A LOWER BOUND FOR THE REMAINDER IN WEYL'S LAW ON NEGATIVELY CURVED SURFACES

DMITRY JAKOBSON, IOSIF POLTEROVICH, AND JOHN A. TOTH

ABSTRACT. We obtain an estimate from below for the remainder in Weyl's law on negatively curved surfaces. In the constant curvature case, such a bound was proved independently by Hejhal and Randol in 1976 using the Selberg zeta function techniques. Our approach works in arbitrary negative curvature, and is based on wave trace asymptotics for long times, equidistribution of closed geodesics and small-scale microlocalization.

1. Introduction and main results

1.1. **Weyl's law.** Let X be a compact negatively curved surface of area A with the Riemannian metric $\{g_{ij}\}$. We assume that the Gaussian curvature $K(x)$ satisfies

$$-K_1^2 \le K(x) \le -K_2^2 \tag{1.1.1}$$

at every point $x \in X$. Let Δ be the Laplacian on X with the eigenvalues $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \dots$ and the corresponding orthonormal basis $\{\phi_i\}$ of eigenfunctions: $\Delta\phi_i = \lambda_i\phi_i$. Let $N(\lambda) = \#\{\sqrt{\lambda_i} \le \lambda\}$ be the eigenvalue counting function. The asymptotic behavior of $N(\lambda)$ is given by *Weyl's law* ([Ho1]):

$$N(\lambda) = \frac{A}{4\pi}\lambda^2 + R(\lambda), \qquad R(\lambda) = O(\lambda). \tag{1.1.2}$$

It shown in [Ber] that on a nonpositively curved surface $R(\lambda) = O\left(\frac{\lambda}{\ln \lambda}\right)$. In the present paper we study lower bounds for $R(\lambda)$. As in [J-P], one of our tools is thermodynamic formalism for hyperbolic flows.

1.2. **Thermodynamic formalism.** Let G^t be the geodesic flow on the unit tangent bundle SX and let E^u_ξ be the (one-dimensional) unstable subspace for G^t, $\xi \in SX$. The *Sinai-Ruelle-Bowen potential* is a Hölder continuous function $\mathcal{H} : SX \to \mathbb{R}$ which for any $\xi \in SX$ is defined by the formula (see [B-R], [Sin])

$$\mathcal{H}(\xi) = \frac{d}{dt}\bigg|_{t=0} \ln\det dG^t|_{E^u_\xi}, \tag{1.2.1}$$

For any continuous function $f : SX \to \mathbb{R}$ one can define the *topological pressure*

$$P(f) = \sup_\mu \left(h_\mu + \int f d\mu\right), \tag{1.2.2}$$

2000 *Mathematics Subject Classification.* Primary: 58J50 Secondary: 35P20, 37D20.

Key words and phrases. Weyl's law, wave trace, negative curvature, thermodynamic formalism, small-scale microlocalization.

D.J. and J.T. are supported by NSERC, FQRNT and Dawson Fellowships. I.P. is supported by NSERC and FQRNT.

where the supremum is taken over all G^t-invariant measures μ and h_μ denotes the *measure-theoretical entropy* of the geodesic flow (see [Bow]). In particular $P(0) = h$, where h is the *topological entropy* of the flow. It is well-known that for the Sinai-Ruelle-Bowen potential $P(-\mathcal{H}) = 0$ and the corresponding equilibrium measure (i.e. the measure on which the supremum is attained) is the Liouville measure μ_L on the unit tangent bundle:

$$h_{\mu_L} = \int_{SX} \mathcal{H} d\mu_L \tag{1.2.3}$$

1.3. **Main result.** Recall that $f_1(\lambda) = \Omega(f_2(\lambda))$ for a function f_1 and a positive function f_2 means that $\limsup_{\lambda\to\infty} |f_1(\lambda)|/f_2(\lambda) > 0$.

Theorem 1.3.1. *Let X be a compact surface of negative curvature. Then*

$$R(\lambda) = \Omega\left((\ln\lambda)^{\frac{P(-\mathcal{H}/2)}{h}-\varepsilon}\right) \quad \forall\, \varepsilon > 0. \tag{1.3.2}$$

As was shown in [J-P, section 1.4], the power of the logarithm in (1.3.2) is always positive:

$$\frac{P\left(-\mathcal{H}/2\right)}{h} \geq \frac{K_2}{2K_1} > 0.$$

Moreover, if the curvature is constant, the bound (1.3.2) reads

$$R(\lambda) = \Omega\left((\ln\lambda)^{\frac{1}{2}-\varepsilon}\right) \tag{1.3.3}$$

for any $\varepsilon > 0$. The estimate (1.3.3) was obtained in [Ran] and, in a slightly stronger form, in [Hej, section 17], using the Selberg zeta function techniques. Our approach, based on the Duistermaat-Guillemin wave trace formula, thermodynamic formalism and semiclassical analysis, allows us to treat the *variable* curvature case as well.

Theorem 1.3.1 agrees with a "folklore" conjecture that on a generic negatively curved surface

$$R(\lambda) = O(\lambda^{\varepsilon}) \quad \forall\, \varepsilon > 0. \tag{1.3.4}$$

Genericity is important, since on arithmetic surfaces corresponding to quaternionic lattices one can prove a much better lower bound $R(\lambda) = \Omega\left(\frac{\sqrt{\lambda}}{\ln\lambda}\right)$ (see [Hej]).

Remark 1.3.5. One may compare Theorem 1.3.1 with the lower bound for the *pointwise* error term obtained in [J-P]:

$$R_x(\lambda) = \sum_{\sqrt{\lambda_i}\leq\lambda} |\phi_i(x)|^2 - \frac{\lambda^2}{4\pi} = \Omega\left(\sqrt{\lambda}\,(\ln\lambda)^{\frac{P(-\mathcal{H}/2)}{h}-\varepsilon}\right) \quad \forall\, \varepsilon > 0. \tag{1.3.6}$$

Estimates (1.3.2) and (1.3.6) are independent, and, in particular, (1.3.2) can not be deduced from (1.3.6). Indeed, cancellations may occur when $R_x(\lambda)$ is integrated over a negatively curved surface (for instance, according to the conjecture (1.3.4), $\sqrt{\lambda}$ should cancel out in the generic case). Also, the sequence of λ–s yielding the Ω–bound (1.3.6) depends on the lengths of the geodesic loops at x (see [J-P, section 5.2]), and hence for each point x such a sequence is apriori different.

1.4. **Wave trace asymptotics for long times.** Consider the spectral distribution

$$\mathrm{Tr}\ e(t) \ = \ \sum_{i=0}^{\infty} \cos(\sqrt{\lambda_i} t) \tag{1.4.1}$$

which is the even part of the wave trace on X.

To prove Theorem 1.3.1 we use a modification of the Duistermaat-Guillemin asymptotic formula for the wave trace [D-G]. Originally, this formula captures the contribution of a single closed geodesic, however for the proof of Theorem 1.3.1 we need to take into account the contributions of *all* closed geodesics of length $T_0 < L \le T(\lambda)$, where $T_0 > 0$ is some constant (see Lemma 2.2.1), and $T(\lambda) \to \infty$ at an appropriate rate as the spectral parameter $\lambda \to \infty$.

Let $\chi(t, T)$ be a cut-off function

$$\chi(t, T) = (1 - \psi(t))\, \hat{\rho}\left(\frac{t}{T}\right), \tag{1.4.2}$$

where $\rho \in \mathcal{S}(\mathbb{R})$ is an even, non-negative Schwartz function such that $\operatorname{supp} \hat{\rho} \subset [-1, +1]$, and $\psi(t) \in C_0^\infty(\mathbb{R})$ with $\psi(t) \equiv 1$ when $t \in [-T_0, T_0]$ and $\psi(t) \equiv 0$ when $|t| \ge 2T_0$.

The long-time version of the Duistermaat-Guillemin trace formula is given by

Theorem 1.4.3. *Let $T(\lambda) \to \infty$ as $\lambda \to \infty$ with $T(\lambda) \le \epsilon \ln \lambda$ for $\epsilon > 0$ small enough, and let L_γ (resp. $L_\gamma^\sharp$) denote the length (resp. the primitive period) of a periodic geodesic γ. Then the asymptotics of the smoothed Fourier transform of the wave trace is given by*

$$\int_{-\infty}^{\infty} \mathrm{Tr}\ e(t) \chi(t, T) \cos \lambda t\, dt = \sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T(\lambda)} \frac{L_\gamma^\sharp \cos(\lambda L_\gamma)\, \chi(L_\gamma, T)}{\sqrt{|\det(I - \mathcal{P}_\gamma)|}} + \mathcal{O}(1)\,. \tag{1.4.4}$$

Here, $\mathcal{P}_\gamma$ is the linearized Poincaré map corresponding to γ.

The proof of Theorem 1.4.3 is the most technically difficult part of the paper. We use semiclassical microlocal analysis on small scales, see section 3. In order to avoid the "accumulation" of singularities in the wave trace and to make the stationary phase method work, we separate the contributions of each closed geodesic using Lemma 2.2.1. Although the idea of working to (suitably scaled) Ehrenfest times $T(\lambda) \sim \epsilon \ln \lambda$ is well-established (see [Ber, Zel2, Fa]), rigorously separating out large exponential sums with $T(\lambda)$-terms from the wave-trace using small-scale $\hbar$-microlocalization appears to be a novel approach to estimating remainders in the negatively-curved case.

The main term in the asymptotics proved in Theorem 1.4.3 is given by the sum of the principal wave invariants at each closed geodesic. As follows from Lemma 2.1.6, to prove Theorem 1.3.1 it is sufficient to show that this sum grows at the rate given by (1.3.2). First, we use the Parry-Pollicott equidistribution result [P-P] to calculate the asymptotics of the sum not taking into account its oscillatory nature. Then we deal with the oscillations in the wave invariants (the difficulty is that oscillating terms may cancel out in the sum) using a "straightening the phases" argument based on the Dirichlet box principle (cf. [J-P]).

1.5. **Lower bound for $R(\lambda)$ in higher dimensions.** Lower bounds for the error term in Weyl's law on higher-dimensional manifolds are in fact much simpler than on surfaces. The reason is that the contribution from the "singularity at zero" to the remainder dominates the contribution from the periodic geodesics.

We write $f(\lambda) \gg \lambda^k$ for a function $f(\lambda)$ if there exist a constant $c_0 > 0$ and a number λ_0, such that $f(\lambda) > c_0\lambda^k$ for any $\lambda > \lambda_0$. The following L^1-estimate (which is stronger than an Ω-bound) holds for the remainder in Weyl's law on a rather general class of Riemannian manifolds containing manifolds of negative curvature.

Theorem 1.5.1. *Let X be a manifold of dimension $n \geq 3$, such that $\int_X \tau \neq 0$, where τ is the scalar curvature. Then,*

$$\frac{1}{\lambda}\int_0^\lambda |R(t)|dt \gg \lambda^{n-2}, \tag{1.5.2}$$

Theorem 1.5.1 is proved using the asymptotics of the Riesz means (see [Saf]) in section 5.1. One can also prove Theorem 1.5.1 using the standard $t \to 0^+$ heat trace asymptotics, see [J-P, §2.1]. In order to get more refined information about $R(\lambda)$ on negatively curved manifolds of higher dimension it is natural to study the *oscillatory* error term, $R^{osc}(\lambda)$, see section 5.2.

2. Two auxiliary lemmas

2.1. **Smoothed Fourier transform of the wave trace.** In the notations of section 1.4 let

$$k(\lambda, T) = \int_{-\infty}^{\infty} \frac{\hat\rho(t/T)}{T}\cos(\lambda t)\,\mathrm{Tr}\; e(t)dt, \tag{2.1.1}$$

where, $T = T(\lambda)$ will be chosen appropriately later on. Substituting (1.4.1) into (2.1.1) we obtain

$$k(\lambda, T) = \sum_{i=0}^{\infty} H_{\lambda,T}(\sqrt{\lambda_i}), \tag{2.1.2}$$

where, for $r \geq 0$,

$$H_{\lambda,T}(r) = \int_{-\infty}^{\infty} \frac{\hat\rho(t/T)}{T}\cos(\lambda t)\cos(rt)dt = \frac{\rho(T(\lambda - r)) + \rho(T(\lambda + r))}{2} = \frac{\rho(T(\lambda - r))}{2} + \mathcal{O}(\lambda^{-\infty}), \tag{2.1.3}$$

Here,

$$\widehat{\rho}(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-is\zeta}\rho(\zeta)d\zeta$$

is the Fourier transform of ρ and the $\mathcal{O}(\lambda^{-\infty})$-error in (2.1.3) follows from the fact that $\rho(\lambda + r) = \mathcal{O}(\lambda^{-\infty})$ uniformly for $r \geq 0$, since $\rho \in \mathcal{S}(\mathbb{R})$. Replacing the sum in (2.1.2) by an integral, we get the following representation of $k(\lambda, T)$:

$$k(\lambda, T) = \int_0^{\infty} \frac{\rho(T(\lambda - r)) + \rho(T(\lambda + r))}{2}dN(r) = \int_0^{\infty} H_{\lambda,T}(r)dN(r) \tag{2.1.4}$$

Formula (2.1.4) plays a key role in our analysis. We shall also use the following notation:

$$\int_0^\infty H_{\lambda,T}(r)dR(r) = \kappa(\lambda, T) \tag{2.1.5}$$

Note that the contribution of the main term in Weyl's law has been subtracted from $k(\lambda, T)$ to get $\kappa(\lambda, T)$.

