

THE INVERSE SPECTRAL PROBLEM FOR SURFACES OF REVOLUTION

STEVE ZELDITCH

Abstract

We prove that isospectral simple analytic surfaces of revolution are isometric.

0. Introduction

This article is concerned with the inverse spectral problem for metrics of revolution on S^2 . We will assume that our metrics are real analytic and belong to a class \mathcal{R}^* of rotationally invariant metrics which are of ‘simple type’ and satisfy some generic non-degeneracy conditions (see Definition (0.1)). In particular, we will assume they satisfy the generalized ‘simple length spectrum’ condition that the length functional on the loop space is a clean Bott-Morse function which takes on distinct values on distinct components of its critical set (up to orientation). Denoting by $Spec(S^2, g)$ the spectrum of the Laplacian Δ_g , our main result is the following:

Theorem I. *Spec: $\mathcal{R}^* \rightarrow \mathbb{R}^{+\mathbb{N}}$ is 1-1.*

Thus, if $(S^2, g), (S^2, h)$ are isospectral surfaces of revolution in \mathcal{R}^* , then g is isometric to h . It would be very desirable to strengthen this result by removing the assumption that $h \in \mathcal{R}^*$, thereby showing that metrics in \mathcal{R}^* are spectrally determined within the entire class of analytic metrics on S^2 with simple length spectra. The only metric on S^2

Received April 7, 1997, and, in revised form, October 17, 1997. Research partially supported by NSF grants #DMS-9404637 and #DMS-9703775.

presently known to be spectrally determined in this sense is the standard one (which is known to be spectrally determined among all C^∞ metrics). A metric h satisfying $\text{Spec}(S^2, h) = \text{Spec}(S^2, g)$ for some $g \in \mathcal{R}^*$ must have many properties in common with a surface of revolution of simple type; see a joint paper (in preparation) by G. Forni and S. Zelditch.

Let us now be more precise about the hypotheses. First, we will assume that there is an effective action of S^1 by isometries of (S^2, g) . The two fixed points will be denoted N, S , and (r, θ) will denote geodesic polar coordinates centered at N , with $\theta = 0$ some fixed meridian γ_M from N to S . The metric may then be written in the form

$$g = dr^2 + a(r)^2 d\theta^2,$$

where $a : [0, L] \rightarrow \mathbb{R}^+$ is defined by $a(r) = \frac{1}{2\pi} |S_r(N)|$, with $|S_r(N)|$ the length of the distance circle of radius r centered at N . For any smooth surface of revolution, the function a satisfies $a^{(2p)}(0) = a^{(2p)}(L) = 0$, $a'(0) = 1$, $a'(L) = -1$ and two such surfaces (S^2, g_i) ($i = 1, 2$) are isometric if and only if $L_1 = L_2$ and $a_1(r) = a_2(r)$ or $a_1(r) = a_2(L - r)$. We will then assume that the metrics belong to the following class \mathcal{R} of *simple analytic* surfaces of revolution [5]:

(0.1) Definition. \mathcal{R} is the moduli space of metrics of revolution (S^2, g) with the properties:

- (i) g (equivalently a) is real analytic;
- (ii) a has precisely one critical point $r_o \in (0, L)$, with $a''(r_o) < 0$, corresponding to an ‘equatorial geodesic’ γ_E ;
- (iii) the (non-linear) Poincaré map \mathcal{P}_{γ_E} for γ_E is of twist type (cf. §1).

We denote by $\mathcal{R}^* \subset \mathcal{R}$ the subset of metrics with ‘simple length spectra’ in the sense above.

Regarding ‘simple length spectra,’ we recall that the closed geodesics of a surface of revolution come in one-parameter families of a common length, filling out invariant torii \mathcal{T} for the geodesic flow. The canonical involution $\sigma(x, \xi) = (x, -\xi)$ takes \mathcal{T} to its ‘time reversal’ $-\mathcal{T}$, and takes the closed geodesics of \mathcal{T} to their reversals on $-\mathcal{T}$. A closed geodesic and its reversal have the same length, so the length spectrum is automatically double except for the length $2L$ of the torus of meridians,

which is σ -invariant. The simple length spectrum hypothesis is that up to time reversal, the common lengths of the closed geodesics on distinct torii are distinct (cf. Definition 1.2.2). In fact, it would be sufficient for the proof of Theorem I that the length $2L$ of the ‘meridian’ closed geodesics is not the length of closed geodesics on any other torus. In any case, it is not hard to show that \mathcal{R}^* is residual in \mathcal{R} (cf. Proposition 1.2.4).

The condition (iii) appears in the proof in the following way: the quadratic coefficient ‘ $\alpha := h''(0)$ ’ of the classical Birkhoff normal form of the metric $|\xi|_g$ at the torus of meridian geodesics must be non-vanishing (see Definition (1.4.5) for the precise meaning of $h(\xi)$). This condition is used in Proposition (4.1.2) and Corollary (4.1.3) to evaluate the wave invariants for the meridian torus.

As will be explained further below, the main purpose of the non-degeneracy and simplicity conditions is to ensure that there are global action-angle variables for the geodesic flow. These conditions rule out several types of surfaces of revolution: First, they rule out Zoll surfaces of revolution, which are degenerate in every possible sense. It is indeed unknown at this time whether real analytic Zoll surfaces of revolution are determined by their spectra. They also rule out ‘peanuts of revolution’ (which have hyperbolic waists) and other natural rotational surfaces such as Liouville torii.

To our knowledge, the strongest prior result on the inverse spectral problem for surfaces of revolution is that of Bruning-Heintze [7]: smooth surfaces of revolution with a mirror symmetry thru the $x - y$ plane are spectrally determined among metrics of this kind. There are also a number of proofs that a surface of revolution is determined by the joint spectrum of Δ and of $\partial/\partial\theta$, the generator of the rotational symmetry [21], [3], [20].

The method-of-proof of Bruning-Heintze was based on the observation that the invariant spectrum can be heard from the entire spectrum. Hence by separating variables the problem can be reduced to the inverse spectral problem for 1D even singular Sturm-Liouville operators, which was solved by Gelfand-Levitan and Marchenko.

Our proof of the Main Theorem is based on different kind of method, and the inverse result presented here is hopefully just one illustration of it. We begin with the facts that simple surfaces of revolution are

completely integrable on both the classical and quantum levels, and that the Laplacian has a global quantum normal form in terms of action operators. We then study the trace of the wave group and prove that from its singularity expansion we can reconstruct the global quantum normal form. Finally we show that this normal form determines the metric.

This approach is suggested by the recent inverse result of Guillemin [17], which shows that the microlocal normal form of Δ around each non-degenerate elliptic closed geodesic can be determined from the wave invariants (see also [25], [26]). However, there is but one non-degenerate closed geodesic on a simple surface of revolution, so the direct application of this inverse result does not take full advantage of the situation. Rather, it is natural to start from the fact that the wave group is completely integrable in the following strong sense: namely, it commutes with an effective action of the torus $S^1 \times S^1$ by Fourier Integral operators on $L^2(S^2)$. That is, there exist global *action operators* \hat{I}_1, \hat{I}_2 and a polyhomogeneous symbol \hat{H} of degree 1 on $\mathbb{R}^2 - 0$ such that $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$. This is the global quantum Birkhoff normal form alluded to above. Our principal tool is the following inverse result:

Main Lemma. *The wave trace invariants of (S^2, g) with $g \in \mathcal{R}^*$ determine the quantum normal form \hat{H} .*

This Lemma does not actually require that (S^2, g) be a surface of revolution, but only that the geodesic flow is toric integrable, i.e., commutes with an effective Hamiltonian torus action. It immediately implies that the principal symbol $H(I_1, I_2)$ of $\hat{H}(\hat{I}_1, \hat{I}_2)$ is a spectral invariant. Since $H(I_1, I_2)$ is essentially a global Birkhoff normal form for the metric, the wave invariants determine the symplectic equivalence of the geodesic flow. Thus we have:

Corollary 1. *From the wave trace invariants of (S^2, g) with $g \in \mathcal{R}^*$ we can determine the symplectic equivalence class of G^t .*

Corollary 1 does not however determine the isometry class of a general $g \in \mathcal{R}^*$: As will be discussed in §1 (see also [12]), simple surfaces of revolution are not symplectically rigid unless they are mirror symmetric. Otherwise put, recovery of the classical Birkhoff normal form only determines the *even part* of the metric in the following sense:

Corollary 2. *A metric $g \in \mathcal{R}$ may be written in the form*

$$g = [f(\cos u)]^2 du^2 + [\sin u]^2 d\theta^2$$

(§1, [4]). *From the wave invariants one can determine the even part of f . In particular if g is mirror symmetric, one can determine g among mirror symmetric metrics in \mathcal{R}^* from its spectrum.*

It is interesting to note that symplectic rigidity in the mirror-symmetric case gives a new and self-contained proof of the Bruning-Heintze result, without the use of Marchenko's inverse spectral theorem for singular Sturm Liouville operators. Although this result is superceded by the Theorem, it may have some future relevance to other inverse problems.

To complete the proof of the Theorem, we therefore have to study the subprincipal terms in \hat{H} . The result is:

Final Lemma. *From \hat{H} one can determine the isometry class of g .*

It is in this last Lemma that we use in full that (S^2, g) is a surface of revolution rather than just a surface with toric integrable geodesic flow. We also use in full that \hat{H} is a global quantum normal form rather than a microlocal one at γ_E . In subsequent work we will investigate the analogues of the Final Lemma for the microlocal normal form at γ_E , for more general toric integrable metrics and metrics isospectral to toric integrable ones.

To close this introduction, we discuss some background and some open problems connected with this work.

The principal motivation for studying the inverse spectral problem for surfaces of revolution is its simplicity. There are to date very few inverse results which determine a metric from its Laplace spectrum within the entire class of metrics, or even within concrete infinite dimensional families. To our knowledge, only the standard S^n for $n \leq 6$ and flat 2-torii are known to be spectrally determined. Hence it is desirable to have a simple model of how an inverse result might go.

A second motivation is a somewhat loose analogy between surfaces of revolution and planar domains. Namely, in both cases the unknown is a function of one variable (the profile curve, resp. the boundary) which completely determines the first return times and angles of geodesics emanating from a transversal. That is, paths of billiard trajectories (broken

geodesics) on a bounded planar domain are determined from collisions with the boundary, while paths of geodesics on surfaces of revolution are determined from collisions with a meridian (or with the equator). In analytic cases, it is plausible that the unknown function may be determined in large part by the spectrum of first return times from the local transversal. In the case of an analytic surface of revolution, this length spectrum determines the corresponding Birkhoff normal form and hence by Corollary 2 it determines the even part of the profile curve. Similarly, in the case of an analytic plane domain, it is proved in [10] that the even part of the boundary of an analytic domain may be determined from the Birkhoff normal form of the billiard map at a bouncing ball orbit. Hence, there is a similarity in the relation between the unknown function and the local classical Birkhoff normal forms. It is interesting to observe in this context that the rigidity result of Colin de Verdiere [10] is quite analogous to the inverse result of Bruning-Heintze.

Some immediate open problems: First, there is the problem mentioned above of removing the assumption that $n \in R^*$. Second, can one relax analyticity to smoothness in the Theorem above? This is likely to follow from a more intensive analysis of the wave invariants. Third, can one extend it to ‘non-simple’ types of surfaces of revolution? The main obstacle is that one will generally not have global action-angle variables or global quantum normal forms. What about completely integrable systems in higher dimensions?

