setminus (\bigcap_{g \in U_\delta} g \cdot B_R))|}{m_{\mathbb{H}}(B_R)} \leq \frac{N(R + \delta) - N(R - \delta)}{m_{\mathbb{H}}(B_R)} \\ &= F_{R+\delta}(\Gamma) \frac{m_{\mathbb{H}}(B_{R+\delta})}{m_{\mathbb{H}}(B_R)} - F_{R-\delta}(\Gamma) \frac{m_{\mathbb{H}}(B_{R-\delta})}{m_{\mathbb{H}}(B_R)}, \end{aligned} \quad (7.12)$$

where the second-to-last inequality follows from (7.8). Choose $R_0 = R_0(\Gamma) > 0$ such that the quantities

$$M = \sup_{r \geq R_0} F_r(\Gamma) \quad \text{and} \quad m = \inf_{r \geq R_0} F_r(\Gamma)$$

are non-zero and finite⁴⁰. Plugging (7.12) into (7.11), we get that, for any $R \geq 2R_0$ and $\delta < R_0$,

$$\begin{aligned} |\mathcal{E}_\delta(R)| &\leq (M - m) \left(\frac{m_{\mathbb{H}}(B_{R+\delta})}{m_{\mathbb{H}}(B_R)} - \frac{m_{\mathbb{H}}(B_{R-\delta})}{m_{\mathbb{H}}(B_R)} \right) \\ &= \frac{(M - m)}{1 - 2e^{-R} + e^{-2R}} \left(e^\delta (1 - 2e^{-(R+\delta)} + e^{-2(R+\delta)}) - e^{-\delta} (1 - 2e^{-(R-\delta)} + e^{-2(R-\delta)}) \right) \\ &\leq (M - m)(ce^\delta - e^{-\delta}), \end{aligned} \quad (7.13)$$

where we might for example take $c = c_\Gamma = \frac{1+e^{-2R_0}}{1-2e^{-R_0}}$.

We now let the parameter δ be a function of the radius R ; for reasons which we will shortly elucidate (see Remark 7.6), we let $\delta = \delta(R) = e^{-\eta R}$ in (7.10), for a certain $\eta > 0$ to be determined later on. In this way, we obtain an expression of the form

$$\begin{aligned} F_R(\Gamma) &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} + \mathcal{E}_{e^{-\eta R}}(R) \\ &\quad + \text{covol}_K(\Gamma \cap K) \left(e^{-\frac{R}{2}} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) + \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) \right. \\ &\quad + \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{-\frac{1+\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) + e^{-\frac{1-\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) \\ &\quad \left. + \varepsilon_0 \left(e^{-\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^+(R) + R e^{-\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^-(R) \right) + \gamma_{\psi_{e^{-\eta R}}}(R) \right), \end{aligned} \quad (7.14)$$

which does not depend on the parameter δ any longer.

Multiplying by $m_{\mathbb{H}}(B_R)$ on both sides of (7.14) yields

$$\begin{aligned} N(R) &= \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} m_{\mathbb{H}}(B_R) + \pi(1 - 2e^{-R} + e^{-2R}) e^R \mathcal{E}_{e^{-\eta R}}(R) \\ &\quad + \pi(1 - 2e^{-R} + e^{-2R}) \text{covol}_K(\Gamma \cap K) \left(e^{\frac{R}{2}} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) + \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) \right. \\ &\quad + \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{\frac{1-\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) + e^{\frac{1+\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) \\ &\quad \left. + \varepsilon_0 \left(e^{\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^+(R) + R e^{\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^-(R) \right) + e^R \gamma_{\psi_{e^{-\eta R}}}(R) \right). \end{aligned} \quad (7.15)$$

In order to reach an accurate upper bound for the error

$$E(R) = \left| N(R) - \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} m_{\mathbb{H}}(B_R) \right|,$$

in our counting problem, it remains to determine which are the highest-order terms in the expansion (7.15). To this end, it is relevant to estimate the Sobolev norms of the functions $\psi_{e^{-\eta R}}$ for $R > 0$, because of the bounds (1.26) and (1.28).