We shall use the following:

Lemma 2.1.6. *Let $R(\lambda) = o((\ln \lambda)^b)$, $b > 0$. Then $\kappa(\lambda, T) = o((\ln \lambda)^b)$ uniformly in T for $T > T_0$, where T_0 is an arbitrary positive number.*

Proof. By the assumption of the lemma, for any $\epsilon > 0$, $R(\lambda) < \epsilon(1 + \ln \lambda)^b$ for large enough λ. Consider the left hand side of (2.1.5):

$$\int_0^\infty H_{\lambda,T}(r)dR(r). \tag{2.1.7}$$

Taking into account (2.1.3) and integrating (2.1.7) by parts we obtain

$$\kappa(\lambda, T) \le \frac{\epsilon T}{2}\int_0^\infty |\rho'(T(r-\lambda))|(\ln(1+r))^b dr + \frac{\epsilon T}{2}\int_0^\infty |\rho'(T(r+\lambda))|(\ln(1+r))^b dr. \tag{2.1.8}$$

Since ρ' is Schwartz class, the second term of (2.1.8) is $O(1)$. Changing variables in the first term of (2.1.8), we obtain

$$\frac{\epsilon T}{2}\int_{-\lambda}^\infty |\rho'(Ts)|(\ln(1+\lambda+s))^b ds =$$

$$\frac{\epsilon(\ln\lambda)^b}{2}\int_{-\lambda T}^\infty |\rho'(u)|\left(1+\frac{\ln(1+\frac{u+T}{\lambda T})}{\ln\lambda}\right)^b du \le C\epsilon(\ln\lambda)^b$$

for some constant $C > 0$, where the last inequality again follows from the fact that ρ' is Schwartz class. Clearly, the constant C can be chosen uniformly in T for $T > T_0 > 0$. Since ϵ can be taken arbitrarily small, we get $\kappa(\lambda, T) = o((\ln \lambda)^b)$, and this completes the proof of the lemma. □

Remark 2.1.9. Lemma 2.1.6 is proved similarly to the results of [J-P, section 2.2], see also [K]. One may also compare it to [Sar, Proposition 3.1]. Sarnak's argument gives the lower bound $R(\lambda) = \Omega(\sqrt{\lambda})$ for the Weyl error on a surface under the assumption that the geodesic flow G^t has a fixed point set of dimension two for some $t > 0$. This condition holds, for example, when G^t is completely integrable, but it is not satisfied on a negatively curved surface.

2.2. **Separation of periodic orbits.** In this section we shall prove the following dynamical lemma:

Lemma 2.2.1. *Let X be a negatively curved surface and let $\Omega(\gamma, \varepsilon)$ denote the ε-neighborhood of a geodesic γ in SX with respect to the Sasaki metric. Then there exist positive constants T_0, B and δ' (depending only on the injectivity radius* $\mathrm{inj}(X)$ *and the lower curvature bound K_1) such that for any $T > T_0$ the sets $\Omega(\gamma, e^{-BT})$ are disjoint for all pairs of closed geodesics γ on X with length $L_\gamma \in [T - \delta', T]$.*

Note that since there are exponentially many closed geodesics on X of length $L_\gamma \in [T - \delta', T]$, disjoint neighborhoods have to be of exponentially small size.

Remark 2.2.2. Here and further on we write $\Omega(Y,d)$ for the d-neighborhood of the set Y. It should not be confused with the Ω notation for the lower bounds.

Proof. Choose $B \geq 2K_1$, where K_1^2 is the curvature bound and K_1 is an upper bound for Lyapunov exponents, cf. (4.1.4). Also, choose $\delta' < \mathrm{inj}(X)/3$, and let T_0 be such that $2e^{-K_1T_0} < \delta'$. Assume for contradiction that there exist two closed geodesics γ_1 and γ_2 with $T-\delta' \leq L_{\gamma_1} \leq L_{\gamma_2} \leq T$ such that the corresponding neighborhoods intersect, and that the geodesics are not inverses of each other (note that due to the restriction $L_\gamma \in [T-\delta', T]$ in the conditions of the lemma, the geodesics cannot be integer multiples of each other unless they are inverses).

Denote the geodesics on X by $\gamma_j(t), 0 \leq t \leq L_{\gamma_j}$, and their lifts to SX by $\tilde{\gamma}_j(t) = (\gamma_j(t), \gamma_j'(t))$; $\gamma_j(t) \subseteq X$ is sometimes called a *footprint* of $\tilde{\gamma}_j(t)$. Without loss of generality we may assume that

$$\mathrm{dist}_{SX}(\tilde{\gamma}_2(0), \tilde{\gamma}_1(0)) \leq 2e^{-2K_1T}.$$

Since for any $0 \leq t \leq L_{\gamma_2}$

$$\mathrm{dist}_{SX}(\tilde{\gamma}_2(t), \tilde{\gamma}_1(t)) = \mathrm{dist}_{SX}(G^t\tilde{\gamma}_2(0), G^t\tilde{\gamma}_1(0)) \leq 2e^{-2K_1T}e^{K_1t} \leq 2e^{-K_1T},$$

and hence

$$\mathrm{dist}_X(\gamma_2(t), \gamma_1(t)) \leq 2e^{-K_1T}. \tag{2.2.3}$$

In other words, the entire geodesic γ_2 lies in the $2e^{-K_1T}$-neighborhood of the geodesic γ_1 and vice versa.

For convenience, we reparametrise γ_1 and define

$$\beta_1(s) := \gamma_1(L_1s/L_2), \qquad 0 \leq s \leq L_2.$$

By triangle inequality and the definition of β_1,

$$\begin{aligned} &\mathrm{dist}(\gamma_2(t), \beta_1(t)) \leq \mathrm{dist}(\gamma_2(t), \gamma_1(t)) + \mathrm{dist}(\gamma_1(t), \beta_1(t)) \leq \\ &2e^{-K_1t} + t\left(1 - \frac{L_1}{L_2}\right) \leq 2e^{-K_1T} + \delta' < \frac{2\cdot \mathrm{inj}(X)}{3}. \end{aligned} \tag{2.2.4}$$

Accordingly, for any $0 \leq t \leq L_2$ there exists a unique shortest geodesic $\alpha_t(s)$ in X connecting $\gamma_2(t)$ and $\beta_1(t)$. We shall choose the parameter $s \in [0,1]$ so that $\alpha_t(0) = \gamma_2(t)$ and $\alpha_t(1) = \beta_1(t)$.

Define the mapping $\Phi(t,s) : [0, L_2] \times [0,1] \to X$ by the formula

$$\Phi(t,s) = \alpha_t(s).$$

We claim that Φ defines a homotopy between $\gamma_2(t)$ and $\beta_1(t)$. Indeed, $\Phi(t,0) = \gamma_2(t)$, $\Phi(t,1) = \beta_1(t)$. Moreover, since both γ_2 and β_1 have period L_2, we have

$$\alpha_0(s) = \alpha_{L_2}(s), \ \forall\, s \in [0,1],$$

and so $\Phi(\cdot, s)$ is a closed curve in X. Finally, $\Phi(t,s)$ is continuous since the function $\mathrm{dist}(\gamma_2(t), \beta_1(t))$ is a continuous function of t.

On the other hand, β_1 is just a reparametrization of γ_1, hence γ_1 and γ_2 lie in the same free homotopy class, contradicting the fact that on a negatively-curved surface there is a unique closed geodesic in each free homotopy class. The contradiction completes the proof of the lemma. □

Remark 2.2.5. Lemma 2.2.1 is proved for the tangent bundle. The tangent and the cotangent bundles can be identified using the Riemannian metric, and therefore Lemma 2.2.1 holds in the cotangent bundle as well. In this setting it will be used in the next section.

Remark 2.2.6. The proof of Lemma 2.2.1 generalizes verbatim to higher dimensions.

Remark 2.2.7. In the literature on wave invariants [D-G, Don, Zel1] it is customary to choose a cut-off function in trace formulas in the length spectrum so that only a single length of a closed geodesic is contained in its support. Since we go to $T(\lambda) \sim \ln\ln\lambda$ times, it is not enough to localize exclusively in the length spectrum. Indeed, when we localize around a geodesic(s) of length L_i, error terms in expressions like (3.2.2) are of the order $O(1/|L_{i+1} - L_i|)$, where $\ldots < L_i < L_{i+1} < \ldots$ denote distinct lengths of closed geodesics on X (ignoring the multiplicity). Therefore, the error would be large in the presence of "near-multiplicities". We can not control the gaps in the length spectrum on a generic negatively curved surface, and hence we have to localize in the phase space as it is done in the next section.

3. Wave trace asymptotics and small-scale microlocalization

In this section we give a proof of Theorem 1.4.3 which is quite technical. Let us note that assuming Proposition 3.2.1, the subsequent sections can be read independently of section 3.

3.1. Plan of the argument.

We choose a parameter $T = T(\lambda) > 0$ satisfying:

- (i) $T(\lambda) \to \infty$ as $\lambda \to \infty$
- (ii) $T(\lambda) \leq \epsilon \ln \lambda$.

Here, $\epsilon > 0$ is a small constant that will be chosen later on. For the applications in this paper, it will only be necessary to take $T(\lambda)$ of order $\ln\ln\lambda$, so we will not be concerned with determining the best possible constant $\epsilon > 0$ in (3.1).

To localize the contribution to the wave trace and to $\kappa(\lambda, T)$ from a given closed geodesic γ of length $L_\gamma \leq T$ it is important to microlocalize the wave trace $\operatorname{Tr} e(t)$ to a neighborhood of γ and then sum over all the different γ's. The complication here is that we want to take into account the contributions of *all* closed geodesics γ with $T_0 < L_\gamma \leq T(\lambda)$. Since $T(\lambda) \to \infty$ as $\lambda \to \infty$, the number of these geodesics blows up. As a result, simply summing stationary phase expansions for each of the γ's (which is automatic when $T = \mathcal{O}(1)$) is impossible when one needs to work with such long period intervals. The way to deal with this is to microlocalize on neighborhoods of the γ's that shrink fast enough as $\lambda \to \infty$ (but not too fast) and then split up the time interval $[T_0, T(\lambda)]$ into short "windows" of fixed size $\delta' > 0$. In this context, it is natural to work with semiclassical pseudodifferential and Fourier integral operators. The crucial dynamical result we need here is Lemma 2.2.1. Roughly speaking, this lemma says that there exist neighborhoods $\Omega(\gamma, e^{-BT})$ of γ of size e^{-BT} in phase space T^*X with the property that, for appropriate geometric constant $\delta' > 0$, no other periodic geodesic with period in the window $[L_\gamma - \delta', L_\gamma + \delta']$ intersects $\Omega(\gamma, e^{-BT})$. Since $T(\lambda) \leq \epsilon \ln \lambda$, this clearly suggests microlocalizing the trace to $e^{-B\epsilon \ln \lambda} = \lambda^{-B\epsilon}$-neighborhoods of γ.

3.2. A reformulation of Theorem 1.4.3.

For the purposes of the proof of Theorem 1.3.1, it is convenient for us to reformulate Theorem 1.4.3 in the following equivalent form:

Proposition 3.2.1. *Let $T(\lambda) \to \infty$ as $\lambda \to \infty$ with $T(\lambda) \le \epsilon \ln \lambda$. Then, for $\epsilon > 0$ small enough,*

$$\kappa(\lambda, T(\lambda)) = \sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T(\lambda)} \frac{L_\gamma^\sharp \cos(\lambda L_\gamma)\, \chi\,(L_\gamma, T(\lambda))}{T(\lambda)\sqrt{|\det(I - \mathcal{P}_\gamma)|}} + \mathcal{O}\left(\frac{1}{T(\lambda)}\right). \tag{3.2.2}$$

Here, $\mathcal{P}_\gamma$ is the linearized Poincaré map corresponding to γ and $\chi(t, T)$ is a cut-off function defined by (1.4.2)*.*

The proof of Proposition 3.2.1 is divided in several steps that are carried out in sections 3.4–3.8.

Equivalence of Theorem 1.4.3 and Proposition 3.2.1 immediately follows from the following

Lemma 3.2.3. *Let $T(\lambda) \to \infty$ as $\lambda \to \infty$. Then*

$$\kappa(\lambda, T) = \frac{1}{T(\lambda)} \int_{-\infty}^{\infty} \mathrm{Tr}\ e(t)\, \chi(t, T) \cos(\lambda t)\, dt + \mathcal{O}\left(\frac{1}{T(\lambda)}\right). \tag{3.2.4}$$

Proof. Combining (2.1.1), (2.1.4), (2.1.5) and using that ρ is a Schwartz function we get

$$\begin{aligned} \kappa(\lambda, T) = & \int_{-\infty}^{\infty} \frac{\hat\rho(t/T)}{T} \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt - \\ & \frac{A}{4\pi} \int_0^\infty r(\rho(T(\lambda - r)) + \rho(T(\lambda + r)))dr = \\ & \int_{-\infty}^{\infty} \frac{\hat\rho(t/T)}{T} \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt - \frac{A}{4\pi} \int_0^\infty r\rho(T(\lambda - r))dr + \mathcal{O}((\lambda T)^{-\infty}) = \\ & \int_{-\infty}^{\infty} \frac{\hat\rho(t/T)}{T} \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt - \frac{A}{2T}\lambda\hat\rho(0) + \mathcal{O}\left(\frac{1}{T^2}\right) + \mathcal{O}((\lambda T)^{-\infty}). \end{aligned} \tag{3.2.5}$$

Therefore, one may rewrite $\kappa(\lambda, T)$ as

$$\begin{aligned} \kappa(\lambda, T) = & \frac{1}{T} \int_{-\infty}^{\infty} \hat\rho(t/T)\psi(t) \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt + \\ & \frac{1}{T} \int_{-\infty}^{\infty} \hat\rho(t/T)(1 - \psi(t)) \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt - \frac{A}{2T}\lambda\hat\rho(0) + \mathcal{O}\left(\frac{1}{T^2}\right) \end{aligned} \tag{3.2.6}$$

Consider now the first term on the right-hand side of (3.2.6):

$$\frac{1}{T} \int_{-\infty}^{\infty} \hat\rho(t/T)\psi(t) \cos(\lambda t)\, \mathrm{Tr}\ e(t) dt \tag{3.2.7}$$

It follows from the trace formula [D-G] that the contributions to (3.2.7) from the non-trivial periods $t = L_\gamma \ne 0$ with $0 < L_\gamma < T_0$ are $\mathcal{O}(1/T)$. At the same time, the wave trace at $t = 0$ has the singularity expansion [Zel1]: $\mathrm{Tr}\, e(t) = -\frac{1}{2\pi} t^{-2} + a_1 t^{-1} + a_2 + \dots$, where the leading coefficient can be computed by integrating the principal on-diagonal term of the parametrix for the wave kernel, see [J-P, section 3.1].