Finally, we would like to thank D.Kosygin for several helpful conversations on this paper, and B.Kleiner for giving us up to date information about the status of the conjugacy problem for geodesic flows on surfaces of revolution.

1. Classical dynamics

1.1. Global action-angle variables.

From a geometric (or dynamical) point of view, the principal virtue of metrics in \mathcal{R} is described by the following:

(1.1.1) Proposition. *Suppose g is a real analytic metric of revolution on S^2 such that a has precisely one (non-degenerate) critical point*

at some $r_o \in (0, L)$. Then the Hamiltonian $|\xi|_g := \sqrt{\sum g^{ij} \xi_i \xi_j}$ on T^*S^2 is completely integrable and possesses global real analytic action-angle variables.

Proof. The complete integrability of $|\xi|_g$ (i.e., of the geodesic flow) is classical, and follows from the existence of the Clairaut integral $p_\theta(v) := \langle v, \frac{\partial}{\partial \theta} \rangle$. Since the Poisson bracket $\{p_\theta, |\xi|_g\} = 0$, the geodesics are constrained to lie on the level sets of p_θ ; and since both $|\xi|_g$ and p_θ are homogeneous of degree one, the behaviour of the geodesic flow is determined by its restriction to $S_g^*S^2 = \{|\xi|_g = 1\}$. With the assumption on a , the level sets are compact, and the only critical level is that of the equatorial geodesics $\gamma_E^\pm \subset S_g^*S^2$ (traversed with either orientation). The other level sets are well-known to consist of two-dimensional torii. (Had we allowed the existence of at least two critical points in a , there would exist a saddle level, i.e., an embedded non-compact cylinder).

The existence of global action-angle variables follows from the general results of [8], [18] and they have been constructed explicitly for simple surfaces of revolution in [9]. The general formula is as follows: Let

$$(1.1.2) \quad \begin{aligned} P &= (|\xi|_g, p_\theta) : T^*S^2 \rightarrow B \\ &:= \{(b_1, b_2) : |b_2| \leq a(r_o)b_1\} \subset \mathbb{R} \times \mathbb{R}^+ \end{aligned}$$

be the moment map of the Hamiltonian \mathbb{R}^2 -action defined by the geodesic flow and rotation. The singular set of P is the closed conic set $Z := \{(r_o, \theta, 0, p_\theta) : \theta \in [0, 2\pi), p_\theta \in \mathbb{R}\}$, i.e., Z is the cone thru the equatorial geodesic (in either orientation). The image of Z is the boundary of B ; the map $P|_{T^*S_gS^2 - Z}$ is a trivial $S^1 \times S^1$ bundle over the open convex cone B_o (the interior of B). For each $b \in B_o$, let $H_1(F_b, \mathbb{Z})$ denote the homology of the fiber $F_b := P^{-1}(b)$. This lattice bundle is trivial since B is contractible, so there exists a smoothly varying homology basis $\{\gamma_1(b), \gamma_2(b)\} \in H_1(F_b, \mathbb{Z})$ which equals the unit cocircle $S_N^*S^2$ together with the fixed closed meridian γ_M when b is on the center line $\mathbb{R}^+ \cdot (1, 0)$. The action variables are given by [9, S 6]

$$(1.1.3) \quad \begin{aligned} I_1(b) &= \frac{1}{2\pi} \int_{\gamma_1(b)} \xi dx = p_\theta, \\ I_2(b) &= \frac{1}{2\pi} \int_{\gamma_2(b)} \xi dx = \frac{1}{\pi} \int_{r_-(b)}^{r_+(b)} \sqrt{b_1^2 - \frac{b_2^2}{a(r)^2}} dr + |b_2|, \end{aligned}$$

where $r_{\pm}(b)$ are the extremal values of r on the annulus $\pi(F_b)$ (with $\pi : S_g^*S^2 \rightarrow S^2$ the standard projection). On the torus of meridians in $S_g^*S^2$, the value of I_2 equals $\frac{L}{\pi}$, and it equals one on the equatorial geodesic. So extended, I_1, I_2 are smooth homogeneous functions of degree 1 on T^*S^2 , and generate 2π -periodic Hamilton flows.

It follows that the pair $\mathcal{I} := (I_1, I_2)$ generates a global Hamiltonian torus $(S^1 \times S^1)$ -action commuting with the geodesic flow. The singular set of \mathcal{I} equals $\mathcal{Z} := \{I_2 = \pm p_\theta\}$, corresponding to the equatorial geodesics. The map

$$(1.1.4) \quad \mathcal{I} : T^*S^2 - \mathcal{Z} \rightarrow \Gamma_o := \{(x, y) \in \mathbb{R} \times \mathbb{R}^+ : |x| < y\}$$

is a trivial torus fibration. Henceforth we often write T_I for the torus $\mathcal{I}^{-1}(I)$ with $I \in \Gamma_o$, and let Γ denote the closure of Γ_o as a convex cone. The symplectically dual angle variables $(\phi_1 = \theta, \phi_2)$ then give, by definition, the flow times (mod 2π) along the orbits of (I_1, I_2) from a fixed point on F_b , which we may take to be the unique point lying above the intersection of the equator and the fixed meridian on F_b with the geodesic pointing into the northern hemisphere.

So far, we have only assumed the metric to be C^∞ . We now observe that if g is real analytic, then so are I_1, I_2 . This is obvious in the case of I_1 and follows from the explicit formula (1.1.3) for I_2 . q.e.d.

Since the metric norm function $|\xi|_g$ commutes with I_1, I_2 , it may be expressed as a function $H(I)$ of the action variables. Hamilton's equations for the geodesic flow in action-angle variables then take the form

$$\dot{I}_k = 0, \quad \dot{\phi}_k = \omega_k(I), \quad (k = 1, 2),$$

where

$$(1.1.5) \quad \omega(I) = \nabla_I H(I)$$

is the *frequency vector* of the torus T_I with action coordinates I . The geodesic flow on T_I is thus given by

$$(1.1.5a) \quad G^t(I, \phi) = (I, \phi + t\omega_I)$$

so that all the geodesics are quasi-periodic in action-angle coordinates (cf. [1, §50]).

The frequency vector ω_I is homogeneous of degree 0 on $T^*S^2 - 0$, and hence is constant on rays of torii \mathbb{R}^+T_I . To break the \mathbb{R}^+ symmetry we restrict to the level set $\{H(I) = 1\} \subset \Gamma_o$ in action space and view the frequency vector as the map:

$$\omega : \{H = 1\} \rightarrow \mathbb{R}^2.$$

Since $\nabla_I H(I) \perp T_I(\{H = 1\})$, the frequency map is more or less the Gauss map of $\{H = 1\}$, although it is not normalized to be of unit length. As a map of the global action space, the frequency map is the Legendre transform associated to H (cf. [15, p.338]).

(1.1.6) Definition. We say that the simple surface of revolution (S^2, g) is *globally non-degenerate* if $\omega|_{\{H=1\}}$ is an embedding.

This is the natural homogeneous analogue of the non-degeneracy condition of [15, loc.cit.] to the effect that the Legendre transform be a global diffeomorphism, and has previously been studied in some detail by Bleher in the setting of simple surfaces of revolution [5]. As will be seen in §1.3, the curve $\{H = 1\}$ is the graph of a smooth function of the form $I_2 = F(I_1)$ in the cone Γ_o , and the non-degeneracy condition (1.1.6) will follow as long as F is a convex or concave function. In fact, in the proof of the Theorem we will only need to use that $\{H = 1\}$ is non-degenerate at the one point $(I_1, I_2) = (0, 1)$. This is sufficient because we assume the metric to be real analytic.

1.2. Length spectrum and periodic torii.

We now come to the definition of length spectrum and simple length spectrum for a completely integrable geodesic flow. We first observe that the orbit thru (I, ϕ) is periodic of period L if and only if

$$(1.2.1a) \quad L\omega_I = M \in \mathbb{Z}^2$$

for some $M \neq 0$. The minimal positive such L will be called the primitive period; the corresponding M is known as the vector of winding numbers of the torus T_I . M parametrizes the homology class of the closed orbit γ since the latter has the form $\sum_{j=1}^2 M_j \gamma_j(I)$ relative to the homology basis $\gamma_j(I)$.

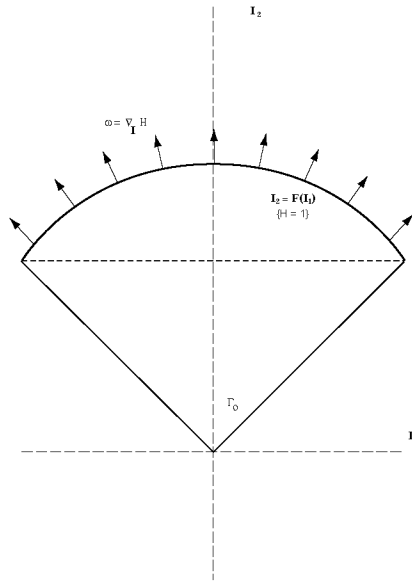


FIGURE 1.

Due to the homogeneity, the period and vector of winding numbers are constant on the ray $\mathbb{R}T_I$. By Euler's formula we then have

$$\nabla_I H \cdot I = \omega_I \cdot I = H,$$

hence the length is given in terms of the winding vector by

$$(1.2.1b) \quad L = \frac{M \cdot I}{H(I)}$$

or simply $L = M \cdot I$ on the unit tangent bundle $H = 1$.

It is clear that the periodicity condition $L\omega = M$ is independent of ϕ . Hence, all of the geodesics on T_I are closed if any of them are. (This also follows, of course, from the transitivity of the torus action on each invariant torus.) We therefore have:

(1.2.2) Definition.

- (a) A torus T_I is a *periodic torus* if all the geodesics on it are closed.
- (b) The *period* L of the periodic torus is then the common period of its closed geodesics.
- (c) The *length spectrum* \mathcal{L} of the completely integrable system is the set of these lengths.

(d) The completely integrable system has a *simple length spectrum* if there exist a unique periodic torus (up to time reversal) of each length $L \in \mathcal{L}$.

In the last statement (d) we are referring to the canonical involution $\sigma : (x, \xi) \rightarrow (x, -\xi)$, which reverses the orientation of the geodesics. It is obvious that if T_I is a periodic torus of period L , then so is $\sigma(T_I)$.

Let $Per \subset S_g^*S^2 - 0$ denote the set of periodic points for the geodesic flow on $S_g^*S^2$, i.e., the set of points which lie on a closed geodesic. It is a union of periodic torii in $S_g^*S^2$ together with points along the equatorial geodesics (which are degenerate torii). The set of all periodic points in $T^*S^2 - 0$ is then equal to $\mathbb{R}^+ Per$. Since the invariant torii are parametrized by the points $I \in \Gamma$ of action space, it is convenient to parametrize Per by a subset of the level set $\{H(I) = 1\} \subset \Gamma$.

(1.2.3) Definition. The set of points $I \in \{H = 1\} \subset \Gamma$ such that $T_I \subset Per$ will be called, with a slight abuse of notation, the set of *periodic* points on $\{H = 1\}$ and will be denoted by \mathcal{P} . That is, $\mathcal{P} = \{I \in \{H = 1\} : \exists L \in \mathbb{R}^+, L\omega_I \in \mathbb{Z}^2\}$.