⁴⁰A straightforward modification of the effective argument we are running leads to the well-known non-effective convergence

$$F_R(\Gamma) \xrightarrow{R \rightarrow \infty} \frac{\text{covol}_K(\Gamma \cap K)}{\text{covol}_G(\Gamma)} \in \mathbb{R}_{>0}.$$

Lemma 7.5. *For any $0 < \delta < 1$, the function ψ_δ can be chosen to satisfy*

$$\|\psi_\delta\|_{W^s} \leq \delta^{-(1+s)} \|\psi_1\|_{W^s} \quad (7.16)$$

for any $s > 0$.

Proof. Recall that ψ_δ is assumed to be K -invariant or, in other words, a smooth compactly supported function on the two-dimensional manifold $K \backslash G / \Gamma$. Since any Riemannian metric on $K \backslash G / \Gamma$ is equivalent, on a fixed compact coordinate ball containing the identity coset $Ke\Gamma$, to the Euclidean metric on a compact neighborhood of the origin in \mathbb{R}^2 (cf. [42, Lem. 13.28]), the problem of constructing ψ_δ so to meet our requirement can be transferred to the Euclidean plane. Specifically, we would like to construct a collection $(\psi_\delta)_{0 < \delta \leq 1}$ of mollifiers (cf. [9, Sec. 4.4]) so that (7.16) is satisfied, where $\|\cdot\|_{W^s}$ are now the standard fractional Sobolev norms on \mathbb{R}^2 . A straightforward computation allows to ascertain that the customary choice

$$\psi_\delta(x) = \frac{1}{\delta^2} \psi_1\left(\frac{x}{\delta}\right), \quad x \in \mathbb{R}^2,$$

where ψ_1 is a fixed compactly supported smooth nonnegative function with unit average over \mathbb{R}^2 , fulfills (7.16). \square

Henceforth, we assume that the collection $(\psi_\delta)_{0 < \delta < 1}$ satisfies the condition in Lemma 7.5.

We remind the reader that we indicate with μ_* the spectral gap of the hyperbolic surface $S = \Gamma \backslash \mathbb{H}$, that is, the infimum of the set $\text{Spec}(\square) \cap \mathbb{R}_{>0}$. Also, we denote by ν_* the complex number uniquely determined by the properties $\nu_* \in \mathbb{R}_{\geq 0} \cup i\mathbb{R}_{>0}$ and $1 - \nu_*^2 = 4\mu_*$.

Since $e^\delta - e^{-\delta} \sim 2\delta$ for $\delta \sim 0$, we deduce from (7.13) that the term $e^R \mathcal{E}_{e^{-\eta R}}(R)$ is at most of order $e^{(1-\eta)R}$. On account of Lemma 7.5, the highest-order term in the expression

$$\begin{aligned} & e^{\frac{R}{2}} \sum_{\mu \in \text{Spec}(\square), \mu > 1/4} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) + \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) + \sum_{\mu \in \text{Spec}(\square), 0 < \mu < 1/4} e^{\frac{1-\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^+(R) \\ & + e^{\frac{1+\nu}{2}R} \beta_{\psi_{e^{-\eta R}, \mu}}^-(R) + \varepsilon_0 \left(e^{\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^+(R) + R e^{\frac{R}{2}} \beta_{\psi_{e^{-\eta R}, 1/4}}^-(R) \right) + e^R \gamma_{\psi_{e^{-\eta R}}}(R) \end{aligned}$$

is $e^{\frac{1+\Re\nu_*}{2}R} \beta_{\psi_{e^{-\eta R}, \mu_*}}(R)$; because of Lemma 7.5 and the bound (1.26), the latter is at most of order

$$e^{\frac{1+\Re\nu_*}{2}R} e^{(1+s)\eta R} = e^{\frac{1+\Re\nu_*+2(1+s)\eta}{2}R}.$$

Remark 7.6. The reason for choosing δ to decay exponentially fast with R becomes now apparent: it is the only way to get a sensible comparison between the orders of the two terms considered above.

Ostensibly, the optimal choice of the parameter η for our purposes is

$$\eta = \frac{1 - \Re\nu_*}{2(2+s)},$$

which realizes the equality of exponents

$$1 - \eta = \frac{1 + \Re\nu_* + 2(1+s)\eta}{2}.$$

Bearing in mind that the K -invariance of $\psi_{e^{-\eta R}}$ allows to choose s can arbitrarily close to $9/2$ (see Theorem 1.12), it is straightforward to deduce that, setting $\eta_* = \frac{1}{13}(1 - \Re\nu_*)$, we have

$$\lim_{R \rightarrow \infty} \frac{E(R)}{e^{(1-\eta_*+\varepsilon)R}} = 0$$

for any $\varepsilon > 0$, which establishes Theorem 1.24.