Taking the contribution of singularity at zero into account, we obtain that (3.2.7) can be represented as

$$\frac{A}{2T}\hat\rho(0)\lambda + \mathcal{O}\left(\frac{1}{T}\right). \tag{3.2.8}$$

Therefore, the leading term in (3.2.7) cancels the λ-term in (3.2.6) (cf. [J-P, Lemma 3.2.1] where a local analogue of such a result is established). To complete the proof of the lemma we note that by definition $\chi(t,T)=\hat{\rho}(t/T)(1-\psi(t))$. □

3.3. **Preliminaries and notations.** We now briefly recall the calculus of small-scale $\hbar$-pseudodifferential operators [D-S, Sj] that will be needed to carry out the various microlocalizations. In the following we use the notation $\hbar=\lambda^{-1}$.

Given $0\leq\delta<\frac{1}{2}$, we say that $a(x,\xi;\hbar)\in C_0^\infty(T^*X\times(0,\hbar_0])$ is in the symbol class $S_\delta^m(T^*X)$ provided

$$|\partial_x^\alpha\partial_\xi^\beta a(x,\xi;\hbar)|\leq C_{\alpha,\beta}\hbar^{-m-\delta(|\alpha|+|\beta|)}.$$

Given $a\in S_\delta^m(T^*X)$, one can define the corresponding $\hbar$-pseudodifferential operator $Op_\hbar(a)$ invariantly in terms of the Schwartz kernel:

$$Op_\hbar(a)(x,y)=(2\pi\hbar)^{-n}\int_{T_x^*X}e^{-i\exp_x^{-1}(y)\cdot\xi/\hbar}\,a(x,\xi;\hbar)\,\zeta(r^2(x,y))d\xi. \tag{3.3.1}$$

Here, $\exp:T_xX\to X$ is the geodesic exponential map, $\xi\in T_x^*X$, $r(x,y)$ is geodesic distance between x and y and $\zeta\in C_0^\infty(\mathbb{R})$ is supported in a ball $B_\epsilon(0)$ and equal to 1 in $B_{\epsilon/2}(0)$ with $\epsilon>0$ sufficiently small (one can take here $\epsilon<\mathrm{inj}(X,g)$. Such operators form a calculus with $Op_\hbar(S_\delta^m)\circ Op_\hbar(S_\delta^{m'})\subset Op_\hbar(S_\delta^{m+m'})$. Calderon-Vaillancourt L^2-boundedness and the $\hbar$-Egorov theorem also hold [Sj, section 2]. Moreover, let $x,y\in\mathbb{R}^n$ be the coordinates of the points $x,y\in X$ in some local coordinate system (here we abuse notation slightly and denote points on the manifold and their coordinates by the same letters). By using the Taylor expansion $-\exp_x^{-1}(y)=x-y+\mathcal{O}(|x-y|^2)$, to make the Kuranishi change of variables $\xi\mapsto(1+\mathcal{O}(x-y))\xi$ in (3.3.1) and integrating by parts in ξ, one can locally rewrite (3.3.1) in the somewhat more familiar form $Op_\hbar(a)(x,y)=(2\pi\hbar)^{-n}\int_{\mathbb{R}^n}e^{i\langle x-y,\xi\rangle/\hbar}\,a(x,\xi;\hbar)(1+\mathcal{O}(\hbar^{1-2\delta}))\,\zeta(|x-y|^2)d\xi$. However, it is useful here to work with the invariantly defined operators in (3.3.1), and we will do so without further comment.

We will also use the following geometric notations. Let M be the universal cover of X. The fundamental group of X is denoted by Γ. For $\omega\in\Gamma$, let $L_\omega=L_\gamma$ be the length of the unique closed geodesic γ on X that corresponds to the conjugacy class $[\omega]\subset\Gamma$. The lifted bicharacteristic curve in T^*M projecting to γ is denoted by $\tilde{\gamma}\subset T^*M$. The standard cotangent projection map is denoted by $\pi:T^*M\to M$. We use analogous notation on T^*X. We will also denote a periodic geodesic on X (resp. T^*X) and a choice of lift on M (resp. T^*M) generally by the same letter when the choice of lift is uniquely specified. Similarly, functions in $C^\infty(T^*X)$ will be identified with their lifts to $C^\infty(T^*M)$.

3.4. **Hadamard parametrix.** One needs to get an asymptotic exponential sum formula for the leading term in (3.2.4) just like in the standard case where $T(\hbar)\sim 1$. As we have already indicated, we will need to microlocalize the wave trace on small-scales with respect to $\hbar=\lambda^{-1}$, so it's useful to introduce the parameter $\hbar$ in the lifted wave operator. We do this by writing the real part of the wave operator on M in the form $\tilde{e}(t)=\cos t\sqrt{\tilde{\Delta}}=\cos t\left(\frac{\hbar\sqrt{\tilde{\Delta}}}{\hbar}\right)$. Here we denote by $\tilde{\Delta}$ the Laplacian on M and by $\tilde{e}(t)$ the real part of the wave operator on M.

Even though introducing $\hbar$ amounts to simply rescaling the Hadamard parametrix approximations $\tilde{E}_N(t)$ to $\tilde{e}(t)$ (see 3.4.1 below), it's useful to think of $\tilde{E}_N(x,y,t;\hbar)$; as the Schwartz kernels of a family of $\hbar$-Fourier integral operators (albeit, with trivial dependence on $\hbar$) where $\hbar\sqrt{\tilde{\Delta}}$ is a classical $\hbar$-pseudodifferential operator [D-S, chapter 7]. Since $\mathrm{inj}(M)=\infty$, the Hadamard parametrix approximation $\tilde{E}_N(t,x,y;\hbar)$ is valid for *all* $t>0$ and is given by the following well-known formula: [Ber]:

$$\tilde{E}_N(t,x,y;\hbar) = \hbar^{-\frac{1}{2}}\frac{\partial}{\partial t}\int_0^\infty e^{i\theta[r^2(x,y)-t^2]/\hbar}\left(\sum_{k=0}^N a_k(x,y)\theta_+^{-k-1/2}\,\hbar^k\right)d\theta, \tag{3.4.1}$$

In the last identity (3.4.1), ∂_t is understood in the distributional sense. Also, for $\alpha\in\mathbb{R}$, $\theta_\pm^\alpha$ denote the standard homogeneous distributions [Ho2, section 3.2] and $a_k\in C^\infty(M\times M)$; $k=0,1,2,....$ It is also well-known (see [Ber, section 39]) that for fixed $y\in M$, $\tilde{U}-\tilde{U}_N\in C^N(M\times\mathbb{R}^+)$ and moreover, for any compact $Y\subset M$, there exists $C_1,C_2>0$ such that

$$\sup_{y\in Y}|\tilde{e}(t,x,y;\hbar)-\tilde{E}_N(t,x,y;\hbar)|\le C_2e^{C_1|t|}\,\hbar^N. \tag{3.4.2}$$

In (3.4.2) the bound is uniform for $(x,y)\in M\times Y$ with $d(x,y)\le d_0$, where d_0 is any constant, and $C_j=C_j(d_0,Y,N)$; $j=1,2$. Then [Ber, CdV], one can take

$$E_N(t,x,y;\hbar)=\sum_{\omega\in\Gamma;L_\omega\le T(\hbar)}\tilde{E}_N(t,x,\omega y;\hbar), \tag{3.4.3}$$

to be the parametrix approximation to the wave operator $e(t)=\cos t\sqrt{\Delta}$ on X.

Moreover, there exists an appropriate cutoff function $\eta\in C_0^\infty(M;[0,1])$ (see [CdV, p. 94], [Ber, Lemma 34]) with $\mathrm{diam}(\mathrm{supp}\,\eta)\le 2\,\mathrm{diam}(X)+1$ so that

$$\mathrm{Tr}\,e(t)=\int_M E_N(t,x,x;\hbar)\,\eta(x)\,dvol(x)+\mathcal{O}(e^{C|t|}\,\hbar^N), \tag{3.4.4}$$

for an appropriate $C>0$.

From now on we put

$$E_{N,\omega}(t,x,y;\hbar):=\tilde{E}_N(t,x,\omega y;\hbar). \tag{3.4.5}$$

and so,

$$E_N(t,x,y;\hbar):=\sum_{\omega\in\Gamma;L_\omega\le T(\hbar)}E_{N,\omega}(t,x,y;\hbar). \tag{3.4.6}$$

Then, since $|t|\le\epsilon|\ln\hbar|$, the RHS in (3.4.2) is $\mathcal{O}(\hbar^{N(1-C_1\epsilon)})$ and so, by choosing $\epsilon<C_1^{-1}$ and N large enough, it follows from Lemma 3.2.3 that

$$\begin{aligned}\kappa(\lambda,T)=&\frac{1}{2T(\hbar)}\sum_{id\neq\omega\in\Gamma}\int_{-\infty}^\infty\int_M E_{N,\omega}(t,x,x;\hbar)\,\eta(x)\,\chi(t;T(\hbar))\,e^{-it/\hbar}\,dvol(x)dt\\&+\frac{1}{2T(\hbar)}\sum_{id\neq\omega\in\Gamma}\int_{-\infty}^\infty\int_M E_{N,\omega}(t,x,x;\hbar)\,\eta(x)\,\chi(t;T(\hbar))\,e^{it/\hbar}\,dvol(x)dt\\&+\mathcal{O}(T(\hbar)^{-1}).\end{aligned} \tag{3.4.7}$$

Remark 3.4.8. Note that here and further on the sum over $\omega\in\Gamma$ is finite due to the presence of the cut-off function $\chi(t;T(\hbar))$: the summation is taken over elements $\omega\neq id$ that such that $L_\omega\le T(\hbar)$.

The integral in (3.4.7) with the the $e^{\pm it/\hbar}$ appearing has a total phase function:

$$\phi^{\pm}(t,\theta;x,y) = (r^2(x,y) - t^2)\theta \pm t.$$

Since $\theta \geq 0$,

$$\partial_t \phi^{\pm}(t,\theta;x,y) = -2t\theta \pm 1,$$

and for any $\alpha \geq 0$,

$$\partial_t^{\alpha} \chi(t;T(\hbar)) = \mathcal{O}_{\alpha}(1),$$

after first integrating by parts in θ for the $k \geq 1$ terms in (3.4.1), followed by repeated integration by parts in t one gets that

$$\begin{aligned}\kappa(\lambda,T) = \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \int_M E^{+}_{N,\omega}(t,x,x;\hbar)\ \chi(t;T(\hbar))\, e^{it/\hbar}\, dvol(x) dt \\ + \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \int_M E^{-}_{N,\omega}(t,x,x;\hbar)\ \chi(t;T(\hbar))\, e^{-it/\hbar}\, dvol(x) dt \\ + \mathcal{O}(T(\hbar)^{-1}),\end{aligned} \tag{3.4.9}$$

where, $E^{\pm}_{N,\omega}(t,x,y;\hbar)$ is defined to be

(3.4.10)

$$\hbar^{-\frac{1}{2}} \eta(x) \frac{\partial}{\partial t} \int_0^{\infty} e^{i\theta[r^2(x,\omega y) - t^2]/\hbar} \left(\sum_{k=0}^{N} a_k(x,\omega y)\, \theta_+^{-k-1/2}\, \hbar^k \right) \psi(-2t\theta \pm 1)\, d\theta.$$

From now on, we denote both the Schwartz kernel and the associated operator by $E^{\pm}_{N,\omega}$. We note that the error term on the RHS of (3.4.9) is the sum of the $\mathcal{O}(T(\hbar)^{-1})$-error in (3.4.7) and an $\mathcal{O}(\hbar^{\infty})$-piece that arises from inserting the cutoff function $\psi(-2t\theta \pm 1)$. To see this, split the θ-integration into the regions where $|\theta| \leq 1$ and $|\theta| \geq 1$. Integrating by parts in t over the first region contributes an $\mathcal{O}(\hbar^{\infty})$-error in (3.4.9). The same thing happens over the region where $|\theta| \geq 1$ by noting that on this set

$$|-2t\theta \pm 1|^{-1} \leq |\theta|^{-1}$$

and that $T(\hbar) = O(|\ln \hbar|)$.

The variables (t,θ) are roughly-speaking dual and the presence of the cutoffs $\psi(-2t\theta \pm 1)$ will be exploited later on when we will need to interchange orders of integration in the expression for $\kappa(\lambda,T)$. The analysis for each of the two integral sums in (3.4.9) is the same and amounts to time-reversal. We will denote the first sum in (3.4.9) by $\kappa^{+}(\lambda,T)$ and the second by $\kappa^{-}(\lambda,T)$.

The next step is to microlocalize on shrinking $\hbar$-scales near individual lifts of periodic geodesics. First, we microlocalize on shrinking scales near S^*X.

3.5. **Small-scale microlocalization near S^*X.** We will need to exploit the separation of geodesics in phase space given by the dynamical Lemma 2.2.1. Since the latter result applies to unit speed geodesics on S^*X, we begin by $\hbar$-microlocalizing the trace $\kappa(\lambda,T)$ near the co-sphere bundle S^*X on small-scales $\sim \hbar^{\delta}$; $0 < \delta < 1/2$. The argument here is quite standard, but for completeness, we sketch the proof.

Let $d : T^*X \times T^*X \to \mathbb{R}$ be the distance function defined by:

$$d_{T^*X}((x_1,\xi_1),(x_2,\xi_2)) = d_{TX}((x_1,\xi_1^*),(x_2,\xi_2^*)),$$

where ξ_1^*, ξ_2^* are dual vectors to ξ_1, ξ_2 and d_{TX} is the Sasaki distance on TX. Lemma 2.2.1 extends in a straightforward way to the cotangent bundle endowed with such a distance function, see Remark 2.2.5.