The following proposition will be needed later on (Proposition (4.1.4)).

(1.2.4) Proposition. *If (S^2, g) is non-degenerate (1.1.6), then \mathcal{P} is dense in $\{H = 1\}$.*

Proof. Let $\mathcal{Q} := \omega(\mathcal{P})$ equal the image of \mathcal{P} under the frequency map ω . Then by definition, \mathcal{Q} (for ‘rational points’) is the projection to the curve $\omega(\{H = 1\})$ of the integer lattice in \mathbb{Z}^2 . It is clear that \mathcal{Q} is a dense set in $\omega(\{H = 1\})$; since ω is an embedding, \mathcal{P} is dense in $\{H = 1\}$. q.e.d.

The next proposition shows that our inverse result is valid for a residual set of simple analytic surfaces of revolution.

(1.2.5) Proposition. *Let $\mathcal{R}^* \subset \mathcal{R}$ be the subset of metrics with simple length spectra. Then \mathcal{R}^* is a residual subset of \mathcal{R} .*

Proof. If $L, L' \in \mathcal{L}$, then there exist $M, M' \in \mathbb{Z}^2$ and $I, I' \in \mathcal{Q}$ such that $L = M \cdot I, L' = M' \cdot I'$. So $L = L'$ implies that $M \cdot I - M' \cdot I' = 0$ hence that the I -coordinates of I, I' are dependent over the rationals. Since the length spectrum moves continuously and non-trivially under

deformations in \mathcal{R} , such a dependence for fixed M, M' can only hold on a closed nowhere dense set. The proposition follows. q.e.d.

1.3. First return times and angles.

Let us consider more carefully the geometric interpretation of the Clairaut integral on a torus $T_I \subset S_g^*S^2$. Since I_1 and $H = |\xi|$ are independent commuting coordinates, and since there are global action-angle variables, the different invariant torii in $S_g^*S^2$ are parametrized by the values $I_1 = \iota \in [-1, 1]$ of the Clairaut integral along $\{H = 1\}$. Indeed by (1.1.3), the second action coordinate I_2 is determined from I_1 on $\{H = 1\}$ by the formula

$$(1.3.1) \quad I_2 = F(I_1) := |I_1| + \frac{1}{\pi} \int_{r_-(I_1)}^{r_+(I_1)} \sqrt{1 - \frac{I_1^2}{a(r)^2}} dr,$$

where $r_{\pm}(I_1)$ are the roots of $I_1^2 = a(r)^2$. The torus in $S_g^*S^2$ with $I_1 = \iota$ is therefore $T_{(\iota, F(\iota))}$, which we will denote simply by T_{ι} . The projection of T_{ι} to S^2 is an annulus of the form $r_+(\iota) < r < r_-(\iota)$. The geodesics on T_I project to S^2 as almost periodic curves oscillating between the two extremal parallels $r = r_+(\iota)$ and $r = r_-(\iota)$.

We observe then that T_{ι} contains a unique geodesic γ_{ι} which passes thru the intersection of the reference geodesics γ_M and γ_E ; ι equals $a(r_o)\cos\alpha$ where $\alpha(\iota)$ is the angle between γ_{ι} and γ_M . In other words, $\alpha(\iota)$ is the common angle with which the geodesics on T_{ι} intersect the meridians as they pass thru the equator in the direction of the northern hemisphere. Since the length L_E of the equator γ_E equals $2\pi\sqrt{a(r_o)}$, and since this length is a symplectic invariant of the geodesic flow, the coordinates ι and α are related in an essentially universal fashion. Hence, either I_1 or $\alpha \in (0, \pi)$ could be used as an action coordinate on $\{H = 1\}$; α is perhaps more geometric, but I_1 is more convenient in calculations.

The picture is the same for any invariant torus T_I : Under the \mathbb{R}^+ action on $T^*S^2 - 0$, it scales to a torus $T_{\iota} \subset S^*S^2$ and all features of its geometry are identical to that of T_{ι} . Thus it carries a unique (parametrized) geodesic γ_I such that $\gamma_I(0)$ is at the intersection $\gamma_E \cap \gamma_M$, and so that $\gamma_I'(0)$ points to the northern hemisphere. The initial angle variables of $(\gamma_I(0), \gamma_I'(0))$ are therefore $(\phi_1(0), \phi_2(0)) = (0, 0)$. At time

t we denote the angle variables of $(\gamma_I(t), \gamma'_I(t))$ by $(\phi_1(t), \phi_2(t))$ where ϕ_1 measures the equatorial angle, and ϕ_2 measures a kind of meridional angle. We then introduce the following ‘first return times’:

(1.3.2) Definition.

(Ei) The equatorial first return time is the minimal time $\tau_E(I) > 0$ such that $\phi_2(\tau_E(I)) = 2\pi$.

(Eii) The equatorial first return angle $\omega_E(I) := \phi_1(\tau_E(I))$ is the change in angle along the equator of a geodesic on T_I leaving $\gamma_E \cap \gamma_M$ at $t = 0$, upon its first return time to γ_E .

(Mi) The meridional first return time is the minimal time $\tau_M(I)$ such that $\phi_1(\tau_M(I)) = 2\pi$.

(Mii) The meridional first return angle is the angle change $\omega_M(I) := \phi_2(\tau_M(I))$ along the meridian of a geodesic on γ_M leaving $\gamma_E \cap \gamma_M$ at $t = 0$, upon its first return time to γ_M .

The terminology ‘first return time’ is taken from dynamics. Note that $\phi_2 \equiv 0 \pmod{2\pi}$ is the equation of the curve on T_I which lies over the equator. Hence, $\tau_E(I)$ is the time of first return of $\gamma_I(t)$ to the equator in the direction of the northern hemisphere. This is actually the second time of intersection of γ_I with the equator, the first one occurring when $\gamma_I(t)$ is heading to the southern hemisphere. This intersection is not in the projection of $\phi_2 \equiv 0 \pmod{2\pi}$. Similarly, $\phi_1 \equiv 0 \pmod{2\pi}$ is the equation of the fixed meridian γ_M , and so τ_M is the time of first return to the arc γ_M (half of the closed geodesic).

The following gives some relations between the various angles and return times.

(1.3.3) Proposition. *Let $\omega_I = (\omega_1, \omega_2)$ be the frequency vector of the invariant torus T_I . Then:*

- (a) $\tau_E \omega_1 = \omega_E$;
- (b) $\tau_M \omega_2 = \omega_M$;
- (c) $2\pi \frac{\omega_2(I)}{\omega_1(I)} = \omega_M$, $2\pi \frac{\omega_1(I)}{\omega_2(I)} = \omega_E$.

Proof. The equation of the geodesic $(\gamma_I(t), \gamma'_I(t))$ on T_I beginning at (ϕ_1, ϕ_2) in angle variables is $(\phi_1 + t\omega_1(I), \phi_2 + t\omega_2(I))$. By definition of return times,

$$\begin{aligned} (\phi_1 + \tau_E \omega_1, \phi_2 + \tau_E \omega_2) &= (\phi_1 + \omega_E, \phi_2 + 2\pi) \\ (\phi_1 + \tau_M \omega_1, \phi_2 + \tau_M \omega_2) &= (\phi_1 + 2\pi, \phi_2 + \omega_M). \end{aligned}$$

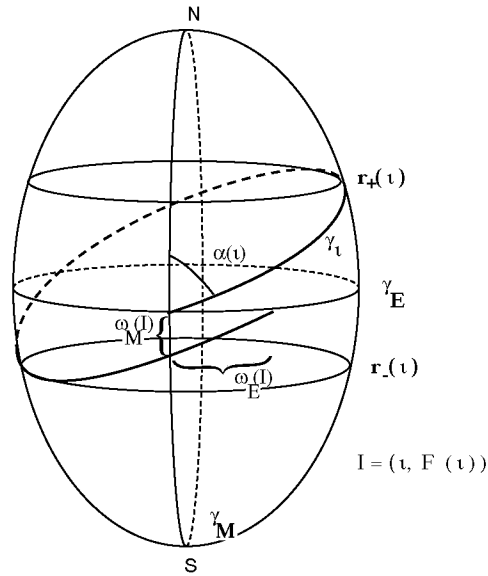


FIGURE 2.

The statements in the proposition follow immediately. \square

These first return times (and angles) are closely related to the (non-linear) Poincaré maps of the geodesic flow. We recall that for each closed geodesic γ , the Poincaré map \mathcal{P}_γ is defined as the first return map of the geodesic flow, restricted to a symplectic transversal $S_\gamma \subset S_g^*S^2$ (surface of section). It is well-known that $\mathcal{P}_\gamma : S_\gamma \rightarrow S_\gamma$ is a symplectic map [22]. Since the symplectic form on a cotangent bundle equals $d\alpha$ (with α the action form), it follows that $\mathcal{P}_\gamma^*(\alpha) - \alpha$ is a closed 1-form on S_γ . Since S_γ may be assumed contractible, there exists a function τ_γ such that $[\mathcal{P}_\gamma^*(\alpha) - \alpha]|_{S_\gamma} = d\tau_\gamma$. Since the integral of α over an arc of a unit speed geodesic just gives its length, τ_γ is the first return time of geodesics near γ to S_γ .

In particular, let $\gamma = \gamma_E$ be the equator (in one of its orientations), and let S_E denote a symplectic transversal at the point $(\gamma_E(0), \gamma_E'(0))$ thru the fixed meridian γ_M . Since γ_M is transverse to the equator, we may define S_E to consist of a small variation of $\gamma_E'(0)$ moved up and down a small arc of γ_M . We see then that the first return time τ_{γ_E} is precisely the first return time τ_M defined above. We also observe that the foliation of S^*S^2 by invariant torii restricts to a foliation of

S_E by invariant circles for \mathcal{P}_{γ_E} , closing in on the center point where γ_E intersects S_E . As noted above, the action coordinate $I_1 = \iota$ gives a natural action (radial) coordinate on S_γ . Since $S_\gamma \subset S_g^*S^2$, the \mathbb{R}^+ homogeneity is broken and we may reformulate the twist condition (0.1 (iii)) as follows:

(1.3.4) Proposition. *The Poincaré map \mathcal{P}_{γ_E} is a twist map of S_E iff $\omega'_M \neq 0$, where $\omega'_M = \frac{\partial}{\partial \iota} \omega_M$.*

Proof. The coordinates (I_2, ϕ_2) restrict to a system of symplectic coordinates on S_E , in terms of which the Poincaré map takes the form

$$\mathcal{P}_{\gamma_E}(I_2, \phi_2) = (I_2, \phi_2 + \tau_M \omega_2) = (I_2, \phi_2 + \omega_M).$$

By definition it is a twist map if $\omega'_M \neq 0$ in a neighborhood of $\iota = 0$ in S_E . q.e.d.

For background on the twist condition in a related context see [15], [24].

The situation for the other periodic orbits is different since they come in one-parameter families. Thus, for the (closed) meridian geodesic γ_M , a transversal S_M is given by the equator γ_E and a small variation of $\gamma'_M(0)$ along it. The foliation by invariant torii restricts to S_M to a foliation by invariant lines (non-closed curves), including the curve of closed geodesics thru γ_M . The Poincaré map \mathcal{P}_{γ_M} is then of parabolic type; the first return time is τ_E above.