REFERENCES

- [1] R.A. Adams. *Sobolev spaces*. Pure and Applied Mathematics, 65. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.
- [2] T. Aubin. *Nonlinear analysis on manifolds. Monge-Ampère equations*. Grundlehren der mathematischen Wissenschaften, 252. Springer-Verlag, New York, 1982.
- [3] V. Bargmann, *Irreducible unitary representations of the Lorentz group*, Ann. of Math. **48** (1947), 568–640.
- [4] M.B. Bekka, M. Mayer. *Ergodic theory and topological dynamics of group actions on homogeneous spaces*. London Mathematical Society Lecture Note Series, 269. Cambridge University Press, Cambridge, 2000.
- [5] Y. Benoist, H. Oh, *Effective equidistribution of S -integral points on symmetric varieties*, Ann. Inst. Fourier **62** (2012), 1889–1942.
- [6] N. Bergeron. *The spectrum of hyperbolic surfaces*. Universitext. Springer, Cham; EDP Sciences, Les Ulis, 2016.
- [7] P. Billingsley. *Convergence of probability measures*. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999.
- [8] A. Borel. *Automorphic forms on $SL_2(\mathbb{R})$* . Cambridge Tracts in Mathematics, 130. Cambridge University Press, Cambridge, 1997.
- [9] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer New York, NY, 2011.
- [10] M.R. Bridson, A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 1999.
- [11] A. Bufetov, G. Forni, *Limit theorems for horocycle flows*, Ann. Sci. Éc. Norm. Supér. **47** (2014), 851–903.
- [12] M. Burger, *Horocycle flow on geometrically finite surfaces*, Duke Math. J. **61** (1990), 779–803.
- [13] P. Buser. *Geometry and spectra of compact Riemann surfaces*. Progress in Mathematics, 106. Birkhäuser Boston, Inc., Boston, MA, 1992.
- [14] P. Colognese, M. Pollicott, *The growth and distribution of large circles on translation surfaces*, arXiv:2107.14058 (2021).
- [15] J. Delsarte, *Sur le Gitter Fuchsien*, C. R. Acad. Sci. Paris **214** (1942), 147–149.
- [16] M. Denker, W. Philipp, *Approximation by Brownian motion for Gibbs measures and flows under a function*, Ergodic Theory Dynam. Systems **4** (1984), 541–552.
- [17] D. Dolgopyat, O. Sarig, *Temporal distributional limit theorems for dynamical systems*, J. Stat. Phys. **166** (2017), 680–713,
- [18] W. Duke, Z. Rudnick and P. Sarnak, *Density of integer points on affine homogeneous spaces*, Duke Math. J. **71** (1993), 143–179.
- [19] S. Dyatlov, F. Faure and C. Guillarmou, *Power spectrum of the geodesic flow on hyperbolic manifolds*, Anal. PDE **8** (2015), 923–1000.
- [20] S. Edwards, *On the rate of equidistribution of expanding translates of horospheres in $\Gamma \backslash G$* , Comment. Math. Helv. **96** (2021), 275–337.
- [21] S. Edwards, *On the equidistribution of translates of orbits of symmetric subgroups in $\Gamma \backslash G$* , available at <https://sites.google.com/view/samedwards>, preprint (2018).
- [22] M. Einsiedler, G. Margulis and A. Venkatesh, *Effective equidistribution of closed orbits of semisimple groups on homogeneous spaces*, Invent. Math. **177** (2009), 137–212.
- [23] M. Einsiedler, T. Ward. *Ergodic Theory with a view towards Number Theory*. Graduate Texts in Mathematics, Springer-Verlag, London, 2011.
- [24] A. Eskin, C. McMullen, *Mixing, counting and equidistribution in Lie groups*, Duke Math. J. **71** (1993), 181–209.
- [25] A. Eskin, S. Mozes and N. Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. of Math. **143** (1996), 253–299.
- [26] H. Farkas, I. Kra. *Riemann Surfaces*. Second Edition. Graduate Texts in Mathematics, 71. Springer-Verlag, 1980.
- [27] L. Flaminio, G. Forni, *Invariant distributions and time averages for horocycle flows*, Duke Math. J. **119** (2003), 465–536.
- [28] G.B. Folland. *A Course in Abstract Harmonic Analysis*. Second Edition. Textbooks in Mathematics, CRC Press, Portland, OR, 2015.
- [29] G.B. Folland. *Real Analysis*. Second Edition. Pure and Applied Mathematics, John Wiley & Sons, New York, NY, 1999.
- [30] G. Forni, *Ruelle resonances from cohomological equations*, arXiv:2007.03116 (2020).
- [31] A. Gorodnik, A. Nevo. *The ergodic theory of lattice subgroups*. Annals of Mathematics Studies, 172. Princeton University Press, Princeton, NJ, 2010.