Let $\zeta \in C_0^\infty(\mathbb{R})$ be a cutoff function with $\zeta(x) = 1$ for x near 0 and satisfying $\operatorname{supp}\zeta \subset [-1,1]$. For $0 < \delta < 1/2$ we define

$$\chi^\delta(x,\xi;\hbar) = \zeta(\hbar^{-2\delta} d^2((x,\xi), S^*X)\,).$$

Then, since $\rho \in S(\mathbb{R})$, it follows that, modulo $O(\lambda^{-\infty})$-errors, in $\kappa(\lambda, T)$ one can sum over only the eigenvalues λ_j satisfying $|\sqrt{\lambda_j} - \lambda| \leq \lambda^\epsilon$ for any $\epsilon > 0$. Write $\hbar = \lambda^{-1}$ and from now on, consider only such eigenfunctions, $\phi_j; j = 1, 2, ...$ Since $(\hbar\sqrt{\Delta} - 1)\phi_j = \lambda_j(\hbar)\phi_j$; $j = 1,2,3,...$, where $|\lambda_j(\hbar) - 1| = \mathcal{O}_\epsilon(\hbar^{1-\epsilon})$ and the $\hbar$-pseudodifferential operator $P(\hbar) = \hbar\sqrt{\Delta} - 1$ is $\hbar$-microlocally elliptic off $S^*X = \{(x,\xi) \in T^*X; |\xi|_g = 1\}$, by a parametrix construction in the calculus $Op_\hbar(S^*_\delta)$, it follows that

$$\|Op_\hbar(1 - \chi^\delta(x,\xi;\hbar))\phi_j\|_{L^2} = \mathcal{O}(\hbar^\infty). \tag{3.5.1}$$

Since $\sum_j \rho(T(\lambda)(\lambda_j + \lambda)) = \mathcal{O}(\lambda^{-\infty})$, it then follows from (2.1.4) and (3.5.1) that

$$\begin{aligned}
\kappa(\lambda, T) &= \frac{1}{2}\sum_j \rho(T(\lambda)(\lambda_j - \lambda)) - \frac{A\lambda}{2T}\hat{\rho}(0) + \mathcal{O}(T^{-2}) \\
&= \frac{1}{2}\sum_j \rho(T(\lambda)(\lambda_j - \lambda)) \cdot \langle Op_\hbar(\chi^\delta)\phi_j, \phi_j\rangle - \frac{A\lambda}{2T}\hat{\rho}(0) + \mathcal{O}(T^{-2}). \qquad (3.5.2)
\end{aligned}$$

In analogy with the construction of the small-cutoff function $\chi^\delta \in C_0^\infty(T^*X)$ we choose $\zeta \in C_0^\infty(\mathbb{R})$ as above with $\zeta(u) = 1$ near $u = 0$. We define $\tilde{\chi}^\delta(x,\xi;\hbar) = \zeta(\hbar^{-2\delta}d^2((x,\xi), S^*M)) \cdot \tilde{\chi}_R(x)$, where $\tilde{\chi}_R \in C_0^\infty(M)$ with $\tilde{\chi}_R = 1$ on $\operatorname{supp}(\eta)$. Then, from Lemma 3.2.3 and (3.5.2) it follows that:

$$\begin{aligned}
\kappa^\pm(\lambda, T) = \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \int_M &[\, Op_\hbar^*(\tilde{\chi}^\delta) \cdot E^\pm_{N,\omega} \cdot Op_\hbar(\tilde{\chi}^\delta)\,](x,x,t;\hbar) \\
&\times e^{\pm it/\hbar}\chi(t; T(\hbar))\, dvol(x)\, dt + \mathcal{O}(T(\hbar)^{-1}). \qquad (3.5.3)
\end{aligned}$$

From now on, the cutoff $\tilde{\chi}^\delta$ will be included in all computations and for each $\omega \in \Gamma$, the operators $E^\pm_{N,\omega} : C_0^\infty(M) \to C_0^\infty(M)$ will be replaced by the microlocalizations $Op_\hbar^*(\tilde{\chi}^\delta) \cdot E^\pm_{N,\omega} \cdot Op_\hbar(\tilde{\chi}^\delta) : C_0^\infty(M) \to C_0^\infty(M)$. To simplify the writing, we will continue to denote the latter microlocalized operators simply by $E^\pm_{N,\omega}$.

3.6. **Small-scale microlocalization near periodic geodesics.** The second part of the small-scale $\hbar$-microlocalization involves microlocalizing on shrinking $\hbar$-scales near the lifts to T^*M of individual periodic geodesics in T^*X. In light of the previous section it suffices to take $\gamma \subset \Omega(S^*M, \hbar^\delta)$ for any fixed $\delta \in (0, 1/2)$.

From Lemma 2.2.1, it follows that one can put *disjoint* tubular neighborhoods $\Omega(\gamma, e^{-BT(\hbar)})$ around all lifts of periodic geodesics γ with periods in the time windows $[t, t+\delta']; T_0 \leq t \leq T(\hbar)$. The constant $B > 0$ is *uniform* and depends only on the curvature pinching condition (1.1.1) .

Fix the lift of a periodic geodesic γ_0 on M and choose ϵ so that $\delta := B\epsilon < \frac{1}{2}$. Then, for $(x,\xi) \in T^*M$, we define the small-scale cutoff functions

$$\zeta^\delta_{\gamma_0}(x,\xi;\hbar) = \zeta\left(\hbar^{-2\delta}\, d^2((x,\xi), \gamma_0)\right). \tag{3.6.1}$$

Clearly, $\zeta^{\delta}_{\gamma_0} \in S^0_{\delta}(T^*M)$, where we have arranged that $\delta < 1/2$. Then, for any $\omega \in \Gamma$, we write

$$E^{\pm}_{N,\omega} = Op^*_{\hbar}(\zeta^{\delta}_{\gamma_0}) \cdot E^{\pm}_{N,\omega} \cdot Op_{\hbar}(\zeta^{\delta}_{\gamma_0}) + Op^*_{\hbar}(\zeta^{\delta}_{\gamma_0}) \cdot E^{\pm}_{N,\omega} \cdot Op_{\hbar}(1-\zeta^{\delta}_{\gamma_0})$$
$$+ Op^*_{\hbar}(1-\zeta^{\delta}_{\gamma_0}) E^{\pm}_{N,\omega} \cdot Op_{\hbar}(\zeta^{\delta}_{\gamma_0}) + Op^*_{\hbar}(1-\zeta^{\delta}_{\gamma_0}) \cdot E^{\pm}_{N,\omega} \cdot Op_{\hbar}(1-\zeta^{\delta}_{\gamma_0}),$$

and taking traces of both sides, we split the RHS of (3.4.7) into three different integral sums. Taking into account that $\mathrm{Tr}\, A = TrA^*$ and $\mathrm{Tr}\, AB = \mathrm{Tr}\, BA$, it follows that

$$\kappa^{\pm}(\lambda;T) = \kappa^{\pm}_{11}(\lambda;T) + 2\kappa^{\pm}_{12}(\lambda;T) + \kappa_{22}(\lambda;T) + O(T(\hbar)), \tag{3.6.2}$$

where,

(3.6.3)
$$\kappa^{\pm}_{11}(\lambda;T) = \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \mathrm{Tr}[E^{\pm}_{N,\omega} \cdot Op_{\hbar}(\zeta^{\delta}_{\gamma_0}) \cdot Op^*_{\hbar}(\zeta^{\delta}_{\gamma_0})]\, e^{\pm it/\hbar} \chi(t;T(\hbar))\, dt,$$

(3.6.4)
$$\kappa^{\pm}_{12}(\lambda;T) = \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \mathrm{Tr}[E^{\pm}_{N,\omega} \cdot Op_{\hbar}(\zeta^{\delta}_{\gamma_0}) \cdot Op^*_{\hbar}(1-\zeta^{\delta}_{\gamma_0})]\, e^{\pm it/\hbar} \chi(t;T(\hbar))\, dt,$$

and

(3.6.5)
$$\kappa^{\pm}_{22}(\lambda;T) = \frac{1}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_{-\infty}^{\infty} \mathrm{Tr}[E^{\pm}_{N,\omega} \cdot Op_{\hbar}(1-\zeta^{\delta}_{\gamma_0}) \cdot Op^*_{\hbar}(1-\zeta^{\delta}_{\gamma_0})]\, e^{\pm it/\hbar} \chi(t;T(\hbar))\, dt.$$

We now estimate each of the integral sums in (3.6.3)-(3.6.5) separately. Roughly speaking, one should think of the decomposition in (3.6.3)-(3.6.5) as follows: (3.6.3) gives the microlocal contribution of single periodic geodesic γ_0 to the trace, (3.6.4) consists of cross-terms which we will show are $\mathcal{O}(\hbar^{\infty})$ and finally, (3.6.5) is estimated in the same way as (3.6.3) by successively microlocalizing around all other periodic geodesics $\gamma \neq \gamma_0$ with $L_{\gamma} \leq T(\hbar)$. Due to the small-scale microlocalizations and ultimately, the splitting of the time scale into short time-windows, the expansions in $\kappa(\lambda, T)$ are no longer classical polyhomogeneous in $\hbar$. For this reason, it is necessary to give a somewhat different argument than in the classical $T(\hbar) = \mathcal{O}(1)$ case (see [Don]).

3.7. **Studying $\kappa^{\pm}_{11}(\lambda, T)$ in normal coordinates.**

The goal of this section is to prove the formula (3.7.8). For $(x, y) \in \mathrm{supp}(\eta) \times \mathrm{supp}(\eta) \subset M \times M$, we have

$$E^{\pm}_{N,\omega} \cdot Op_{\hbar}(\zeta^{\delta}_{\gamma_0}) \cdot Op^*_{\hbar}(\zeta^{\delta}_{\gamma_0})(x,y;\hbar) = \frac{\partial}{\partial t} I^{\pm}_{N,\omega}(x,y,t;\omega,\hbar). \tag{3.7.1}$$

Here

$$I^{\pm}_{N,\omega}(x,y,t;\omega,\hbar) = (2\pi\hbar)^{-2} \int\int\int e^{i[-t^2\theta + r^2(x,\omega z)\theta - \exp_z^{-1}(y)\cdot\xi]/\hbar} a_N(x,z,\theta;\omega,\hbar)$$
$$\times\, \eta(x)\, \tilde{\zeta}^{\delta}_{\gamma_0}(z,\xi;\hbar)\, \psi(-2t\theta \pm 1)\, \zeta(r^2(z,y))\, dz d\xi d\theta,$$

where $a_N(x,z,\theta;\omega,\hbar) = \hbar^{-\frac{1}{2}} \theta_+^{-\frac{1}{2}} \sum_{k=0}^{N} a_k(x,\omega z) \hbar^k \theta_+^{-k}$. By symbolic calculus for $\hbar$-pseudodifferential operators, $\tilde{\zeta}^{\delta}_{\gamma_0} \in S^0_{\delta}$ satisfies ([D-S, p. 78])

$$\tilde{\zeta}^{\delta}_{\gamma_0}(z,\xi;\hbar) = |\zeta^{\delta}_{\gamma_0}(z,\xi)|^2 + \hbar^{1-2\delta} \zeta^{\delta,-1}_{\gamma_0}(z,\xi;\hbar) \tag{3.7.2}$$

with $|\zeta^{\delta}_{\gamma_0}|^2, \zeta^{\delta,-1}_{\gamma_0} \in S^0_{\delta}$. Since $\mathrm{inj}(M,g)=\infty$, to simplify the writing somewhat, we put here $\zeta(r^2(z,y))=1$ in the $\hbar$-pseudodifferential cutoffs (see (3.3.1)).

Since $\omega \in \Gamma$ acts by by isometries on (M,g), it follows that $\exp^{-1}_{\omega z}(\omega y)\cdot(d\omega^{-1})^t\xi = \exp_z^{-1}(y)\cdot\xi$. and so the expression for $I^{\pm}_{N,\omega}(x,y,t;\omega,\hbar)$ above can be rewritten as

$$(2\pi\hbar)^{-2}\int\int\int e^{i[-t^2\theta + r^2(x,\omega z)\theta - \exp^{-1}_{\omega z}(\omega y)\cdot(d\omega^{-1})^t\xi]/\hbar} a_N(x,z,\theta;\omega,\hbar)$$
$$\times\, \eta(x)\, \tilde{\zeta}^{\delta}_{\gamma_0}(z,\xi;\hbar)\, \psi(-2t\theta \pm 1)\, dz d\xi d\theta,$$

Changing the variables $\omega z \mapsto z$, $(d\omega^{-1})^t\xi \mapsto \xi$ and using that $\omega \in \Gamma$ acts by isometries, one gets

$$I^{\pm}_{N,\omega}(x,y,t;\omega,\hbar) = (2\pi\hbar)^{-2}\int\int\int e^{i[-t^2\theta + r^2(x,z)\theta - \exp_z^{-1}(\omega y)\cdot\xi]/\hbar} a_N(x,z,\theta;\hbar)$$
$$\times\, \eta(x)\, \tilde{\zeta}^{\delta}_{\gamma_0}(\omega^{-1}z,(d\omega)^t\xi;\hbar)\, \psi(-2t\theta \pm 1)\, dz d\xi d\theta. \tag{3.7.3}$$

Here, ω has been scaled out of the amplitude a_N so that $a_N(x,z,\theta;\hbar) = \hbar^{-\frac{1}{2}}\theta_+^{-\frac{1}{2}}\cdot \sum_{k=0}^N a_k(x,z)\hbar^k\theta_+^{-k}$.

Let us now fix a global normal coordinate system centered at $x \in M$. With some abuse of notation we identify a point $z \in M$ and a vector of its coordinates in this system: $z=(z_1,z_2) \in T_xM \cong \mathbb{R}^2$. For instance, in the following $z-\omega y$ means a vector $(z_1-(\omega y)_1, z_2-(\omega y)_2)$.

Writing the Taylor expansion

$$-\exp_z^{-1}(\omega y)\cdot\xi = (z-\omega y)\cdot(1+\mathcal{O}(z-\omega y))\xi$$

and making the corresponding Kuranishi change of variables $\xi \mapsto \xi(1+\mathcal{O}(z-\omega y))$ in the above integral, we get a coordinate expression for (3.7.3):

$$I^{\pm}_{N,\omega}(x,y,t;\omega,\hbar) = (2\pi\hbar)^{-2}\int\int\int e^{i[-t^2\theta + r^2(x,z)\theta + (z-\omega y)\xi]/\hbar} a'_N(x,z,\theta;\omega,\hbar)$$
$$\times\, \eta(x)\, \tilde{\zeta}^{\delta}_{\gamma_0}(\omega^{-1}z,(d\omega)^t\xi;\hbar)\, \psi(-2t\theta \pm 1)\, dz d\xi d\theta, \tag{3.7.4}$$

where $a'_N(x,z,\theta;\omega,\hbar) = a_N(x,z,\theta;\hbar)\cdot(1+\mathcal{O}(z-\omega y))$. Here ω (again, with some abuse of notation) is understood as the transformation $\mathbb{R}^2 \to \mathbb{R}^2$, mapping the coordinates of $z \in M$ to the coordinates of $\omega z \in M$ in the normal coordinate system centered at $x \in M$; $d\omega$ is understood as the Jacobian of this mapping. Note that in formula (3.7.5) and Lemma 3.7.6 below, ω and $d\omega$ are also understood in this sense.