For the inverse problem it will be necessary to have expressions for these return times and angles in terms of the metric. This will also make the twist condition more transparent. For ease of quotation, it is convenient to make a change of dependent and independent variables, following [4] and Darboux [8]. Equivalent expressions in the original polar coordinates can also be easily derived, and will be given below.

(1.3.5) Proposition. *Suppose that (S^2, g) is a simple surface of revolution. Then there exists a coordinate system (u, θ) on U , with $\sin u = a(r)$ and a smooth function f on $[-1, 1]$ such that $f(1) = 1, f(-1) = 1$ and such that*

$$g = [f(\cos u)]^2 du^2 + \sin^2 u d\theta^2.$$

Proof. First, define

$$b(r) = \begin{cases} \sin^{-1}(a(r)), & r \in [0, r_o], \\ \pi - \sin^{-1}(a(r)), & r \in [r_o, L], \end{cases}$$

and

$$c(v) = \begin{cases} (a|_{[0, r_o]})^{-1}(\sqrt{1-v^2}), & v \in [0, 1], \\ (a|_{[r_o, L]})^{-1}(\sqrt{1-v^2}), & v \in [-1, 0]. \end{cases}$$

Then define $f : (-1, 1) \rightarrow \mathbb{R}$ by

$$\begin{cases} f(v) = \frac{v}{a'[c(v)]}, & (v \neq 0), \\ f(0) = \frac{1}{-a''(r_o)}. \end{cases}$$

Since $b(r) = u, c(\cos u) = b^{-1}(u) = r$, we have $a(r) = \sin u$, $a'[c(\cos u)]dr = \cos u du$, and $b'(r)dr = du$. The smoothness properties of f follow from those of a and the fact that $a(r)$ and $\sin u$ have the same qualitative shapes [4, loc.cit.]. q.e.d.

The geometric result is the following [4, Theorem 4.11]:

(1.3.6) Proposition. *In the above coordinates, the equator $\gamma_E(s) = (u(s), \theta(s))$ has the equation*

$$u(s) \equiv \frac{\pi}{2}, \quad \theta(s) = s.$$

Any other geodesic $\gamma(s) = (u(s), \theta(s))$ in U is contained between two parallels $u = i$ and $u = \pi - i$, and the angle $\theta(i)$ between two consecutive points of contact with these parallels is given by:

$$\theta(i) = \sin i \int_i^{\pi-i} \frac{f(\cos u)}{\sin u (\sin^2 u - \sin^2 i)^{\frac{1}{2}}} du.$$

The length of this arc of γ is given by

$$s(i) = \int_i^{\pi-i} \frac{\sin(u) f(\cos u)}{(\sin^2 u - \sin^2 i)^{\frac{1}{2}}} du.$$

Sketch of Proof. Using the Clairaut integral, the equations of the geodesic have the form:

$$\begin{cases} \frac{d\theta}{ds} = \frac{\sin i}{\sin^2 u}, \\ \frac{d\theta}{du} = \frac{\sin(i) f(\cos u)}{\sin u (\sin^2 u - \sin^2 i)^{\frac{1}{2}}}, \\ \frac{ds}{du} = \frac{\sin(u) f(\cos u)}{(\sin^2 u - \sin^2 i)^{\frac{1}{2}}}. \end{cases}$$

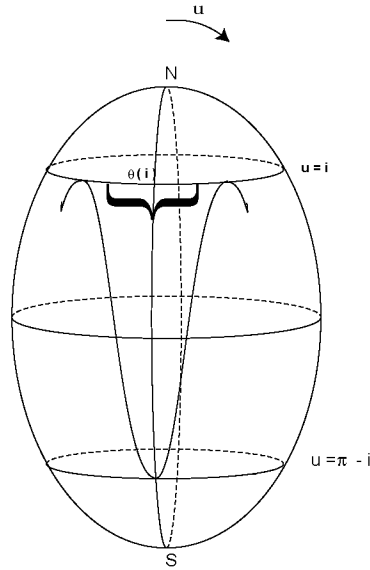


FIGURE 3.

The formulae above follow by integration. q.e.d.

We then have:

(1.3.7) Proposition. *Let $T_I \subset S_g^* S^2$ denote an invariant torus with $H(I) = 1$, and let $i(I)$ be the u -coordinate of the extremal parallel closest to N in the projection of T_I to S^2 . Then:*

- (i) $\tau_E(I) = 2s(i(I));$
- (ii) $\omega_E(I) = 2(\theta(i(I)) - \pi);$
- (iii) $\tau_E(I) = \frac{1}{\pi} \int_{r_-(I_1)}^{r_+(I_1)} (1 - \frac{I_1^2}{a(r)^2})^{-\frac{1}{2}} dr$
- (iv) $\frac{1}{2\pi} \omega_E(I) = -1 + \frac{I_1}{\pi} \int_{r_-(I_1)}^{r_+(I_1)} a(r)^{-2} (1 - \frac{I_1^2}{a(r)^2})^{-\frac{1}{2}} dr, (I_1 > 0).$

Proof.

(i) By definition, $\omega_E(I)$ is the change in angle along the equator between a geodesic γ_I on T_I , starting at the equator on a fixed meridian and heading towards the northern hemisphere, and the fixed meridian, upon second intersection with the equator. We can view this arc of the geodesic as consisting of three pieces: one from the equator to the

northern extremal parallel, one on the ‘back-side’ between the two extremal parallels, and one on the ‘front-side’ from the southern parallel to the equator. Since the lengths of the ‘front-side’ arcs are unchanged by rotation, we can rotate one until the two make up a smooth geodesic arc between the parallels. The length of the geodesic arc is therefore twice that of an arc between the parallels, i.e., $\tau_E(I) = 2s(i(I))$.

(ii) For the angle change: In the same way, the change in θ along this arc of γ_I is the change in θ of two arcs between the extremal parallels. We subtract π since ω_I measures the addition to one full revolution.

(iii) Since $\sin(u) = a(r)$ and $f(\cos u)du = dr$, we have

$$s(i(I)) = \int_{i(I)}^{\pi-i(I)} \frac{dr}{\sqrt{1 - \frac{I_1^2}{a(r)^2}}}.$$

(iv) From Propositions 1.3.3 we have that $\frac{\omega_1}{\omega_2} = \frac{1}{2\pi}\omega_E$. Since $H(I_1, I_2) = 1$ implies that $\omega_1 + F'(I_1)\omega_2 = 0$, we get from (1.3.1) that $\omega_E = -2\pi F'(I_1)$. q.e.d.

(1.3.8) Corollary. *The non-degeneracy condition (1.1.6) is satisfied if \mathcal{P}_{γ_E} is globally twisted, i.e., if $\omega'_M > 0$ or $\omega'_M < 0$ or equivalently of $\omega'_E > 0$ or $\omega'_E < 0$.*

Proof. In these cases, F is convex (or concave). Since the set $\{H = 1\}$ is the graph of F in Γ_o , the Gauss map (and hence the frequency map) is an embedding. q.e.d.

Remark. Both cases of concavity and convexity, occur for ellipsoids of revolution; see [5]. The separating case of the round sphere is of course degenerate.

1.4. Classical Birkhoff invariants.

The classical Birkhoff normal form of Hamiltonian H near a non-degenerate periodic orbit γ is a germ of a completely integrable system to which H is symplectically equivalent in a ‘formal neighborhood’ of γ (see e.g. [15] for a detailed discussion). In the case at hand, where H is already completely integrable, the Birkhoff normal form is simply H itself expressed in terms of action-angle variables. The ‘Birkhoff invariants’ of the Hamiltonian at a torus T_{I_o} may then be identified with the germ of $H(I_1, I_2)$ or of ω_I at $I = I_o$.

Since H is homogeneous of degree 1, it is equivalent and somewhat clearer to define the Birkhoff invariants after first breaking the \mathbb{R}^+ symmetry. That is, we would like to introduce a ‘base’ to the cone Γ_o . The most natural one may appear to be the energy level $\{H = 1\}$; but for the purpose of calculating wave invariants at T_{I_o} it is more convenient to use the tangent line $\omega_{I_o} \cdot (I - I_o) = 0$ at a point $I_o \in \{H = 1\}$.

Let us first consider the level set $\{H = 1\}$ as the transversal. The homogeneous function H is obviously determined by the curve $H(I_1, I_2) = 1$ whose equation is given by (1.3.1). Hence we can define the Birkhoff invariants at a torus T_{I_o} to be the Taylor coefficients of the function F' at $I_o \in \{H = 1\}$. By (1.3.7 (iv)) it is equivalent to put:

(1.4.1) Definition. The (first) Birkhoff invariants of H at an invariant torus T_ι with $\iota \in \{H = 1\}$ are the Taylor coefficients of ω_E at ι in the coordinate I_1 .

Secondly, let us consider the tangent lines as transversals: We fix a point $I^o \in \{H = 1\}$, let ω^o denote the common frequency vector of the ray of torii $\mathbb{R}^+T_{I^o}$ and put

$$(1.4.2) \quad I \cdot \omega^o := \sum_{k=1}^2 \omega_k^o I_k.$$

The equation of the tangent line to $\{H = 1\}$ at I^o in the action cone Γ_o is then given by $\omega^o \cdot I = 1$. Note that $I \cdot \omega^o$ is homogeneous of degree 1 and hence equals $H(I)$ along the ray $\mathbb{R}I^o$; consequently it is elliptic (non-vanishing) in a conic neighborhood $W_o \subset \Gamma_o$ of it. The conic neighborhood will be parametrized in the following way: we fix a basis (i.e., a non-zero vector) v of the line $I \cdot \omega^o = 0$, and define the map

$$(1.4.3a) \quad (\rho, \xi) \rightarrow \rho(I^o + \xi v), \quad \xi \in (-\epsilon, \epsilon).$$

For sufficiently small ϵ , this map sweeps out a conic neighborhood W_o of I^o with inverse given by

$$(1.4.3b) \quad \rho = \omega^o \cdot I, \quad \xi v_j := \frac{I_j}{I \cdot \omega^o} - I_j^o.$$

Since

$$H(I_1, I_2) = (\omega^o \cdot I) \ H\left(\frac{I_1}{\omega^o \cdot I}, \frac{I_2}{\omega^o \cdot I}\right)$$

and since

$$\left(\frac{I_1}{\omega^o \cdot I}, \frac{I_2}{\omega^o \cdot I}\right) \in \{\omega^o \cdot I = 1\}$$

we may write

$$(1.4.4a) \quad H(I_1, I_2) = \rho h_{I^o}(\xi),$$

where h_{I^o} is the function on $W_o \cap \{\omega^o \cdot I = 1\}$ defined by

$$(1.4.4b) \quad h_{I^o}(\xi) := H(I^o + \xi v).$$

The C^∞ Taylor expansion of $h_{I^o}(\xi)$ around $\xi = 0$ is then a symplectic invariant of H .

(1.4.5) Definition. The second (tangential) classical Birkhoff invariants of H associated to the periodic torus T_{I^o} are the Taylor coefficients $h_{I^o}^\alpha(0)$.