- [32] E. Hebey. *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*. Courant Lecture Notes in Mathematics, 5. New York University, Courant Institute of Mathematical Sciences, New York. American Mathematical Society, Providence, RI, 1999.
- [33] A. Ivić, E. Krätzel, M. Kühleitner and W.G. Nowak. *Lattice points in large regions and related arithmetic functions: recent developments in a very classical topic*. Elementare und analytische Zahlentheorie, 89–128, Franz Steiner Verlag Stuttgart, Stuttgart, 2006.
- [34] H. Iwaniec. *Spectral methods of automorphic forms*. Second edition. Graduate Studies in Mathematics, 53. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002.
- [35] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995.
- [36] S. Katok. *Fuchsian Groups*. Chicago Lecture Notes in Mathematics, University of Chicago Press, Chicago, IL, 1992.
- [37] D. Kleinbock, G.A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, Amer. Math. Soc. Transl. **171** (1996), 141–172.
- [38] A.W. Knap. *Lie groups beyond an introduction*. Second Edition. Progress in Mathematics, Birkhäuser Boston, New York, NY, 2002.
- [39] B. Kra, N.A. Shah and W. Sun, *Equidistribution of dilated curves on nilmanifolds*, J. Lond. Math. Soc. **98** (2018), 708–732.
- [40] S. Lang. $SL_2(\mathbf{R})$. Graduate Texts in Mathematics, 105. Springer-Verlag, New York, 1985.
- [41] P. Lax, R. Phillips, *The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces*, J. Funct. Anal. **46** (1982), 280–350.
- [42] J.M. Lee. *Introduction to Smooth Manifolds*. Second Edition. Graduate Texts in Mathematics, 218. Springer, New York, 2013.
- [43] J.M. Lee. *Introduction to Riemannian Manifolds*. Second Edition. Graduate Texts in Mathematics, 176. Springer, Cham, 2018.
- [44] J.-F. Le Gall. *Brownian motion, martingales, and stochastic calculus*. Graduate Texts in Mathematics, 274. Springer, Cham, 2016.
- [45] A. Livsic, *Certain properties of the homology of Y -systems*, Mat. Zametki **10** (1971), 555–564.
- [46] G.W. Mackey. *The theory of unitary group representations*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, Ill.-London, 1976.
- [47] G.A. Margulis, *Applications of ergodic theory to the investigation of manifolds of negative curvature*, Funct. Anal. Appl. **4** (1969), 335.
- [48] G.A. Margulis. *On some aspects of the theory of Anosov systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2004.
- [49] J. Marklof. *Selberg’s trace formula: an introduction*. Hyperbolic geometry and applications in quantum chaos and cosmology, 83–119. London Math. Soc. Lecture Note Ser., 397. Cambridge Univ. Press, Cambridge, 2012.
- [50] J.R. Munkres. *Topology*. Second Edition. Prentice Hall, Upper Saddle River, NJ, 2000.
- [51] W. Philipp, W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Mem. Amer. Math. Soc. **2** no. 161 (1975), iv+140pp.
- [52] R. Phillips, Z. Rudnick, *The circle problem in the hyperbolic plane*, J. Funct. Anal. **121** (1994), 78–116.
- [53] M.S. Raghunathan. *Discrete subgroups of Lie groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68. Springer-Verlag, New York-Heidelberg, 1972.
- [54] J.G. Ratcliffe. *Foundations of hyperbolic manifolds*. Third Edition. Graduate Texts in Mathematics, 149. Springer, Cham, 2019.
- [55] B. Randol, *The behaviour under projection of dilating sets in a covering space*, Trans. Amer. Math. Soc. **285** (1984), 855–859.
- [56] M. Ratner, *The central limit theorem for geodesic flows on n -dimensional manifolds of negative curvature*, Israel J. Math. **16** (1973), 181–197.
- [57] M. Ratner, *The rate of mixing for geodesic and horocycle flows*, Ergodic Theory Dynam. Systems **7** (1987), 267–288.
- [58] D. Ravotti, *Quantitative equidistribution of horocycle push-forwards of transverse arcs*, Enseign. Math. **66** (2020), 135–150.
- [59] D. Ravotti, *Asymptotics and limit theorems for horocycle ergodic integrals à la Ratner*, arXiv:2107.02090 (2021).
- [60] P. Sarnak, *Spectra of hyperbolic surfaces*, Bull. Amer. Math. Soc. **40** (2003), 441–478.
- [61] S. Sasaki, *On the differential geometry of tangent bundles of Riemannian manifolds*, Tohoku Math. J. **10** (1958), 338–354.
- [62] L. Schwartz. *Analyse. Topologie générale et analyse fonctionnelle*. Collection Enseignement des Sciences, No. 11. Hermann, Paris, 1970.