Recall from (3.7.2) that one can write $\tilde{\zeta}^{\delta}_{\gamma_0} = |\zeta^{\delta}_{\gamma_0}|^2 + \hbar^{1-2\delta}\zeta^{\delta,-1}_{\gamma_0}$, where the second term $\zeta^{\delta,-1}_{\gamma_0} \in S^0_{\delta}$ satisfies estimates of the form

$$\partial_z^{\alpha}\partial_{\xi}^{\beta}\zeta^{\delta,-1}_{\gamma_0}(\omega^{-1}z,(d\omega)^t\xi;\hbar) = O_{\alpha,\beta}\left(e^{CL_\omega}\hbar^{-\delta(|\alpha|+|\beta|)}\right). \tag{3.7.5}$$

Here, C denotes possibly different positive constants not depending on ω. The constants C may depend on α and β, but this dependence can be ignored since we need to take into account only a finite number of derivatives: $|\alpha+\beta| \le 2(m+1)$, where m is given by (3.8.4).

To prove (3.7.5) we first note that since $\zeta^{\delta,-1}_{\gamma_0} \in S^0_{\delta}$, by the chain rule, differentiating the symbol gives the negative powers of $\hbar$. The estimate then follows from Lemma 3.7.6 below, since $\omega^{-1}z \in \pi(\mathrm{supp}\,\zeta^{\delta,-1}_{\gamma_0})$ and therefore $r(x,\omega^{-1}z) < C$.

Lemma 3.7.6. *Let M and ω be as above. Fix a normal coordinate system centered at $x \in M$, and let $z \in M$ be such that $r(x, \omega^{-1}z) < C$. Let $z = (z_1, z_2)$, $\omega^{-1}z = ((\omega^{-1}z)_1, (\omega^{-1}z)_2)$ be, respectively, coordinates of z and $\omega^{-1}z$ in this system, and let $\xi = (\xi_1, \xi_2) \in T_z^*M$ and $d\omega^t\xi = ((d\omega^t\xi)_1, (d\omega^t\xi)_2) \in T^*_{\omega^{-1}z}M$ be the corresponding covectors. Then $\partial_z^\alpha(\omega^{-1}z) = O_\alpha(e^{CL_\omega})$ and $\partial_\xi^\beta(d\omega^t\xi) = O_\beta(e^{CL_\omega})$.*

Proof. Since ω is an isometry, we can identify the tangent spaces $T_xM \cong \mathbb{R}^2$ and $T_{\omega x}M \cong \mathbb{R}^2$ using $d\omega$, so that for any point $y \in M$, the coordinates of the point y in a normal coordinate system centered at x coincide with the coordinates of ωy in the normal coordinate system centered at ωx. Taking this identification into account we obtain

$$((\omega^{-1}z)_1, (\omega^{-1}z)_2) = \exp^{-1}_{\omega x} \circ \exp_x(z_1, z_2). \tag{3.7.7}$$

As was shown in [Ber, Appendix, Propositions 1 and 3], the Jacobian of the exponential map is bounded away from zero and the derivatives of the Jacobi fields have at most exponential growth at infinity. At the same time, the derivatives of the exponential map could be expressed in terms of the derivatives of the Jacobi fields [Ch, p. 103], and hence also have at most exponential growth at infinity. Therefore, the derivatives of the map $\exp^{-1}_{\omega x}$ in (3.7.7) are $\mathcal{O}(1)$ since by assumption of the lemma $r(z, \omega x) < C$. The derivatives of the map $\exp_x$ in (3.7.7) grow exponentially in $r(x,z) \le r(x, \omega x) + r(\omega x, z) \le r(x, \omega x) + C$. At the same time, $r(x, \omega x) \le L_\omega + C$ by triangle inequality. This completes the proof of the first estimate in Lemma 3.7.6.

Consider now the ξ-derivatives and let $\beta = (\beta_1, \beta_2)$. Note that $\partial_\xi^\beta(d\omega^t\xi) = 0$ if $\beta_1 \ge 2$ or $\beta_2 \ge 2$. The first derivatives of $d\omega^t\xi$ pull out components of $d\omega^t$, which is the transpose matrix of the differential $d\omega$. Let us recall in what sense $d\omega$ is understood here: it is the Jacobian of the map $\omega : \mathbb{R}^2 \to \mathbb{R}^2$ given by (3.7.7). But we have already proved above that the derivatives of this map (in particular, the first derivatives that are components of the Jacobian) grow at most exponentially in L_ω. This completes the proof of Lemma 3.7.6. □

Let us go back to the formula (3.7.4). By integration by parts in ξ in (3.7.4), modulo $\mathcal{O}(\hbar^\infty)$-errors, one can assume that $r(z, \omega y) = r(\omega^{-1}z, y) \le \frac{1}{C}$. By carrying out the same argument as for the leading term $|\zeta^\delta_{\gamma_0}|^2$ below and using the derivative bounds (3.7.5), it follows that by taking $\delta > 0$ sufficiently small, the contribution to $\kappa_{11}(\lambda, T)$ of the remainder term $\hbar^{1-2\delta}\zeta^{\delta,-1}_{\gamma_0}(\omega^{-1}z, (d\omega)^t\xi; \hbar)$ is $\mathcal{O}(\hbar^{C(\delta)})$ for some $C(\delta) > 0$. From now on, with a slight abuse of notation, we ignore this remainder and rewrite $|\zeta^\delta_{\gamma_0}|^2$ as $\tilde{\zeta}^\delta_{\gamma_0}$. So, from (3.7.4), we need to study the integral

$$\begin{aligned} I^{\pm}_{N,\omega}(x, y, t; \omega, \hbar) = (2\pi\hbar)^{-2} \int\int\int e^{i[-t^2\theta + r^2(x,z)\theta + (z - \omega y)\xi]/\hbar} a'_N(x, z, \theta; \omega, \hbar) \\ \times \eta(x)\, \tilde{\zeta}^\delta_{\gamma_0}(\omega^{-1}z, (d\omega)^t\xi; \hbar)\, \psi(-2t\theta \pm 1)\, dz d\xi d\theta. \end{aligned} \tag{3.7.8}$$

By integration by parts in the θ-variable in (3.7.8) it suffices modulo $\mathcal{O}(\hbar^\infty)$-errors to assume that for any fixed $\epsilon > 0$,

$$|r^2(x,z) - t^2| < \epsilon. \tag{3.7.9}$$

3.8. **Stationary phase in (z, ξ)-variables and an expansion for $\kappa^{\pm}_{11}(\lambda, T)$.** The goal of this section is to prove Lemma 3.8.9. We would like to apply stationary phase with parameters in the (z, ξ)-variables in (3.7.8) (see [Ho2] Theorem

7.7.5). For this, we need the following estimates for derivatives of the phase function $\phi(z,\xi;x,y,\omega) := \theta r^2(x,z)+(z-\omega y)\xi$. First, we have that for $(\omega^{-1}z, d\omega^t\xi; x,y,\omega) \in \operatorname{supp}(\tilde{\zeta}^\delta_{\gamma_0}) \times \operatorname{supp}(\eta) \times \operatorname{supp}(\eta) \times \{\omega; L_\omega \leq T(\hbar)\}$,

$$\partial_z^\alpha \partial_\xi^\beta \phi(z,\xi;x,y,\omega) = \mathcal{O}_{\alpha,\beta}(T(\hbar)^2). \tag{3.8.1}$$

The estimate (3.8.1) follows from the fact that $T(\hbar)^{-1} \leq \theta \leq T_0^{-1}$ on supp $\psi(-2t\theta \pm 1)$ and the following bounds:

$$\partial_z^\alpha \phi(z,\xi;x,y,\omega) = \partial_z^\alpha(\theta r^2(x,z)) = \theta \partial_z^\alpha r^2(x,z) = \theta \partial_z^\alpha(|z|^2)$$
$$= \mathcal{O}_\alpha(\theta\, r^2(x,z)) = \mathcal{O}_\alpha(1) T(\hbar)^2; \quad |\alpha| \geq 0.$$

Here, we have used (3.7.9) as well as the fact that $r^2(x,z) = z_1^2 + z_2^2$ since the z_j's are geodesic normal coordinates at $x \in M$. The mixed (z,ξ)-derivatives of ϕ are pointwise $\mathcal{O}(1)$. Finally,

$$\partial_\xi \phi(z,\xi;x,y,\omega) = z - \omega y = \mathcal{O}(L_\omega) = \mathcal{O}(T(\hbar)),$$

where the last estimate follows from the triangle inequality for the distance function. Moreover, the $\mathcal{O}_{\alpha,\beta}(1)$-constants appearing on the RHS in (3.8.1) are all uniform in $\omega \in \Gamma$.

The required lower bound on the norm of the gradient says that for $(\omega^{-1}z, d\omega^t\xi, x, y, \omega) \in \operatorname{supp}(\tilde{\zeta}^\delta_{\gamma_0}) \times \operatorname{supp}(\eta) \times \operatorname{supp}(\eta) \times \{\omega; L_\omega \leq T(\hbar)\}$, there exists another constant $C > 0$ (uniform in all parameters including ω), such that

$$|\partial_z \phi|^2(z,\xi;x,y) + |\partial_\xi \phi|^2(z,\xi;x,y) \geq \frac{1}{C}(|\xi + \partial_z r^2(x,z)\theta|^2 + |z-\omega y|^2), \tag{3.8.2}$$

Indeed, the lower bound in (3.8.2) follows by Taylor expansion around the critical point $\xi = -\partial_z r^2(x,z)\theta, z = \omega y$ and the Hessian lower bound $(\partial^2_{z,\xi}\phi) \gg Id$. The latter follows from the fact that $\partial_z^2 \phi = \theta \partial_z^2 r^2(x,z)) = \theta \partial_z^2(z_1^2 + z_2^2) = \mathcal{O}(\theta) = \mathcal{O}(1)$, $\partial_z \partial_\xi \phi = 1$ and finally, $\partial_\xi^2 \phi = 0$. Given (3.8.1) and (3.8.2), by Hörmander's interpolation proof of stationary phase ([Ho2] Theorem 7.7.5),

$$\begin{aligned} I^\pm_{N,\omega}(x,y,t;\hbar) = (2\pi\hbar)^{-1/2} \sum_{k=0}^N \int_0^\infty & e^{i[-t^2+r^2(x,\omega y)]\theta/\hbar} \\ & \times b_N^k(x,y,\theta;\omega,\hbar)\, \psi(-2t\theta \pm 1)\, \theta_+^{-1/2-k} d\theta. \end{aligned} \tag{3.8.3}$$

Set

$$m = \left[\frac{4}{1-2\delta}\right] \tag{3.8.4}$$

In (3.8.3) we have

$$\begin{aligned} b_N^k(x,y,\theta;\omega,\hbar) = \hbar^k \sum_{j=0}^m \frac{\hbar^j (D_z D_\xi)^j}{(j+1)!} & [a_k(x,z) \cdot \tilde{\zeta}^\delta_{\gamma_0}(\omega^{-1}z, d\omega^t\xi)]|_{\xi=-\partial_z r^2(x,z)\theta, z=\omega y} \\ & + \mathcal{O}(\hbar^{k+m-1}\, T(\hbar)^2 \sup_{|\alpha+\beta| \leq 2(m+1)} |\partial_z^\alpha \partial_\xi^\beta (a_k \cdot \tilde{\zeta}^\delta_{\gamma_0})|). \end{aligned} \tag{3.8.5}$$

Let us now estimate the contribution to $\kappa^\pm_{11}(\lambda, T)$ coming from the error in (3.8.5). Since it is well-known that $\partial_z^\alpha a_k = \mathcal{O}_\alpha(e^{CL_\omega})$ ([Ber, Appendix, Lemma 4]), we have that

$$\hbar^{m-1} \sup_{|\alpha+\beta|\leq 2(m+1)} |\partial_z^\alpha \partial_\xi^\beta (a_k \cdot \tilde{\zeta}_{\gamma_0}^\delta)| \ll e^{CL_\omega} \hbar^{-2(m+1)\delta+m-1})$$

$$\ll e^{C\epsilon|\ln\hbar|}\, \hbar^{(m+1)(1-2\delta)-2} \ll \hbar^{(m+1)(1-2\delta)-2-C\epsilon} \ll \hbar^{2-C\epsilon} \tag{3.8.6}$$

Consequently, from (3.8.3), (3.7.8) and (3.6.3) it follows that the contribution to $\kappa_{11}^{\pm}(\lambda,T)$ coming from the remainder in (3.8.5) for each $k=0,...,N$, is bounded by

(3.8.7)

$$\frac{T(\hbar)^2}{T(\hbar)} \sum_{T_0\leq L_\omega\leq T(\hbar)} \int_{-\infty}^{\infty}\int_0^\infty |\psi(-2t\theta\pm 1)| \chi(t;T(\hbar))| \partial_t e^{\pm it/\hbar}| \hbar^{2-C\epsilon}\, \hbar^k\, \theta_+^{-1/2-k}\, d\theta dt$$

$$\ll \hbar^{k+1-C\epsilon} T(\hbar) \sum_{L_\omega\leq T(\hbar)} \int_{T_0}^{T(\hbar)} \int_{1/T(\hbar)}^{1/T_0} \theta_+^{-1/2-k}\, d\theta\, dt$$

$$\ll \hbar^{k+1-C\epsilon}\, T(\hbar)\, e^{C|T(\hbar)|} \int_{T_0}^{T(\hbar)} |T(\hbar)|^{-1/2+k}\, dt \ll \hbar^{k+1-C\epsilon}\, |T(\hbar)|^{3/2+k}.$$

In the second to last estimate in (3.8.7) we have used the exponential bounds for the growth of the number of periodic geodesics [M-S].