In the real analytic case the Taylor coefficients determine h_{I^o} and hence H by homogeneity. It is more or less obvious that the first and second Birkhoff invariants also carry the same information in the C^∞ case. To be sure, we prove:

(1.4.6) Proposition. *The first Birkhoff invariants canonically determine the tangential Birkhoff invariants and vice versa.*

Proof. By definition

$$h_{I^o}(\xi) = H(I^o + \xi v).$$

Therefore,

$$H(h_{I^o}(\xi)^{-1}(I^o + \xi v)) = 1$$

or equivalently

$$h_{I^o}(\xi)^{-1}(I_2^o + \xi v_2) = F(h_{I^o}(\xi)^{-1}(\xi v_1 + I_1^o)).$$

Writing $u = \frac{I_1^o + \xi v_1}{h_{I^o}}$, this says

$$\frac{F(u)}{u} = -\frac{I_2^o + \xi v_2}{I_1^o + \xi v_1}.$$

Hence the knowledge of Taylor coefficients of F is equivalent to the knowledge of the Taylor coefficients of h_{I^o} . q.e.d.

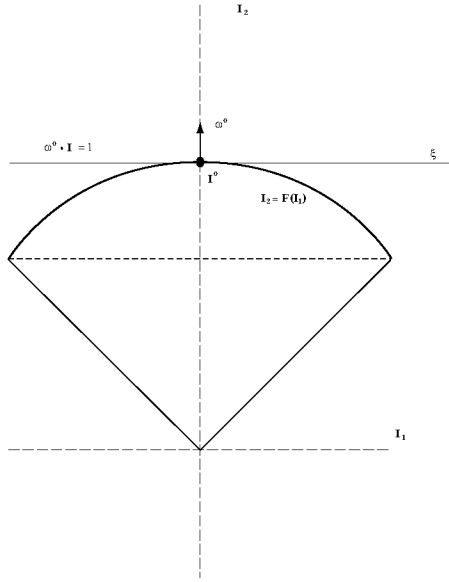


FIGURE 4.

We may reformulate the non-degeneracy condition (0.1 (iii)) in terms of h_{I^o} where T_{I^o} is the meridian torus:

(1.4.7) Proposition. *(S^2, g) satisfies the non-degeneracy condition (0.1 (iii)) as long as $\alpha := h''_{I^o}(0) \neq 0$.*

Proof. By definition, $h_{I^o}(\xi) = H(I^o + \xi v)$. At the meridian torus, $v = (1, 0)$ so

$$h_{I^o}(\xi) = H(I^o + (\xi, 0))$$

and

$$h'_{I^o}(\xi) = \frac{\partial}{\partial I_1} H(I^o + (\xi, 0)) = \omega_1(I^o + (\xi, 0)).$$

Hence

$$h''_{I^o}(0) = \frac{\partial}{\partial I_1} \omega_1(I^o).$$

Since $\omega_1(I^o) = 0$ it follows that

$$\frac{\partial}{\partial I_1} \omega_E(I^o) = \frac{\omega'_1(I^o)}{\omega_2(I^o)}.$$

Hence, the condition $h''_{I^o}(0) \neq 0$ is equivalent to the condition that $\frac{\partial}{\partial I_1} \omega_E(I^o) \neq 0$. The equivalence of this to (0.1 (iii)) is proved in Propositions (1.3.4) and again in Corollary (1.3.8). q.e.d.

1.5. Symplectic conjugacy of geodesic flows.

The Birkhoff normal form of a Hamiltonian H at a closed orbit (or family of closed orbits) is a symplectic conjugacy invariant of H in a neighborhood of the orbit(s). Hence the global Birkhoff normal form $H(I_1, I_2)$ of a completely integrable Hamiltonian is a symplectic conjugacy invariant. The purpose of this section is to show that it is a complete conjugacy invariant. We begin by showing that the homogeneous Hamiltonian torus actions commuting with geodesic flows of simple surfaces of revolution are all symplectically equivalent:

(1.5.1) Proposition. *Suppose (S^2, g_1) and (S^2, g_2) are smooth surfaces of revolution of simple type, and let (I_1, I_2) resp. (J_1, J_2) denote their global action variables as above. Then there exists a homogeneous symplectic diffeomorphism $\chi : T^*S^2 \rightarrow T^*S^2$ such that $\chi^*J_i = I_i$.*

Sketch of Proof. Let ϕ_i (resp. ψ_i) be the dual angle variables on (S^2, g_1) (resp. (S^2, g_2)). Except on the equators of (S^2, g_i) a point in T^*S^2 is uniquely specified by its action-angle coordinates. Define χ to be the identity map in action-angle coordinates with respect to the two metrics. It is obvious that χ is a homogeneous symplectic diffeomorphism on the complement of the equators, so to prove the Corollary it suffices to show that χ extends to the equators with this property.

Since χ is homogeneous of degree 1, it is necessary and sufficient to define it on the unit cotangent bundles. Moreover, since it commutes with the Hamilton flow of p_θ on the regular set, its extension must also do so. Hence it must be the lift of a diffeomorphism $\bar{\chi}$ on the orbit space $\mathcal{O} := S^*S^2/S^1$ of the rotation action. This action is free, so the natural projection $p : S^*_g S^2 \rightarrow \mathcal{O}$ must be diffeomorphic to the standard projection from $S^0(3) \rightarrow S^2$. The image of the torus foliation defined by level sets of p_θ is a singular foliation of \mathcal{O} formed by level sets of the function \bar{p}_θ induced by p_θ on \mathcal{O} , and the two singular points o^\pm are the images of the equators γ_E^\pm . Since \bar{p}_θ is, by assumption, a perfect Morse function on $\mathcal{O} \sim S^2$ for each metric, the quotient map $\bar{\chi}$ on the punctured quotient $\mathcal{O} - \{o^\pm\}$ of $\bar{\chi}$ extends smoothly to the completion. It follows that χ extends smoothly as a rotationally equivariant map on the completion of $S^*S^2 - \{\gamma_E^\pm\}$, and the homogeneous extension must be symplectic. (See [9] for more on the behaviour near the poles). q.e.d.

Remark. In fact, this proposition can be sharpened to say: there exists only one homogeneous Hamiltonian $S^1 \times S^1$ action on $T^*S^2 - 0$ (up to symplectic equivalence). This follows from a homogeneous analogue of the Delzant classification of completely integrable torus action on compact Kähler manifolds. We hope to give more general results of this kind at a later time.

Thus, the torus actions defined above on the cotangent bundles of simple surfaces of revolution are always symplectically equivalent. The question arises when the geodesic flows are symplectically equivalent. The answer can be given by expressing the norm functions $|\xi|_g$ of the metrics in terms of the global action variables. Before doing so, we note that the action variables are not quite uniquely defined above because the choice of generators $\gamma_i(b)$ is not unique. For instance, one might have permuted the roles of N and S . Hence we have:

(1.5.2) Proposition. *Let (S^2, g_i) be a simple surfaces of revolution, and let (I_1, I_2) resp. (J_1, J_2) be the global action variables defined above. Let $|\xi|_{g_1} = H_1(I_1, I_2)$ resp. $|\xi|_{g_2} = H_2(J_1, J_2)$ be the expressions of the metric norms of g_i in terms of action variables. Then the geodesic flows of (S^2, g_i) are homogeneously symplectically equivalent if and only if there a linear map $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in SL(2, \mathbb{Z})$ such that $H_1 = H_2 \circ A$.*

Proof. If such a choice exists, then the map χ above obviously defines a symplectic conjugacy.

Conversely, suppose the geodesic flows are symplectically conjugate by a homogeneous symplectic diffeomorphism χ , i.e., $\chi^*H_2(J_1, J_2) = H_1(I_1, I_2)$. Then χ^*J_i are global action variables for the geodesic flow of (S^2, g_1) . But global action variables are almost unique; they correspond to a trivialization of the lattice bundle $H_1(F_b, \mathbb{Z})$ [8], [18]. Therefore there exists $A \in SL(2, \mathbb{Z})$ so that $\chi^*J = A \cdot I$. Hence $H_1(I) = H_2(A \cdot I)$.
q.e.d.

The following gives a geometric interpretation of the conjugacy condition:

(1.5.3) Proposition. *Suppose $g_1, g_2 \in \mathcal{R}$. Then their geodesic flows are symplectically equivalent if and only if their equatorial first return times τ_E and angles ω_E are equal.*

Proof. Suppose first that the flows are symplectically equivalent. After choosing compatible bases for the homology, we may then assume that the expressions for H_1 and H_2 in global action-angle variables are the same. Thus the frequency maps are the same, and by Proposition (1.3.3) the equatorial return angles are the same. Also, the Poincaré maps \mathcal{P}_{γ_E} are conjugate, and hence the equatorial return times are equal.

Conversely, if the equatorial return times and angles are the same, then the flows have the same frequency vectors (as functions of the global action angle variables) and thus have the same global Birkhoff normal forms. By Proposition (1.5.2) the flows are symplectically equivalent.

q.e.d.

2. Quantum dynamics and normal form

It is owing to the following notion that simple surfaces of revolution are so manageable.

(2.1) Definition. The wave group $e^{it\sqrt{\Delta}}$ of a compact, Riemannian n -manifold (M, g) is *quantum torus integrable* if there exists a unitary Fourier-Integral representation

$$\hat{\tau} : T^n \rightarrow U(L^2(M)), \quad \hat{\tau}_{(t_1, \dots, t_n)} = e^{i(t_1 \hat{I}_1 + \dots + t_n \hat{I}_n)}$$

of the n -torus and a symbol $\hat{H} \in S^1(\mathbb{R}^n - 0)$ such that

$$\sqrt{\Delta} = \hat{H}(I_1, \dots, I_n).$$

In the above formula, we follow physics notation in indicating operators (as opposed to their symbols) with a ‘hat’. Thus, the generators \hat{I}_j are first order pseudodifferential operators with the property that $e^{2\pi i \hat{I}_j} = C_j Id$ for some constant C_j of modulus one.

Since \hat{H} is a first order elliptic symbol on $\mathbb{R}^n - 0$, it has an asymptotic expansion in homogeneous functions of the form:

$$(2.2) \quad \hat{H} \sim H_1 + H_0 + H_{-1} + \dots, \quad H_j(rI) = r^j H_j(I).$$

The quantum action operators are uniquely defined up to a choice of basis of $H^1(\mathbb{R}^n/\mathbb{Z}^n, \mathbb{Z})$, and the terms H_j are uniquely determined up to the same ambiguity.

The principal symbols I_j of the \hat{I}_j 's generate a classical Hamiltonian torus action, so any quantum torus action automatically induces a classical one. Conversely, it is a theorem of Boutet de Monvel-Guillemin and Weinstein that any classical Hamiltonian torus action can be quantized [6, Appendix, Proposition 6.6]. Since metrics in \mathcal{R} give rise to quantum torus actions, we have, in particular:

(2.3) Proposition (cf. [9]). *Suppose $g \in \mathcal{R}$. Then $\sqrt{\Delta_g}$ is quantum torus integrable.*

We also observe that the following holds for any Laplacian commuting with a quantum torus action:

(2.4) Proposition. *For any Δ , $H_o = 0$.*

Proof. Since the subprincipal symbol $\sigma_{sub}(\sqrt{\Delta})$ equals zero, the same is true; $\hat{H} := \hat{H}(\hat{I}_1, \hat{I}_1)$. Now $\sigma_{sub}(\hat{H})$ is invariantly defined (independent of a choice of symplectic coordinates); hence it may be expressed in action-angle coordinates in the form

$$0 = \sigma_{sub}(\hat{H}) = H_o - \frac{1}{i} \sum_j \frac{\partial^2 H_1}{\partial I_j \partial \phi_j}.$$

But the mixed derivative term automatically vanishes since H_1 is a function only of the classical action variables. q.e.d.