- [63] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. **20** (1956), 47–87.
- [64] A. Selberg, *Equidistribution in discrete groups and the spectral theory of automorphic forms*, <https://publications.ias.edu/selberg/section/2491>.
- [65] N.A. Shah, *Limit distributions of expanding translates of certain orbits on homogeneous spaces*, Proc. Indian Acad. Sci. Math. Sci. **106** (1996), 105–125.
- [66] N.A. Shah, *Limiting distributions of curves under geodesic flow on hyperbolic manifolds*, Duke Math. J. **148** (2009), 251–279.
- [67] N.A. Shah, *Equidistribution of expanding translates of curves and Dirichlet’s theorem on Diophantine approximation*, Invent. Math. **177** (2009), 509–532.
- [68] N.A. Shah, *Expanding translates of curves and Dirichlet-Minkowski theorem on linear forms*, J. Amer. Math. Soc. **23** (2010), 563–589.
- [69] Ya.G. Sinai, *The central limit theorem for geodesic flows on manifolds of constant negative curvature*, Soviet Math. Dokl., **1** (1960), 938–987.
- [70] V. Strassen, *An invariance principle for the law of the iterated logarithm*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **3** (1964), 211–226.
- [71] V. Strassen, *Almost sure behaviour of sums of independent random variables and martingales*. Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, pp. 315–343. Univ. California Press, Berkeley, Calif., 1967.
- [72] R.S. Strichartz, *Analysis of the Laplacian on the complete Riemannian manifold*, J. Funct. Anal. **52** (1983), 48–79,
- [73] R.S. Strichartz, *Magnified curves on a flat torus, determination of almost periodic functions, and the Riemann-Lebesgue lemma*, Proc. Amer. Math. Soc. **107** (1989), 755–759,
- [74] A. Strömbergsson, *On the uniform equidistribution of long closed horocycles*, Duke Math. J. **123** (2004), 507–547,
- [75] A. Strömbergsson, *On the deviation of ergodic averages for horocycle flows*, J. Mod. Dyn. **7** (2013), 291–328.
- [76] A. Terras. *Harmonic analysis on symmetric spaces – Euclidean space, the sphere, and the Poincaré upper half plane*. Second edition. Springer, New York, 2013.
- [77] W.P. Thurston. *The Geometry and Topology of Three-Manifolds*. Electronic version. <https://www.msri.org/publications/books/gt3m>, 2002.
- [78] H. Triebel. *Theory of function spaces II*. Monographs in Mathematics, 84. Birkhäuser Verlag, Basel, 1992.
- [79] A. Venkatesh, *Sparse equidistribution problems, period bounds and subconvexity*, Ann. of Math. **172** (2010), 989–1094.
- [80] P. Yang, *Equidistribution of expanding translates of curves and Diophantine approximation on matrices*, Invent. Math. **220** (2020), 909–948.

(E.C.) ETH ZÜRICH, RÄMISTRASSE 101 CH-8092 ZÜRICH SWITZERLAND

Email address: corsoemilio2@gmail.com

(D.R.) UNIVERSITÄT WIEN, DEPARTMENT OF MATHEMATICS, OSKAR-MORGENSTERN-PLATZ 1, 1090 WIEN, AUSTRIA

Email address: davide.ravotti@gmail.com