Choosing $\epsilon>0$ small enough and taking into account that $T(\hbar)\leq \epsilon|\ln\hbar|$, it follows that the error terms for $k\geq 0$ in (3.8.5) are $\mathcal{O}(\hbar^{\alpha_1})$ for some $\alpha_1>0$. We are only interested here in working up to such errors. So, it is enough to consider only the principal term in (3.8.5). Next, note that for each $j\geq 1$,

(3.8.8)

$$\hbar^j \frac{(D_z D_\xi)^j}{(j+1)!} [a_k(x,z)\cdot \tilde{\zeta}_{\gamma_0}^\delta(\omega^{-1}z, d\omega^t\xi)]|_{\xi=-\partial_z r^2(x,z)\theta,\, z=\omega y} = \mathcal{O}(\hbar^{j(1-2\delta)} e^{C_j L_\omega}),$$

and so again after possibly shrinking $\epsilon>0$ in the $T(\hbar)\leq \epsilon\ln\hbar$, it suffices modulo $\mathcal{O}(\hbar^{\alpha_2})$ to restrict the analysis to the $j=0$ case. By the same token, it suffices to restrict to the $k=0$ case in (3.8.3).

To get $\kappa_{11}^{\pm}(\lambda,T)$, we put $y=x$ in (3.8.3), integrate over $x\in \text{supp}(\eta)$ and sum over $\omega\in\Gamma; T_0<L_\omega<T(\hbar)$. Substituting the formula (3.8.3) in (3.6.3) and taking into account the estimate in (3.8.8), for appropriate $\alpha_3>0$, we get the following

Lemma 3.8.9.

$$\kappa_{11}^{\pm}(\lambda,T) = \frac{1}{2T(\hbar)} \sum_{id\neq\omega\in\Gamma} \int_0^\infty \int_M \int_{-\infty}^{\infty} e^{i\theta[-t^2+r^2(x,\omega x)]/\hbar} b_N^0(x,x,\theta;\omega,\hbar)(\hbar\theta_+)^{-\frac{1}{2}}$$

$$\times\eta(x)\,\psi(-2t\theta\pm 1)\frac{\partial}{\partial t}\left[e^{\pm it/\hbar}\,\chi(t;T(\hbar))\right]\, dtdvol(x)d\theta + \mathcal{O}(\hbar^{\alpha_3})$$

$$= \frac{1}{2\hbar T(\hbar)} \sum_{id\neq\omega\in\Gamma} \int_0^\infty \int_M \int_{-\infty}^{\infty} e^{i\theta[-t^2+r^2(x,\omega x)\pm t]/\hbar} b_N^0(x,x,\theta;\omega,\hbar)(\hbar\theta_+)^{-\frac{1}{2}}$$

$$\times\eta(x)\,\psi(-2t\theta\pm 1)\chi(t;T(\hbar))\, dtdvol(x)d\theta + \mathcal{O}(\hbar^{\alpha_3}).$$

In the last line in (3.8.9), we have used that the leading term comes from applying ∂_t to the exponential $e^{\pm it/\hbar}$ since $\partial_t\chi(t;T(\hbar)) = \partial_t[\hat{\rho}(t/T)\,(1-\psi)(t)] = \mathcal{O}(1)$.

For a fixed $\omega\in\Gamma$, every term in (3.8.9) can be represented as a sum using the expansion (3.8.5) for b_N^0. Let us estimate each term in this sum separately.

Given the estimate in (3.8.8) the $j = 0$ – term dominates in (3.8.5) and so, up to $\mathcal{O}(\hbar^{\alpha_4})$-errors, it suffices to estimate:

$$\int\int\int e^{i\theta[-t^2+r^2(x,\omega x)]/\hbar \pm it/\hbar}\,\tilde{\zeta}^{\delta}_{\gamma_0}(x, -d\omega^t \cdot \partial_z r^2(x,z)|_{z=\omega x}\,\theta) \tag{3.8.10}$$
$$\times \eta(x)\,a_0(x,\omega x)\,(\hbar\theta_+)^{-\frac{1}{2}}\,\chi(t;T(\hbar))\,\psi(-2t\theta \pm 1)\,dt\,dvol(x)\,d\theta.$$

3.9. **Splitting into short time-windows.** The goal of this section is to prove Lemma 3.9.6. We split up the iterated time integral in (3.8.10) into a sum over 'time-windows' of size $\frac{\delta'}{T(\hbar)}$. The reason for this is that Lemma 2.2.1 only controls the splitting of geodesics in a time interval of size $\frac{\delta'}{T(\hbar)}$ (note that the original time variable t has already been rescaled to $\frac{t}{T(\hbar)}$ at this point). Consider a covering of $[T_0, T(\hbar)]$ by open intervals I_j of length δ', $j = 0, \dots [\frac{T(\hbar)-T_0}{\delta'}]$, and let $\eta_j \in C_0^\infty(\mathbb{R})$ be a partition of unity subordinate to this covering. For $j = 0, ..., [\frac{T(\hbar)-T_0}{\delta'}]$, we will need the following cutoff functions:

$$\chi_j(t;T(\hbar)) = \eta_j(T(\hbar)t)\cdot\chi(t;T(\hbar))$$
$$= \eta_j(T(\hbar)t)\cdot\hat{\rho}(t/T(\hbar))\cdot(1-\psi)(t). \tag{3.9.1}$$

Since we have already inserted the cutoff function $\psi(-2t\theta \pm 1)$ in (3.8.10) this allows us to apply Fubini and do the (t,θ)-iterated integrals in (3.8.10) first and the x-integration last. Indeed, we just rewrite the total phase in (3.8.10) as a sum:

$$r(x,\omega x) + \Phi^{\pm}(t,\theta;x,\omega), \tag{3.9.2}$$

where,

$$\Phi^{\pm}(t,\theta;x,\omega) = (\pm t - r(x,\omega x)) - \theta\,(t^2 - r^2(x,\omega x)).$$

At this point, we would like to do stationary phase in (t,θ), treating $(x,\omega,\hbar)$ as parameters. Just as in (3.7.8) we need to establish a couple of estimates for derivatives of the phase $\Phi^{\pm}$. The first estimate

$$\partial_\theta^\alpha \partial_t^\beta \Phi^{\pm}(t,\theta;x,\omega) = \mathcal{O}_{\alpha,\beta}(r^2(x,\omega x)) \tag{3.9.3}$$

with $\mathcal{O}_{\alpha,\beta}$-constants uniform in all parameters is immediate. Since $\partial_t\partial_\theta\Phi^{\pm} = -2t$, $\partial_\theta^2\Phi^{\pm} = 0$ and $\partial_t^2\Phi^{\pm} = -2\theta$ one gets the following lower bound for the (t,θ)-Hessian of $\Phi^{\pm}$:

$$(\partial^2_{t,\theta}\Phi^{\pm})^t\cdot(\partial^2_{t,\theta}\Phi^{\pm}) \geq \frac{1}{C}\,Id,$$

and so, by Taylor expansion around the critical points $t = \pm r(x,\omega x)$ and $\theta = \frac{1}{2r(x,\omega x)}$, one gets the uniform lower bound

$$|\partial_t\Phi^{\pm}|^2 + |\partial_\theta\Phi^{\pm}|^2 \geq \frac{1}{C}\left(|t \pm r(x,\omega x)|^2 + |\theta - \frac{1}{2r(x,\omega x)}|^2\right). \tag{3.9.4}$$

So, by [Ho2] Theorem 7.7.5, it follows that for the expression in (3.8.10):

$$\int\int\int e^{i\theta[-t^2+r^2(x,\omega x)]/\hbar \pm it/\hbar}\,\tilde{\zeta}^{\delta}_{\gamma_0}(x, -d\omega^t \cdot \partial_z r^2(x,z)|_{z=\omega x}\,\theta)\cdot\eta(x)\,a_0(x,\omega x)$$
$$\times(\hbar\theta_+)^{-\frac{1}{2}}\,\chi(t;T(\hbar))\,\psi(-2t\theta\pm 1)\,dt\,d\theta\,dvol(x)$$
$$= 2\pi\hbar^{1/2}\int_M e^{ir(x,\omega x)/\hbar}\eta(x)\,a_0(x,\omega x)\cdot\tilde{\zeta}^{\delta}_{\gamma_0}(x, -d\omega^t\cdot\partial_z r(x,z)|_{z=\omega x})\,\cdot\,r(x,\omega x)^{-\frac{1}{2}}$$
$$\times\psi(\pm r(x,\omega x)/T)\,(1-\psi)(\pm r(x,\omega x))\,dvol(x)\,+\mathcal{O}(\hbar^{2-2\delta}\,e^{CL_\omega}). \tag{3.9.5}$$

In the last line of (3.9.5), we have again used the exponential bounds for the derivatives of a_0 [Ber] as well as the uniform lower bounds for $\Phi^{\pm}$ in (3.9.4). The contribution of the $\mathcal{O}(\hbar^{2-2\delta}\, e^{CL_\omega})$ error term to $\kappa_{11}(\lambda,T)$ is then

$$\ll \frac{\hbar^{2(1-\delta)}}{\hbar T(\hbar)} \sum_{\omega; L_\omega \leq T(\hbar)} e^{CL_\omega} \ll \hbar^{1-2\delta} e^{C|T(\hbar)|} |T(\hbar)|^{-2}.$$

After possibly reducing the size of $\epsilon > 0$ further, this term is then $\mathcal{O}(\hbar^{\alpha_5})$ for some $\alpha_5 > 0$.

We summarize what was shown so far in the following

Lemma 3.9.6.

$$\kappa_{11}^{\pm}(\lambda,T) = \frac{\hbar^{-1/2}}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \sum_{j=0}^{[\frac{T(\hbar)-T_0}{\delta'}]} \int_M e^{ir(x,\omega x)/\hbar} a_0(x,\omega x) \tilde{\zeta}^{\delta}_{\gamma_0}(x, -d\omega^t \cdot \partial_z r(x,z)|_{z=\omega x})$$
$$\cdot\, \chi_j(\pm r(x,\omega x); T(\hbar))\, r(x,\omega x)^{-\frac{1}{2}}\, \eta(x)\, dvol(x) + \mathcal{O}(\hbar^{\alpha_5}),$$

for some $\alpha_5 > 0$, where $\eta \in C_0^\infty(M;[0,1])$ is an appropriate cutoff function with $\mathrm{diam}(\mathrm{supp}\ \eta) \leq 2\,\mathrm{diam}(X) + 1$.

3.10. **Stationary phase in the x-variables.** The last step involves expanding each term in the ω-sum in (3.9.6) separately. For fixed $id \neq \omega \in \Gamma$, one carries out stationary phase in (3.9.6) in the x-variables transverse to the lift of the periodic geodesic on X given by

$$\gamma(\omega) = \{x \in \mathrm{supp}(\eta) \subset M; \nabla_x f_\omega(x) = 0\}, \tag{3.10.1}$$

where $f_\omega(x) = r(x,\omega x)$ is the displacement function.

Since the dynamical Lemma 2.2.1 only controls separation of geodesics in time intervals of size $\frac{\delta'}{T(\hbar)}$, we estimate the summands:

$$\frac{\hbar^{-1/2}}{2T(\hbar)} \sum_{id \neq \omega \in \Gamma} \int_M e^{ir(x,\omega x)/\hbar} a_0(x,\omega x) \cdot \tilde{\zeta}^{\delta}_{\gamma_0}(x, -d\omega^t \cdot \partial_z r(x,z)|_{z=\omega x}) \tag{3.10.2}$$
$$\times\ \chi_j(\pm r(x,\omega^k x); T(\hbar))\, r(x,\omega x)^{-\frac{1}{2}}\, \eta(x)\, dvol(x); \quad j = 0, ..., [\frac{T(\hbar)-T_0}{\delta'}]$$

separately. For fixed $id \neq \omega \in \Gamma$ and $j \in \{0,...,[\frac{T(\hbar)-T_0}{\delta'}]$, we introduce Fermi coordinates $(u,s) \in \mathbb{R} \times [0,L]$ in (3.10.2) (see [CdV]) centered on the geodesic segment $\gamma(\omega)$ (i.e. $u = 0$ on $\gamma(\omega)$) and apply stationary phase in the u-variables just as in the $T(\hbar) \sim 1$ case (see [CdV, section 2]). In terms of the (u,s)-coordinates on M

$$dvol(x) = J(u,s;\omega)\, duds, \tag{3.10.3}$$

where, $|\partial_u^\alpha \partial_s^\beta J(u,s;\omega)| = \mathcal{O}_{\alpha,\beta}(e^{C(\alpha,\beta)L_\omega})$ [CdV].

In the case at hand, it is necessary to control the dependence of the phase and amplitude of (3.9.6) on the parameters, ω.

We now need the following estimates: The first is a (uniform) Hessian lower bound ([CdV, Lemma 4]) which says that

$$\nabla_u^2 f_\omega(u,s) \geq C_0 > 0. \tag{3.10.4}$$

The constant $C_0 > 0$ in (3.10.4) is *uniform* in the parameters $\omega \in \Gamma$ and depends only the curvature pinching conditions $-K_1^2 \leq K \leq -K_2^2$. In addition, one has the upper bounds [Ber, CdV]:

$$\partial_u^\alpha a_0(x, \omega x) = \mathcal{O}_\alpha(e^{CL_\omega}),\ \partial_u^\alpha f_\omega(u,s) = \mathcal{O}_\alpha(e^{C'L_\omega}), \\ \partial_u^\alpha J(u,s) = \mathcal{O}_\alpha(e^{C''L_\omega}), \tag{3.10.5}$$

where the estimates in (3.10.5) are all uniform in the Fermi coordinates $(u,s) \in \mathbb{R} \times [0,L]$. In order to apply stationary phase with parameters, we need to carefully analyze the critical sets of the phase function $f_\omega(u,s)$ in (3.10.2) for all $\omega \in \Gamma$ with $L_\omega \ll |\ln \hbar|$. We do this, by combining the estimates (3.10.4) and (3.10.5) with the dynamical Lemma 2.2.1.

3.11. **Application of Lemma 2.2.1.** As above, let $\gamma(\omega)$ be the unique lifted geodesic invariant under the action of $\omega \in \Gamma$. As is well-known [B-O, Proposition 4.2], the displacement function $f_\omega(x)$ on a negatively curved surface is strictly convex, except on $\gamma(\omega)$ where it is constant: $f_\omega(x) = L_\omega$ when $x \in \gamma(\omega)$. The geodesic $\gamma(\omega)$ is also the critical set of $f_\omega(x)$, see (3.10.1).