Let us specialize to the case of $\sqrt{\Delta_g}$ with $g \in \mathcal{R}$. From the fact that $e^{2\pi i \hat{I}_j} = C_j Id$ for a quantum torus action, it follows that the joint spectrum of the quantum moment map

$$Sp(\mathcal{I}) \subset \mathbb{Z}^2 \cap \Gamma + \{\mu\}$$

is the set of integral lattice points, translated by μ , in the closed convex conic subset $\Gamma \subset \mathbb{R}^2$. The vector $\mu = (\mu_1, \mu_2)$ can be identified with the Maslov indices of the homology basis of the invariant torii. In the case of $\sqrt{\Delta_g}$, $\mu = (0, 1/2)$ [9].

Expressing $\sqrt{\Delta_g}$ in the form $\hat{H}(\hat{I}_1, \hat{I}_2)$ we have that

$$Sp(\sqrt{\Delta_g}) = \{\hat{H}(N + \mu) : N \in \mathbb{Z}^2 \cap \Gamma_o\}.$$

Thus, the eigenvalues of $\sqrt{\Delta_g}$ take the form:

$$\lambda_N \sim H_1(N + \mu) + H_{-1}(N + \mu) + \dots$$

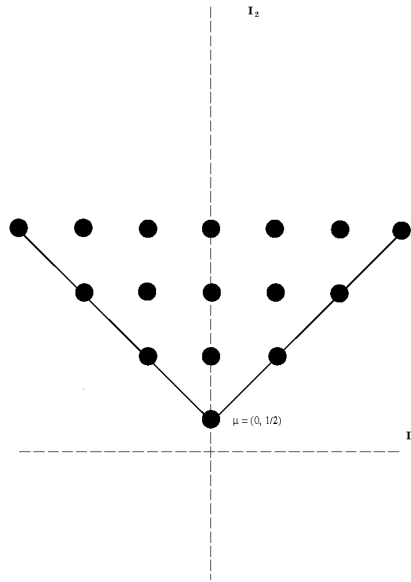


FIGURE 5.

and the wave trace is

$$(2.5) \quad \text{Tr} e^{it\sqrt{\Delta_g}} = \sum_{N \in \mathbb{Z}^2} e^{it\hat{H}(N+\mu)}.$$

The symbol \hat{H} may be regarded as a global quantum Birkhoff normal form. As in the classical case, it has a germ at any periodic torus, so these may be regarded as the microlocal Birkhoff canonical forms. To be more precise, we imitate the definition of the classical tangential Birkhoff normal forms and write

$$(2.6) \quad H_j(I_1, I_2) = (\omega^o \cdot I)^j H_j\left(\frac{I_1}{\omega^o \cdot I}, \frac{I_2}{\omega^o \cdot I}\right) := (\omega^o \cdot I)^j h_j(\xi),$$

where ξ is a linear coordinate relative to a vector v generating $\omega^o \cdot I = 1$. We then Taylor expand $h_j(\xi)$ around $\xi = 0$:

$$h_j(\xi) = \sum_{\alpha \geq 0} h_j^\alpha(0) \xi^\alpha.$$

(2.7) Definition. The homogeneous quantum Birkhoff normal form coefficients of \hat{H} at the periodic torus T_{I^o} are the Taylor coefficients h_j^α for $j = 1, 0, -1, \dots$

(2.8) Proposition. *Assume that $g \in \mathcal{R}^*$. Then all of the functions H_j are real analytic in the interior of the cone Γ_o and all of the functions h_j are real analytic near $\xi = 0$.*

Proof. First consider $|\xi|_g = H_1(I_1, I_2) = H(I_1, I_2)$. We know that H is a C^∞ homogeneous function on Γ_o . On the other hand, from the formula $I_2 = G(|\xi|_g, I_1)$ we see that I_2 is a real analytic function of $|\xi|_g, I_1$. Since G is the inverse function to H with respect to the first variable, H must also be a real analytic function of I_1, I_2 .

Next recall that \hat{I}_2 is defined (up to a smoothing term) as the function $\hat{G}(\sqrt{\Delta}, \hat{I}_1)$ which has principal symbol $G(|\xi|_g, I_1)$ and which has integral spectrum. More precisely, one begins with $G(\sqrt{\Delta}, \hat{I}_1)$, which has the property that $\exp(2\pi i G(\sqrt{\Delta}, \hat{I}_1) + \mu_j) = I + K$ with K of order -1 . One defines $R = \frac{-1}{2\pi i} \text{Log}(I + K)$ and puts $\hat{G}(\sqrt{\Delta}, \hat{I}_1) = G(\sqrt{\Delta}, \hat{I}_1) + \mu_j + R_j$.

Since G is a real analytic function and we only apply the holomorphic functional calculus in the steps of the construction, it follows that \hat{G} is an analytic function of $(\sqrt{\Delta}, \hat{I}_1)$.

Since $\hat{I}_2 = \hat{G}(\sqrt{\Delta}, \hat{I}_1)$ has the inverse function $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$, from the inverse function theorem for analytic functions it follows again that \hat{H} is a real analytic function.

The real analyticity of the h_j 's follows from that of the H_j 's. q.e.d.

3. Wave invariants as non-commutative residues

To relate the wave invariants to the coefficient of the normal form, it will be helpful (as in [17], [25], [26]) to express the wave invariants as non-commutative residues of the wave operator and its time-derivatives. Let us recollect how this goes.

The non-commutative residue of a Fourier Integral operator is an extension of the well-known non-commutative residue of a pseudodifferential operator A on a compact manifold M , defined by

$$\text{res}(A) = 2 \text{Res}_{s=0} \zeta(s, A),$$

where

$$\zeta(s, A) = \text{Tr } A \Delta^{-s/2} \quad (\text{Re } s \gg 0).$$

Here, Δ is a Laplacian (or any positive elliptic operator) on M . From the work of Seeley, Wodzicki and Guillemin, one knows that $\zeta(s, A)$ is holomorphic in $\operatorname{Re} s > \frac{1}{2} \dim M + \operatorname{ord}(A)$ and admits a meromorphic continuation to \mathbb{C} , with simple poles at

$$s = \dim M + \operatorname{ord}(A) - k \quad (k = 0, 1, 2, \dots).$$

The residue at $s = 0$ has a number of remarkable properties (not shared by the residues of the other poles):

- $\operatorname{res}(AB) = \operatorname{res}(BA)$, i.e., res is a trace on the algebra $\Psi^*(M)$ of pseudodifferential operators over M ;
- $\operatorname{res}(A)$ is independent of the choice of $\sqrt{\Delta}$;
- there is a local formula for the residue,

$$\operatorname{res}(A) = (2\pi)^{-n} \int_{S^*M} a_{-n}(x, \xi) i_{\mathcal{R}} dx \wedge d\xi,$$

where a_{-n} is the term of degree $(-n)$ in the complete symbol expansion $a \sim \sum_{-\infty}^m a_j$ for A ; $dx \wedge d\xi$ is the canonical symplectic volume measure on T^*M .

These results may be extended to Fourier integral operators as follows: Let A be a Fourier Integral operator in $I^m(M \times M, \Lambda)$ for some homogeneous canonical relation $\Lambda \subset T^*(M \times M) \setminus 0$ and $m \in \mathbb{Z}$. Also, let $\operatorname{diag}(X \times X)$ denote the diagonal in $X \times X$. Below we will sketch a proof of the following:

(3.1) Theorem. *Suppose Λ and $\operatorname{diag}(T^*M \times T^*M)$ intersect cleanly. Then $\zeta(s, A) = \operatorname{Tr} A \Delta^{-s/2}$ ($\operatorname{Re} s \gg 0$) has a meromorphic continuation to \mathbb{C} , with simple poles at $s = m + 1 + \frac{e_0 - 1}{2} - j$, where*

$$e_0 = \dim \Lambda \cap \operatorname{diag}(S^*M \times S^*M),$$

and $j = 0, 1, 2, \dots$.

The clean intersection hypothesis above is that

$$\Lambda \cap \operatorname{diag}(T^*M \times T^*M) \setminus 0$$

is a clean intersection. It is satisfied in the case where $\Lambda = C_t$ if and only if the fixed point set of G^t is clean. Hence, Theorem (3.1) implies:

(3.2) Corollary. *Let $\zeta(s, t) = \text{Tr } U(t)\Delta^{-s/2}$. If the fixed point set of G^t is clean, then $\zeta(\cdot, t)$ has a meromorphic continuation to \mathbb{C} , with simple poles at*

$$s = 1 + \frac{\dim S \text{Fix}(G^t) - 1}{2} - j \quad (j = 0, 1, \dots, 2).$$

Here, $S \text{Fix}(G^t)$ is the set of unit vectors in the fixed point set of G^t . In the case of a completely integrable system, $S \text{Fix}(G^L)$ is the union of the periodic torii with period L . We assume here, and henceforth, that the periodic torii are all clean fixed point sets for G^t on S^*M .

The non-commutative residue of the Fourier integral operator A is then defined by:

(3.3) Definition.

$$\text{res}(A) := \text{Res}_{s=0} \zeta(s, A)$$

just as in the case of pseudodifferential operators. And just as in that case, $\text{res}(A)$ has some remarkable properties:

- it is independent of the choice of Δ ;
- if either A or B is associated to a local canonical graph, then $\text{res}(AB) = \text{res}(BA)$;
- there is a local formula for $\text{res}(A)$.

The basic properties of $\text{res}(A)$ may be deduced from a singularity analysis of the closely related distribution trace $S(t, A) := \text{Tr}AU(t)$ (cf. [28]). Under the cleanliness assumption above, $S(t, A)$ is a Lagrangean distribution on \mathbb{R} with singularities at the set of ‘sojourn times’,

$$\mathcal{ST} = \{T : \exists(x, \xi) \in S^*M : (x, \xi, G^T(x, \xi)) \in \Lambda\}.$$

For a given sojourn time T , the corresponding set

$$W_T = \{(x, \xi) : (x, \xi, G^T(x, \xi)) \in \Lambda\}$$

of sojourn rays fills out a submanifold of S^*M of some dimension e_T . With $N(\lambda, A) = \sum_{\sqrt{\lambda_j} \leq \lambda} (A\varphi_j, \varphi_j)$, we have

(3.4) Proposition [28, Proposition 1.10]. *If $\rho \in C^\infty(\mathbb{R})$ with $\hat{\rho} \in C_0^\infty(\mathbb{R})$, $\text{supp } (\hat{\rho}) \cap \mathcal{ST} = \{0\}$ and $\hat{\rho} \equiv 1$ near 0, then*

$$\rho * dN(\lambda, A) \sim C_n \lambda^{m + \frac{e_0 - 1}{2}} \sum \alpha_j \lambda^{-j},$$

where C_n is a universal constant. The coefficients have the form,

$$\alpha_j = \int_{\Lambda_\Delta} \omega_j d\lambda_\Delta,$$

where $\Lambda_\Delta = \Lambda \cap \text{diag}(S^*M \times S^*M)$, $e_0 = \dim \Lambda_\Delta$, $d\lambda_\Delta$ is a canonical density on Λ_Δ , and the functions ω_j are determined by the j -jet of the (local) complete symbols of AU along Λ_Δ .

Proof of Theorem (3.1). As in the pseudodifferential case [14, Proposition 2.1], we have

$$\begin{aligned} \text{Tr } A\Delta^{-s/2} &= \langle \chi_s, dN(\cdot, A) \rangle \\ &= \langle \chi_s, \rho * dN(\cdot, A) \rangle + \text{entire.} \end{aligned}$$

Hence,

$$\begin{aligned} \zeta(s, A) &= C_n \sum_{j=0}^{\infty} \alpha_j \int_1^{\infty} \lambda^{\frac{e_0-1}{2}+m-s-j} d\lambda + \text{entire} \\ &= C_n \sum_{j=0}^{\infty} \frac{\alpha_j}{m+1+\frac{e_0-1}{2}-(s+j)} + \text{entire}, \end{aligned}$$

completing the proof. q.e.d.

In particular, if $A = U(L)$, then $\{0\}$ is a sojourn time if and only if $L \in \text{Lsp}(M, g)$. If $L \notin \text{Lsp}(M, g)$, $\zeta(s, L)$ is regular at 0.

Now let us return to the case where the geodesic flow is completely integrable. Then the dimension of each periodic torus \mathcal{T} of period L equals $e_o = \dim \mathcal{T} = n$. Hence we have:

$$(3.5a) \quad \text{Tr}U(t) = e_o(t) + \sum_{\mathcal{T}} e_{\mathcal{T}}(t),$$

where the sum runs over the periodic tori in S^*M , and

$$(3.5b) \quad \begin{aligned} e_{\mathcal{T}}(t) &= a_{\mathcal{T};-\frac{n+1}{2}}(t-L+i0)^{-\frac{n+1}{2}} \\ &+ a_{\mathcal{T};-\frac{n+1}{2}+1}(t-L+i0)^{-\frac{n+1}{2}+1} + \dots \end{aligned}$$

More precisely, it takes this form if n is even; if n is odd, the positive powers of $(t-L+i0)$ should be multiplied by $\log(t-L+i0)$. In the following Corollary we use the notation $-\mathcal{T}$ for the time reverse torus $\sigma\mathcal{T}$.

(3.6) Corollary.

$$\sum_{\pm} a_{\pm\mathcal{T}, -(\frac{n+1}{2})+k} = \text{res}(\sqrt{\Delta}^{-\frac{n+1}{2}+k} U(t)|_{t=L}).$$

Proof. In the notation of (possibly negative) fractional derivatives in t , we have $\Delta^\mu U(L) = D_t^\mu U(t)|_{t=L} =$. The claim follows from the facts that

$$D_t^{-\frac{n+1}{2}+k} (t - L + i0)^{-\frac{n+1}{2}+k} = \log(t - L + i0),$$

and that the non-commutative residue is the coefficient of $\log(t - L + i0)$.
q.e.d.

(3.7) Examples.

(a) If $\dim M = 2$, then the wave trace expansion at a torus \mathcal{T} has the form

$$e_{\mathcal{T}}(t) = a_{\mathcal{T}, -\frac{3}{2}}(t - L + i0)^{-\frac{3}{2}} + a_{\mathcal{T}, -\frac{1}{2}}(t - L + i0)^{-\frac{1}{2}} + \dots$$

Thus

$$\sum_{\pm} a_{\pm\mathcal{T}, -\frac{3}{2}} = \text{res}\sqrt{\Delta}^{-\frac{3}{2}} e^{iL\sqrt{\Delta}}, \quad a_{\mathcal{T}, -\frac{1}{2}} = \text{res}\sqrt{\Delta}^{-\frac{1}{2}} e^{iL\sqrt{\Delta}}, \dots$$

(b) If $\dim M = 3$, the wave trace expansion at \mathcal{T} takes the form

$$e_{\mathcal{T}}(t) = a_{\mathcal{T}, -2}(t - L + i0)^{-2} + a_{\mathcal{T}, -1}(t - L + i0)^{-1} + a_{\mathcal{T}, 0}\log(t - L + i0) + \dots$$

Hence

$$\sum_{\pm} a_{\pm\mathcal{T}, -2} = \text{res}\Delta^{-2} e^{iL\sqrt{\Delta}}, \quad a_{\mathcal{T}, -1} = \text{res}\Delta^{-1} e^{iL\sqrt{\Delta}},$$

$$\sum_{\pm} a_{\pm\mathcal{T}, 0} = \text{res}e^{iL\sqrt{\Delta}}, \dots$$

(3.8) Remark. The wave invariants for a closed geodesic γ (or periodic torus T_L) are exactly the same as for its time reversal, hence the same residue formulae also give the individual wave invariants. For this reason it is often not necessary to resolve the ambiguity between a torus and its time reversal.

4. Proof of the Main Lemma

We now prove that the wave trace invariants of a metric $g \in \mathcal{R}^*$ determine its quantum normal form \hat{H} .

4.1. Wave invariants for simple surfaces of revolution.

We begin by specializing Corollary (3.6) to the case of a surface of revolution (S^2, g) in \mathcal{R}^* . Since the length spectrum \mathcal{L} is simple, there is a unique periodic torus $\mathcal{T}_L \subset S_g^*S^2$ of each length $L \in \mathcal{L}$ (up to time reversal). By the existence of global action-angle coordinates, it may be expressed in the form $\mathcal{T}_L = T_{I_L}$ for a unique point $I_L = (I_{L1}, I_{L2}) \in \mathcal{P}$ (up to reflection). Let ω_L denote the frequency vector at I_L . Then we have $L \cdot \omega_L = M_L$ where M_L is the vector of winding numbers of the periodic orbits on \mathcal{T}_L .

The other periodic torii of period L lie on the rays $\mathbb{R}^+ \mathcal{T}_L \cup \mathbb{R}^+ \sigma(\mathcal{T}_L)$. Their action coordinates lie on the rays $\mathbb{R}^+ I_L \cup \mathbb{R}^+ \sigma(I_L)$, and they have the same frequency vector, ω_L , as for I_L . Since the wave invariants at \mathcal{T}_L depend only on the microlocalization of the wave group to a conic neighborhood of $\mathbb{R}^+ \mathcal{T}_L$, we introduce a microlocal cut off operator $\hat{\psi}_L(\hat{I}_1, \hat{I}_2)$ with ψ homogeneous of degree 0, equal to 1 in a small conic neighborhood of the ray $\mathbb{R}^+ I_L$ and zero off of a slightly larger conic neighborhood. The singularity of $TrU(t)$ at $t = L$ is then the same as the singularity of $Tr\hat{\psi}_L U(t)$.

To calculate the wave trace as a residue, we also introduce the first order pseudodifferential operator $\omega_L \cdot \hat{I} := \omega_{L1} \hat{I}_1 + \omega_{L2} \hat{I}_2$. We emphasize that $\omega_L \cdot \hat{I}$ is a linear combination with constant coefficients of the action operators. Since $\omega_I \cdot I = H$, the principal symbol $\omega_L \cdot I$ of $\omega_L \cdot \hat{I}$ takes the value 1 at $I = I_L$ and therefore $\omega_L \cdot \hat{I}$ is elliptic in a conic neighborhood of \mathcal{T}_L . We will use it as the gauging elliptic operator in the residue formula for the wave invariants.

(4.1.1) Proposition. *Let $g \in \mathcal{R}^*$ and let $L \in \mathcal{L}$. Then we have:*

$$\begin{aligned} & \sum_{\pm} a_{\pm T_L, -\frac{1}{2}+k} \\ &= Res_{s=0} \int_{\Gamma} \psi_L(I + \mu) e^{i\langle M_L, I \rangle} e^{-iL\hat{H}(I+\mu)} \\ & \quad \cdot (\hat{H}(I + \mu))^{-\frac{1}{2}+k} (\omega_L \cdot (I + \mu))^{-s} dI, \end{aligned}$$

where as above M_L is the vector of winding numbers of \mathcal{T}_L .

Proof. Since $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$ and $\hat{\psi}_L$ is a function of the action operators, we have by Corollary (3.6) that

$$\begin{aligned} & \sum_{\pm} a_{\pm T_L, -\frac{1}{2}+k} \\ &= \text{Res}_{s=0} \sum_{N \in \mathbb{Z}^2} \psi_L(N + \mu) e^{iL\hat{H}(N+\mu)} (\hat{H}(N + \mu))^{-\frac{1}{2}+k} (\omega_L \cdot (N + \mu))^{-s}. \end{aligned}$$

We then apply the Poisson summation formula for $\text{Re } s \gg 0$ to replace the sum over $N \in \mathbb{Z}^2$ by

$$\text{Res}_{s=0} \sum_{M \in \mathbb{Z}^2} J_{L,M,k}(s),$$

where

$$\begin{aligned} & J_{L,M,k}(s) \\ &:= \int_{\Gamma} \psi_L(I + \mu) e^{-i\langle M, I \rangle} e^{iL\hat{H}(I+\mu)} (\hat{H}(I + \mu))^{-\frac{1}{2}+k} (\omega_L \cdot (I + \mu))^{-s} dI. \end{aligned}$$

From Theorem 3.1 and by the simplicity of the length spectrum it follows that only the term with $M = M_L$ has a pole at $s = 0$. This can be seen more directly from the fact that only in this term does the phase $-\langle M, I \rangle + LH(I)$ have a critical point, since $M = L\nabla_I H(I_M)$ implies that the torus with actions I_M is periodic of period L (cf. §1.2). q.e.d.

To calculate the residue of the integral $J_{Lk}(s) := J_{L,M_L,k}(s)$, we rewrite the integrals in terms of the (ρ, ξ) coordinates introduced in (1.4.3) in a conic neighborhood of the point $I_L \in \{H = 1\}$: Recall that we parametrize the tangent line $T_{I_L}(\{H = 1\})$ by

$$\xi \in \mathbb{R} \rightarrow I_L + \xi v,$$

where v is a non-zero vector along the line $M_L \cdot I = 0$, and parametrize a conic neighborhood of $\mathbb{R}^+ I_L$ by

$$(\rho, \xi) \in \mathbb{R}^+ \times \mathbb{R} \rightarrow \rho(I_L + \xi v).$$

Since I_L is fixed, we abbreviate h_{I_L} by h in the next proposition. We also denote the cutoff in these coordinates by $\psi_L(\xi)$.

(4.1.2) Proposition. *With the above notation, $a_{I_L, -\frac{1}{2}+k}$ equals the term of order ρ^{-1-k} in the asymptotic expansion of the integral*

$$J_{Lk}(\rho) := L^{-2+s-k+\frac{1}{2}} e^{i\langle M, \mu \rangle} \rho^{\frac{1}{2}} \int_{\mathbb{R}} \psi_L(\xi) e^{i\rho\alpha\frac{\xi^2}{2}} e^{i\rho g_3} e^{i\sum_{k=1}^{\infty} L^{k+1} \rho^{-k} h_{-k}(\xi)} [h(\xi) + \sum_{\ell=1}^{\infty} L^{\ell+1} \rho^{-\ell-1} h_{\ell}(\xi)]^{-\frac{1}{2}+k} d\xi.$$

Here, $\alpha = h''(0)$, and g_3 is the third order remainder in the Taylor expansion of $h(\xi)$ at $\xi = 0$.