Fix $j = 0, ..., [\frac{T(\hbar)-T_0}{\delta'}]$ and $\omega \in \Gamma$ and consider the corresponding summand in (3.10.2) given by

$$\frac{\hbar^{-1/2}}{2T(\hbar)} \int_M e^{ir(x,\omega x)/\hbar} a_0(x,\omega x) \cdot \tilde{\zeta}_{\gamma_0}^\delta(x, -d\omega^t \cdot \partial_z r(x,z)|_{z=\omega x})\, \chi_j(\pm r(x,\omega^k x); T(\hbar)) \\ \times r(x,\omega x)^{-\frac{1}{2}}\, \eta(x)\, dvol(x). \tag{3.11.1}$$

We now show that, the integral in (3.10.2) is $\mathcal{O}(\hbar^\infty)$ unless $\omega = \omega_0$, where the latter group element fixes the lifted geodesic $\gamma_0(\omega_0)$. As before, $\pi : T^*M \to M$ denotes the standard cotangent projection map and $\tilde{\gamma} \subset T^*M$ will denote the bicharacteristic curve with $\pi(\tilde{\gamma}) = \gamma$.

To prove the above claim, we first note that on supp $\tilde{\zeta}_{\gamma_0}^\delta(x, -d\omega^t \partial_z r(x,z)|_{z=\omega x})$,

$$d(x, \pi(\tilde{\gamma_0})) = d(x, \gamma_0) = \mathcal{O}(\hbar^\delta). \tag{3.11.2}$$

Here, following our convention we write γ_0 for $\gamma_0(\omega_0)$. Since $d\omega_0^t \cdot \partial_z r(x,z)|_{z=\omega_0 x}$ is the cotangent vector to $\tilde{\gamma_0}$ at $x \in \pi(\tilde{\gamma_0})$, it follows from the small-scale microlocalization in the ξ-variables that for any $x \in \gamma_0$, we have that

$$d\omega^t \cdot \partial_z r(x,z)|_{z=\omega x} = d\omega_0^t \cdot \partial_z r(x,z)|_{z=\omega_0 x} + \mathcal{O}(\hbar^\delta). \tag{3.11.3}$$

From the exponential upper bounds for the derivatives of f_ω in (3.10.5) and by a Taylor expansion around γ_0, it follows that for any $x \in M$ with $d(x,\gamma_0) = \mathcal{O}(\hbar^\delta)$,

$$d\omega^t \cdot \partial_z r(x,z)|_{z=\omega x} = d\omega_0^t \cdot \partial_z r(x,z)|_{z=\omega_0 x} + \mathcal{O}(\hbar^{\delta_0}), \tag{3.11.4}$$

for some $0 < \delta_0 < \delta$.

On the other hand, by taking $\epsilon > 0$ small enough, and again using the upper bounds in (3.10.5), it follows by an integration by parts in (3.9.6) in the transversal u-variable, that modulo $\mathcal{O}(\hbar^\infty)$-errors, one can cut off the integration to values of x satisfying:

$$\partial_x f_\omega(x) = \mathcal{O}(\hbar^\delta), \tag{3.11.5}$$

for any $0 \leq \delta < 1/2$. Then using (3.10.5) again together with the uniform Hessian lower bound in (3.10.4), it follows by a Taylor expansion argument that

$$d(\gamma_0(\omega_0), \gamma(\omega)) = \mathcal{O}(\hbar^{\delta_1}). \tag{3.11.6}$$

Here, $0 < \delta_1 < \delta_0$ is yet another, possibly smaller constant. But then, in view of the estimate (3.11.4) we get that

$$d(\tilde{\gamma}_0(\omega_0), \tilde{\gamma}(\omega)) = \mathcal{O}(\hbar^{\delta_1}). \tag{3.11.7}$$

The presence of the time cutoff χ_j in (3.10.2) ensures that there is only one j-summand in (3.10.2) that contributes in a non-negligible way to $\kappa_{11}^{\pm}$, namely, the interval I_j containing L_{γ_0}.

Finally, by possibly shrinking $\epsilon > 0$ further, it follows from (3.11.7) and the dynamical Lemma 2.2.1 that, up to $\mathcal{O}(\hbar^\infty)$-error in $\kappa_{11}^{\pm}(\lambda, T)$, $\omega = \omega_0$.

One can repeat the above argument for each $j \in \{0, ..., [\frac{T(\hbar) - T_0}{\delta'}]\}$ and $id \neq \omega \in \Gamma$ separately in (3.11.1) and sum up. By taking the exponential bounds in (3.10.5) into account and possibly further decreasing the size of ϵ in $T(\hbar) = \epsilon |\ln \hbar|$ if necessary, it follows by applying stationary phase in u (see [CdV, Don, Sun]) that for appropriate $\alpha_6 > 0$,

$$\kappa_{11}^{\pm}(\lambda, T) = \frac{e^{\pm i L_{\gamma_0}/\hbar}}{2T(\hbar)} \cdot L_{\gamma_0}^{\sharp} \cdot \chi\left(L_{\gamma_0}, T(\hbar)\right) \cdot |\det(I - P_{\gamma_0})|^{-1/2} + E(\hbar), \tag{3.11.8}$$

where,

$$E(\hbar) = \mathcal{O}(\hbar^{\alpha_6}) + \mathcal{O}\left(\frac{e^{CT(\hbar)}\hbar^\infty}{T(\hbar)}\right) = \mathcal{O}(\hbar^{\alpha_6}). \tag{3.11.9}$$

The first term on the RHS of (3.11.9) gives the remainder produced by the stationary phase method for $\omega = \omega_0$. The second term on the RHS of (3.11.9) follows from the bound $\#\{\gamma; L_\gamma \leq T\} = \mathcal{O}(e^{CT(\hbar)}T(\hbar)^{-1})$, and the estimate $\mathcal{O}(\hbar^\infty)$ for each summand in (3.10.2) corresponding to $\omega \neq \omega_0$.

Estimate for $\kappa_{12}^{\pm}(\lambda, T)$: Lemma 2.2.1 implies that there are *no* periodic geodesics in supp $(\,\zeta_{\gamma_0}^{\delta} \cdot (1 - \zeta_{\gamma_0}^{\delta})\,)$. So, by repeated integration by parts in the u-variable in (the analogue of) (3.11.1), it follows that

$$\kappa_{12}(\lambda, T) = \mathcal{O}\left(\frac{e^{CT(\hbar)}\hbar^\infty}{T(\hbar)}\right) = \mathcal{O}(\hbar^\infty), \tag{3.11.10}$$

when $T(\hbar) = \epsilon |\ln \hbar|$.

Estimate for $\kappa_{22}^{\pm}(\lambda, T)$: Here, repeat the microlocalization as in the estimate for $\kappa_{11}(\lambda, T)$ near each periodic geodesic $\gamma \neq \gamma_0$ with $0 < L_\gamma \leq \epsilon |\ln \hbar|$ separately and sum up. By the same argument as for $\kappa_{12}(\lambda, T)$, all cross terms give $\mathcal{O}(\hbar^\infty)$ contributions to $\kappa(\lambda, T)$. Also, we note that since the remainder term is $\mathcal{O}(\hbar^{\alpha_6})$ for some $\alpha_6 > 0$ in (3.11.8), after summing over all $\omega \neq \omega_0$ in $\kappa_{22}^{\pm}$, it follows that, after possibly shrinking $\epsilon > 0$ further, the remainder in the latter is

$$\mathcal{O}(e^{CT(\hbar)}T(\hbar)^{-1}\hbar^{\alpha_6}) = \mathcal{O}(\hbar^{\alpha_7}),$$

for some $\alpha_7 > 0$. This finishes the proof of Proposition 3.2.1 since the $\mathcal{O}(\hbar^{\alpha_7})$-error is absorbed in the $\mathcal{O}(T^{-1})$-error in (3.2.4) for $\kappa(\lambda, T)$. □

4. Proof of Theorem 1.3.1

4.1. **Application of thermodynamic formalism.** The strategy of the proof of Theorem 1.3.1 is to get a contradiction with Lemma 2.1.6 that gives an *upper bound* for the quantity $\kappa(\lambda, T)$ defined in (2.1.4), (2.1.5); that bound holds for *any* λ, T. We use the formula (3.2.2) and the estimate (4.1.3) below to obtain a *lower bound* for $\kappa(\lambda, T)$ which holds for an infinite sequence of pairs (λ, T) such that $T \sim \frac{1}{h} \ln \ln \lambda$, where h is the topological entropy of the geodesic flow on X.

Consider the main term in the formula (3.2.2). This sum is a trigonometric polynomial in λ, while the number of terms and the coefficients depend on T.

We would like to choose λ and T in such a way that the value of the polynomial is large. We first study the rate of growth for the sum of coefficients of this trigonometric polynomial as $T \to \infty$ not taking into account the oscillatory nature of the terms. Let

$$S(T) = \sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T} \frac{L_\gamma}{\sqrt{|\det(I - \mathcal{P}_\gamma)|}}; \tag{4.1.1}$$

It turns out that the asymptotic rate of growth of $S(T)$ can be determined using results from the theory of thermodynamic formalism for Anosov flows:

Proposition 4.1.2. *There exists a constant $C_0 > 0$ such that*

$$S(T) = C_0 e^{P\left(-\frac{\mathcal{H}}{2}\right)\cdot T}(1 + o(1)) \tag{4.1.3}$$

as $T \to \infty$, where P is the topological pressure (1.2.2) *and $\mathcal{H}$ is the Sinai-Ruelle-Bowen potential* (1.2.1).

It was shown in [J-P] that $P\left(-\frac{\mathcal{H}}{2}\right) \ge K_2/2$, hence $S(T)$ grows exponentially in T.

Proof of Proposition 4.1.2. Let $\xi_\gamma \in T_{\gamma(0)}M$ be the tangent vector to γ. The Poincaré map $\mathcal{P}_\gamma$ preserves the unstable subspace $E^u_{\xi_\gamma}$ and the stable subspace $E^s_{\xi_\gamma}$, both of dimension one. The map $\mathcal{P}_\gamma$ is symplectic and has determinant equal to one. The eigenvalues $\mu, 1/\mu$ of $\mathcal{P}_\gamma$, corresponding, respectively, to the unstable and the stable directions, satisfy:

$$\ln|\mu| \in [K_2 L_\gamma, K_1 L_\gamma]. \tag{4.1.4}$$

It follows from (4.1.4) that

$$|\det(I - \mathcal{P}_\gamma)| = (|\mu| - 1)(1 - \frac{1}{|\mu|}) = |\mu|\left(1 + O(e^{-K_2 L_\gamma})\right),$$

On the other hand, by definition of $\mathcal{H}$

$$|\mu| = \exp\left[\int_0^{L_\gamma} \mathcal{H}(G^s \xi_\gamma) ds\right],$$

so we have

$$|\det(I - \mathcal{P}_\gamma)| = \exp\left[\int_0^{L_\gamma} \mathcal{H}(G^s \xi_\gamma) ds\right]\left(1 + O(e^{-K_2 L_\gamma})\right). \tag{4.1.5}$$

Split the sum $S(T)$ into two parts:

$$S(T) = \sum_{L_\gamma \le T/2} + \sum_{T/2 < L_\gamma \le T} = S_1(T) + S_2(T).$$

It follows from (4.1.1) and (4.1.5) that
(4.1.6)
$$S_2(T) = \sum_{L_\gamma \in \mathrm{Lsp}, T/2 \le L_\gamma \le T} L_\gamma \exp\left[-\frac{1}{2}\int_0^{L_\gamma} \mathcal{H}(G^s \xi_\gamma) ds\right] \left(1 + O(e^{-K_2 T/2})\right),$$
and that
$$S_1(T) = O\left(\sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T/2} L_\gamma \exp\left[-\frac{1}{2}\int_0^{L_\gamma} \mathcal{H}(G^s \xi_\gamma) ds\right]\right).$$

It follows easily from the results of [Par], [P-P, (7.1)], [M-S, p. 109] that
$$S_2(T) = \frac{e^{P\left(-\frac{\mathcal{H}}{2}\right)\cdot T}}{P(-\mathcal{H}/2)} (1 + o(1))$$
and that
$$S_1(T) = O\left(e^{P\left(-\frac{\mathcal{H}}{2}\right)\cdot \frac{T}{2}}\right).$$
This finishes the proof of the proposition □

If follows easily from Proposition 4.1.2 that
$$\tilde{S}(t) = \sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T} \frac{L_\gamma^\sharp}{\sqrt{|\det(I - \mathcal{P}_\gamma)|}} = C_0 e^{P\left(-\frac{\mathcal{H}}{2}\right)\cdot T}(1 + o(1)). \tag{4.1.7}$$

Indeed, since each imprimitive geodesic is a multiple of some geodesic of at least twice smaller length, the contribution of imprimitive geodesics into $S(T)$ is $O\left(e^{P\left(-\frac{\mathcal{H}}{2}\right)\cdot \frac{T}{2}}\right)$ and hence could be neglected (cf. [J-P, p. 25]).

4.2. **Preliminary estimates.** We can now finish the proof of Theorem 1.3.1. Assume for contradiction that Theorem 1.3.1 doesn 't hold. Then $R(\lambda)$ satisfies
$$R(\lambda) = O\left((\ln \lambda)^{P(-\mathcal{H}/2)(1-\varepsilon)/h}\right)$$
for some $\varepsilon > 0$. Let $b = P(-\mathcal{H}/2)(1-\varepsilon)/h$ be the exponent $\ln \lambda$ in the previous formula. Then by Lemma 2.1.6 we have
$$\kappa(\lambda, T) = O((\ln \lambda)^b). \tag{4.2.1}$$

We rewrite (4.2.1) as
$$\ln |\kappa(\lambda, T)| \le b \ln\ln \lambda + C_1. \tag{4.2.2}$$
To finish the proof, it suffices to establish a contradiction with (4.2.2). This will be done using Proposition 3.2.1 and the estimate (4.1.3) for a suitable choice of λ and T. In the sequel, we shall let $\lambda, T \to \infty$ while keeping $T \sim \frac{1}{h} \ln\ln \lambda$. This ensures that the hypothesis $T < \epsilon \ln \lambda$ of Proposition 3.2.1 is satisfied.

Denote the main term in the asymptotics of $\kappa(\lambda, T)$ by
$$\Sigma(\lambda, T) = \sum_{L_\gamma \in \mathrm{Lsp}, L_\gamma \le T} \frac{L_\gamma^\sharp \cos(\lambda L_\gamma)\, \chi(L_\gamma, T)}{T\sqrt{|\det(I - \mathcal{P}_\gamma)|}}.$$
It is a trigonometric polynomial in λ. According to (1.4.2), without loss of generality we may assume that
$$\chi(L_\gamma, T) \ge 1/2, \qquad \forall\ L_\gamma \in (T_0, T(1 - \varepsilon/2)]. \tag{4.2.3}$$

The condition $L_\gamma > T_0$ is not essential because it rules out only a finite number of closed geodesics, and their total contribution to (4.1.7) is $O(1)$.

Next, we would like to choose λ so that all the terms $\cos(\lambda L_\gamma)$ will be (say) $\geq 1/2$ for $L_\gamma \in \mathrm{Lsp}, T_0 < L_\gamma \leq T$. Let $\nu(T)$ be the number of *distinct* such L_γ-s, and let $L_1, L_2, \ldots, L_{\nu(T)}$ be the corresponding lengths. It suffices to choose λ so that

$$\mathrm{dist}(\lambda L_j, 2\pi\mathbb{Z}) \ \leq \ 1/2, \qquad 1 \leq j \leq \nu(T). \tag{4.2.4}$$

Assuming (4.2.3) and (4.2.4), we get for large enough T

$$\ln|\Sigma(\lambda, T)| \geq C_2 + \ln|\tilde{S}(T)|.$$

Using the estimate (4.1.7) and Proposition 3.2.1 we conclude that

$$\ln|\kappa(\lambda, T)| \geq P(-\mathcal{H}/2)T(1-\varepsilon/2) - \ln T + C_3. \tag{4.2.5}$$

This formula will be used to get a contradiction with the upper bound (4.2.2) for $|\kappa(\lambda, T)|$.

4.3. **Dirichlet box principle.** We next explain how to choose λ so that (4.2.4) would hold. Let M_1 be a large constant whose value will be specified later. Then by Dirichlet box principle ([J-P], see also [P-Rud, Rub-S]) there exists

$$\lambda \in [M_1, M_1 2^{\nu(T)}]$$

such that (4.2.4) holds. Hence, for *any* choice of M_1 there exists λ satisfying

$$\ln\ln M_1 \leq \ln\ln\lambda \leq \ln\ln M_1 + \ln\nu(T) + \ln\ln 2.$$

for which (4.2.4) holds.

It follows from results of Margulis [M-S] that $\nu(T) = e^{hT}(1+o(1))/hT$ as $T \to \infty$. Therefore, any λ satisfying the previous inequality would also satisfy

$$\ln\ln M_1 \leq \ln\ln\lambda \leq \ln\ln(M_1) + hT - \ln(hT) \tag{4.3.1}$$

We now choose

$$M_1 = \exp(\exp(\alpha T)), \quad \text{where } \alpha < \frac{h\varepsilon}{2(1-\varepsilon)}. \tag{4.3.2}$$

Then (4.3.1) becomes

$$\alpha T \leq \ln\ln\lambda \leq (h+\alpha)T - \ln(hT). \tag{4.3.3}$$

The first inequality in (4.3.3) ensures that the hypothesis of Proposition 3.2.1 is satisfied, implying (3.2.2). By the previous argument, we have shown that (4.3.3) and (4.2.5) implies existence of λ such that the following estimate holds:

$$\ln|\kappa(\lambda, T)| \geq P(-\mathcal{H}/2)T(1-\varepsilon/2) - \ln T + C_3. \tag{4.3.4}$$

To establish a contradiction with the formula (4.2.2), it suffices to have

$$P(-\mathcal{H}/2)T(1-\varepsilon/2) - \ln T + C_3 > b\ln\ln\lambda + C_1 = \frac{P(-\mathcal{H}/2)(1-\varepsilon)}{h}\ln\ln\lambda + C_1$$

or

$$\ln\ln\lambda \leq \frac{h(1-\varepsilon/2)}{1-\varepsilon}\,T + \frac{h}{P(-\mathcal{H}/2)(1-\varepsilon)}(C_4 - \ln T) \tag{4.3.5}$$

If we could show that the inequality (4.3.3) *implies* the inequality (4.3.5), we would be done. Indeed, by Dirichlet box principle there exists *some* λ satisfying (4.3.3), and so (4.3.5) holds for *that* value of λ, establishing a contradiction. The linear function of T is the fastest-growing term in the right-hand side of both inequalities, so it suffices to compare the coefficients of T. The coefficient in (4.3.3) is equal to $h+\alpha$, while that in (4.3.5) is equal to $h(1-\varepsilon/2)/(1-\varepsilon)$. Accordingly, it suffices to have

$$h+\alpha < h(1-\varepsilon/2)/(1-\varepsilon),$$

and this is ensured by the choice of α in (4.3.2). This establishes the desired contradiction and finishes the proof of Theorem 1.3.1. □

5. Remainder estimates in higher dimensions

5.1. **Proof of Theorem 1.5.1.** In this section we assume that X is a compact Riemannian manifold of dimension $n \geq 3$. Recall that the Riesz mean of order k of a function $f(\lambda)$ is defined by

$$\mathcal{R}_k f(\lambda) = \frac{k}{\lambda}\int_0^\lambda \left(1-\frac{t}{\lambda}\right)^{k-1} f(t)dt, \quad k=1,2,\ldots \tag{5.1.1}$$

As was shown in [Saf], on any smooth compact n-dimensional Riemannian manifold

$$\mathcal{R}_2 R(\lambda) = Ca_1\lambda^{n-2} + O(\lambda^{n-3}), \tag{5.1.2}$$

Here C is a non-zero constant depending on the dimension only, and $a_1 = \frac{1}{6}\int_X \tau$, where τ is the scalar curvature of X. Note that a_1 is the first *heat invariant* of X, the coefficient in the short time asymptotics of the heat trace:

$$\sum_i e^{-\lambda_i t} \sim \frac{1}{(4\pi)^{n/2}}\sum_{j=0}^\infty a_j\, t^{j-\frac{n}{2}}. \tag{5.1.3}$$

Therefore, if a_1 does not vanish (which is the assumption of Theorem 1.5.1), $\mathcal{R}_2 R(\lambda) >> \lambda^{n-2}$. Combining (5.1.1) and (5.1.2) we get

$$\frac{1}{\lambda}\int_0^\lambda |R(t)|dt > \frac{1}{\lambda}\int_0^\lambda \left(1-\frac{t}{\lambda}\right) R(t)dt = \frac{1}{2}\mathcal{R}_2 R(\lambda) >> \lambda^{n-2}.$$

This completes the proof of Theorem 1.5.1. □

Remark 5.1.4. As follows from results of [Saf], the first Riesz mean of $R(\lambda)$ satisfies

$$\frac{1}{\lambda}\int_0^\lambda R(t)dt = O(\lambda^{n-2}).$$

At the same time, $R(\lambda) = O(\lambda^{n-1})$. Putting together these two upper bounds and the lower bound (1.5.2) one gets an idea about the amount of cancelations occurring when $R(\lambda)$ is integrated over $[0,\lambda]$.

5.2. **Oscillatory error term.** Following [J-P, section 1.2] one may introduce the *oscillatory error term* $R^{osc}(\lambda)$ in Weyl's law:

$$N(\lambda) = \frac{1}{(4\pi)^{n/2}}\sum_{j=0}^{\left[\frac{n-1}{2}\right]} \frac{a_j}{\Gamma\left(\frac{n}{2}-j+1\right)}\lambda^{n-2j} + R^{osc}(\lambda), \tag{5.2.1}$$

where a_j are defined by (5.1.3). The expression (5.2.1) is *not* an asymptotic expansion, however it often appears in physics literature. Such a representation is quite natural since it allows to separate the "mean smooth part" of the counting function coming from the singularity of the heat trace at zero, and the "oscillating part" produced by the singularities in the wave trace caused by closed geodesics.

We believe that using essentially the same arguments as in the proof of Theorem 1.3.1, one can show that the oscillatory error term on an n-dimensional compact negatively curved manifold satisfies:

$$R^{osc}(\lambda) = \Omega\left((\ln\lambda)^{\frac{P(-\mathcal{H}/2)}{h}-\varepsilon}\right) \quad \forall\,\varepsilon > 0. \tag{5.2.2}$$

In order to prove (5.2.2), one has to extend to dimensions $n \geq 3$ the dynamical part of the proof of Theorem 1.3.1, which is easy, and to generalize Theorem 1.4.3, which requires some work. In particular, one needs higher-dimensional analogues of the results of [CdV] that are used in section 3.9. We plan to carry out the details of this argument elsewhere.

Acknowledgments. The authors would like to thank D. Dolgopyat, M. Brin and L. Polterovich for useful comments regarding the proof of Lemma 2.2.1 and Yu. Safarov for suggesting the use of Riesz means in the proof of Theorem 1.5.1. We also thank Y. Colin de Verdière, A. Grigor'yan and M. Shubin for helpful discussions, and the anonymous referee for useful remarks.

Part of this paper was written while the first author visited IHES and Max Planck Institute for Mathematics in Bonn; their hospitality is greatly appreciated.

References

[Ber] P. Berard. *On the wave equation on a compact riemannian manifold without conjugate points.* Math. Z. 155 (1977), 249–276.

[B-G-M] M. Berger, P. Gauduchon and E. Mazet. Le Spectre d'une Variété Riemannienne. Springer Lecture Notes in Mathematics 194, Springer, 1971.

[B-O] R. Bishop and B. O'Neill. *Manifolds of negative curvature.* Trans. Amer. Math. Soc. 145 (1969), 1-49.

[Bow] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms.* Lecture Notes in Math. 470. Springer, 1975.

[B-R] R. Bowen and D. Ruelle. *The ergodic theory of Axiom A flows.* Invent. Math. 29 (1975), no. 3, 181–202.

[Ch] I. Chavel. *Riemannian geometry - a modern introduction.* Cambridge Tracts in Math. 108, Cambridge Univ. Press, 1995.

[CdV] Y. Colin de Verdière. *Spectre du Laplacien et longueurs des géoésiques périodiques I.* Compositio Math. 27 (1973), 83–106.

[D-S] M. Dimassi and J. Sjöstrand. *Spectral asymptotics in the semiclassical limit.* Cambridge University Press, 1999.

[Don] H. Donnelly. *On the wave equation asymptotics of a compact negatively curved surface.* Invent. Math. 45 (1978), 115–137.

[D-G] J. Duistermaat and V. Guillemin. *The spectrum of positive elliptic operators and periodic bicharacteristics.* Inventiones Math. 29 (1975), 39–75.

[Fa] F. Faure, *Semi-classical formula beyond the Ehrenfest time in quantum chaos. (I) Trace formula.* Annales de l'Institut Fourier (to appear).

[Hej] D. Hejhal. *Selberg trace formula for $PSL(2,\mathbb{R})$, Vol. I.* Lecture Notes in Math. 548, Springer, 1976.

[Ho1] L. Hörmander. *The spectral function of an elliptic operator.* Acta Math. 121 (1968), 193–218.

[Ho2] L. Hörmander. *The analysis of linear partial differential operators (Vol. 1)*, Springer-Verlag, 1987.

[J-P] D. Jakobson, I. Polterovich. *Estimates from below for the spectral function and for the remainder in local Weyl's law*, to appear in Geom. Funct. Anal.

[K] A. Karnaukh. *Spectral count on compact negatively curved surfaces.* Ph.D. thesis, Princeton University (1996), 1-48.

[M-S] G. Margulis. *On Some Aspects of the Theory of Anosov Systems*, Springer, 2004. With a survey by Richard Sharp *Periodic Orbits of Hyperbolic Flows.*

[Par] W. Parry. *Equilibrium states and weighted uniform distribution of closed orbits.* Dynamical Systems (College Park, MD 1986-87), Lecture Notes in Math. 1342, 617–625. Springer, 1988.

[P-P] W. Parry and M. Pollicott. *Zeta functions and closed orbit structure for hyperbolic systems.* Asterisque, 187-188 (1990), 1–256.

[P-Rud] R. Phillips and Z. Rudnick. *The circle problem in the hyperbolic plane.* J. Funct. Anal. 121 (1994), no. 1, 78–116.

[Ran] B. Randol. *The Riemann hypothesis for Selberg's zeta-function and the asymptotic behavior of eigenvalues of the Laplace operator.* Trans. Amer. Math. Soc. 236 (1978), 209–223.

[Rub-S] M. Rubinstein and P. Sarnak. Chebyshev's bias. Experiment. Math. 3 (1994), no. 3, 173–197.

[Saf] Yu. Safarov. *Riesz means of the distribution function of the eigenvalues of an elliptic operator.* J. Sov. Math. 49, No.5 (1990), 1210-1212 (1990); translation from Zap. Nauchn. Semin. Leningr. Otd. Mat. Inst. Steklova 163 (1987), 142-145.

[Sar] P. Sarnak. *Arithmetic quantum chaos.* The Schur lectures (1992) (Tel Aviv), 183–236, Israel Math. Conf. Proc., 8, 1995.

[Sin] Y. Sinai. *Gibbs measures in ergodic theory.* Uspehi Mat. Nauk 27 (1972), 4(166), 21–64. Translation: Russian Math. Surveys 27 (1972), 21–69.

[Sj] J. Sjöstrand. *Semi-excited states in nondegenerate potential wells* Asymp. Anal. 6 (1992), 29-43.

[Sun] T. Sunada, *Trace formula for the heat equation for a non-positively curved manifold.* Amer. J. Math. 104 (1982), 795–812.

[Zel1] S. Zelditch. *Lectures on wave invariants*, in: *Spectral theory and geometry*, edited by B. Davies and Y. Safarov, LMS lecture note series 273, Cambridge University Press, 1999.

[Zel2] S. Zelditch. *On the rate of quantum ergodicity I.* Comm. Math. Phys. 160 (1994), no. 1, 81–92.

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Str. West, Montréal QC H3A 2K6, Canada.

E-mail address: jakobson@math.mcgill.ca

Département de mathématiques et de statistique, Université de Montréal CP 6128 succ Centre-Ville, Montréal QC H3C 3J7, Canada.

E-mail address: iossif@dms.umontreal.ca

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Str. West, Montréal QC H3A 2K6, Canada.

E-mail address: jtoth@math.mcgill.ca