Proof. We first change variables $I + \mu \rightarrow I$ in the expression for $J_{L, M_L, k}(s)$ in the preceding Lemma and then further change variables to $(\rho := \omega_M \cdot I, \xi)$. Thus we get

$$J_{Lk}(s) = e^{i\langle M, \mu \rangle} \int_0^{\infty} \int_{\mathbb{R}} \psi_L(\xi) e^{-i\rho\langle M, (I_L + \xi v) \rangle} e^{iL[\rho h(\xi) + \rho^{-1} h_{-1}(\xi) + \dots]} \cdot [\rho h(\xi) + \rho^{-1} h_{-1}(\xi) + \dots]^{-\frac{1}{2}+k} \rho^{-s+1} d\rho d\xi.$$

Taylor expanding $h(\xi) = h(0) + h'(0)\xi + \frac{1}{2}h''(0)\xi^2 + g_3(\xi)$ and using that $h(0) = H(I_L) = 1$, that $h'(0) = \omega_{I_L} \cdot v = 0$ and that $\langle M, I_L + \xi v \rangle = \langle M, I_L \rangle = L$ by (1.2.1a-b), we get

$$J_{Lk}(s) = e^{i\langle M, \mu \rangle} \int_0^{\infty} \int_{\mathbb{R}} \psi_L(\xi) e^{iL\rho h''(0)\frac{1}{2}\xi^2} e^{iL\rho g_3(\xi)} e^{iL[\rho^{-1} h_{-1}(\xi) + \dots]} \cdot [\rho h(\xi) + \rho^{-1} h_{-1}(\xi) + \dots]^{-\frac{1}{2}+k} \rho^{-s+1} d\rho d\xi.$$

Now change variables again, $\rho \rightarrow L^{-1}\rho$ and pull the factor ρL^{-1} in front of $h(\xi)$ in the bracketed expression outside the $d\xi$ -integral. We get:

$$\begin{aligned} & J_{Lk}(s) \\ &= e^{i\langle M, \mu \rangle} L^{-2+s-k+\frac{1}{2}} \int_0^{\infty} \left\{ \int_{\mathbb{R}} \psi_L(\xi) e^{i\rho h''(0)\frac{1}{2}\xi^2} e^{i\rho g_3(\xi)} e^{iL[\rho^{-1} h_{-1}(\xi) + \dots]} \cdot [h(\xi) + \rho^{-2} L^2 h_{-1}(\xi) + \dots]^{-\frac{1}{2}+k} d\xi \right\} \rho^{-s+1+k-\frac{1}{2}} d\rho. \end{aligned}$$

The pole at $s = 0$ is produced by the terms of order ρ^{-1} in the $d\xi - d\rho$ -integrals, hence by the terms of order $-k - 1 - \frac{1}{2}$ in the asymptotic expansion of the $d\xi$ -integral. q.e.d.

To determine the terms of order ρ^{-k-1} in $J_{L^k}(\rho)$ we apply the method of stationary phase. Cancelling factors of $\rho^{\frac{1}{2}}$ we get:

(4.1.3) Corollary. *In the above notation, $a_{T_L, -\frac{1}{2}+k}$ equals the term of order ρ^{-k-1} in the asymptotic expansion of*

$$\frac{1}{\sqrt{2\pi i\alpha}} L^{-2-k+\frac{1}{2}} e^{i\langle M, \mu \rangle} \sum_{m=0}^{\infty} (2i\rho)^{-m} \alpha^{-m} \partial_{\xi}^{2m} \left\{ e^{i\rho g_3(\xi)} e^{i\sum_{j=1}^{\infty} L^{j+1} \rho^{-j} h_{-j}(\xi)} \cdot \left[h(\xi) + \sum_{\ell=1}^{\infty} L^{\ell+1} \rho^{-\ell-1} h_{-\ell}(\xi) \right]^{-\frac{1}{2}+k} \right\} \Big|_{\xi=0}.$$

4.2. The meridian torus.

We now want to analyse the terms of the expansion coming from the ray of meridian torii with $I_1 = 0$. The special property of the meridian torii is that they are invariant under the canonical involution $\sigma(x, \xi) = (x, -\xi)$ of $T^*S^2 - 0$. As the following proposition shows, this leads to a useful symmetry property of the quantum normal form. As above, the notation $h(\xi), h_{-1}(\xi)$ etc. refer to the functions h_{I_L} (etc.) where I_L is the point on $\{H = 1\}$ corresponding to the meridian torus in $S_g^*S^2$.

(4.2.1) Proposition. *The complete symbol of \hat{H} is invariant under σ . Hence h_j is even for all j .*

Proof. Let C denote the operator of complex conjugation: $C\psi = \bar{\psi}$. Since $\sqrt{\Delta}$ commutes with C , so does $\hat{H}(\hat{I})$. Since C is a conjugate-linear involution, this implies that

$$\hat{H}(\hat{I}) = C^{-1} \hat{H}(\hat{I}) C = \bar{\hat{H}}(C^{-1} \hat{I} C),$$

where the bar denotes complex conjugation.

We claim that $C^{-1} \hat{I}_1 C = -\hat{I}_1$ and that $C^{-1} \hat{I}_2 C = \hat{I}_1$. Moreover, that $I_1 \circ \sigma = -I_1, I_2 \circ \sigma = I_2$. The statements regarding I_1, \hat{I}_1 are obvious since $\hat{I}_1 = \frac{\partial}{i\partial\theta}$ changes sign under complex conjugation and $\sigma_{C^{-1} \hat{I}_1 C} = I_1 \cdot \sigma$. This latter also follows from the fact $I_1(x, \xi) = x_2 \xi_1 - x_1 \xi_2$.

Regarding I_2 we note that its σ -invariance follows immediately from the explicit formula (1.3.1). The non-obvious claim is that \hat{I}_2 is invariant under C . But from the invariance of the principal symbol we have

$$C^{-1} \hat{I}_2 C = \hat{I}_2 + K_1,$$

where K_1 is of order 0. Since \hat{I}_2 is a function of $(\hat{I}_1, \sqrt{\Delta})$, it is clear that C must take joint (\hat{I}_1, \hat{I}_2) eigenfunctions into joint eigenfunctions. Hence C determines an involution on the joint spectrum $\{(n, k + \frac{1}{2}) : |n| \leq 2k + 1, k \geq 0\}$ which, we recall, is simple. Consequently, the involution (still denoted C) must take the form

$$C : (n, k + \frac{1}{2}) \rightarrow (-n, k + \frac{1}{2} + f(n, k)),$$

where f is a bounded function. Moreover, since $f(n, k) = \langle \hat{I}_2 \overline{\phi_{n,k}}, \overline{\phi_{n,k}} \rangle$ and the $\phi_{n,k}$ are quasi-modes associated to Bohr-Sommerfeld-Maslov torii [11], it is clear that $f(n, k + \frac{1}{2})$ must be a polyhomogeneous function of order 0 on Γ_0 . Since it is also integral-valued on the semi-lattice of joint spectral points, it must be constant. Additionally it must satisfy the involution condition, and one sees that the constant must be 0.

Returning to $\hat{H}(\hat{I}_1, \hat{I}_2)$ we have

$$\hat{H}(\hat{I}_1, \hat{I}_2) = C^{-1} \hat{H}(\hat{I}_1, \hat{I}_2) C = \tilde{\hat{H}}(C^{-1} \hat{I}_1 C, C^{-1} \hat{I}_2 C) = \tilde{\hat{H}}(-\hat{I}_1, \bar{I}_2),$$

so that $\hat{H}(a, b) = \tilde{\hat{H}}(-a, b)$. We then observe that \hat{H} is a real function (at least modulo terms of order $-\infty$). To see this, we recall that the eigenvalues $\hat{H}(n, k + \frac{1}{2})$ are real and argue by induction on the symbol expansion using that

$$\hat{H}(n, k + \frac{1}{2}) = H(n, k + \frac{1}{2}) + H_{-1}(n, k + \frac{1}{2}) + \dots \in \mathbb{R}.$$

First, the principal symbol H is real so we may drop it from the expansion without affecting reality. Assuming that $H = H_1, H_{-1}, \dots, H_{-k}$ are real, we may drop them all and get that the tail sum is real. Since it is dominated by H_{-k-1} , this function must be real for all points $(n, k + \frac{1}{2})$ sufficiently far from 0. But by homogeneity, H_{-k-1} is then real on the projection of these lattice points to $H = 1$. The projected points form a dense set by Proposition 1.2.4 and hence H_{-k-1} is everywhere real. We conclude that \hat{H} is σ -invariant.

Now consider the ray of meridian torii, or more precisely the ray $\mathbb{R}^+(0, 1)$ in the action cone Γ . We note that this ray is invariant under the involution $\sigma_\Gamma(a, b) := (-a, b)$ of Γ , and moreover the meridian torus is invariant under σ (it ‘rotates’ the torus by angle π). From the above, the level set $\{H = 1\}$ is invariant under σ_Γ and hence the tangent line

at $(0, 1)$ is invariant. Evidently it is horizontal in the (I_1, I_2) -plane, and σ_Γ restricts to it to the map $\xi \rightarrow -\xi$. Since the complete symbol of \hat{H} is σ_Γ -invariant, the h_j 's must be even. q.e.d.

We now go back to the wave invariants associated to the meridian torus and its iterates. Let us write the amplitude for the k th wave invariant of the iterate of length L , namely

$$(4.2.2) \quad \begin{aligned} A_{Lk}(\xi, \rho) := & e^{i \sum_{j=1}^{\infty} L^{j+1} \rho^{-j} h_{-j}(\xi)} [h(\xi) \\ & + \sum_{\ell=1}^{\infty} L^{\ell+1} \rho^{-\ell-1} h_{-\ell}(\xi)]^{-\frac{1}{2}+k}, \end{aligned}$$

in the form

$$A_{Lk}(\xi, \rho) := A_{Lk0}(\xi) + \rho^{-1} A_{Lk1}(\xi) + \dots$$

Thus, $a_{TL, -\frac{1}{2}+k}$ is the term of order ρ^{-k-1} in

$$(4.2.3) \quad \frac{L^{-k-1}}{\sqrt{2\pi i L \alpha}} \sum_{m=0}^{\infty} \sum_{\ell=0}^{\infty} \alpha^{-m} (2i\rho)^{-m-\ell} \partial_\xi^{2m} [e^{i\rho g_3} A_{Lk\ell}].$$

Expanding the derivatives and using that $g_3(\xi)$ is even and of order $0(\xi^4)$, we may rewrite (4.2.3) in the form

$$(4.2.4) \quad \frac{L^{-k-1}}{\sqrt{2\pi i L \alpha}} \sum_{m=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{j=0}^m \sum_{q \leq \frac{1}{2}m} \sum_{(i_1, \dots, i_q): |i|=j} \{C_{mli} \alpha^{-m} \rho^{-m-\ell+q} \cdot (\partial_\xi^{2i_1} g_3 \dots \partial_\xi^{2i_q} g_3) \partial_\xi^{2(m-j)} A_{Lk\ell}\}.$$

Here, the multi-index i -sum runs over q -tuplets with $i_n \geq 2, |i| = \sum i_n = j \leq m$ with $q \leq \frac{1}{2}m, 2q \leq |i| = j \leq m$. In order that $-m-\ell+q = -k-1$ it is necessary that

$$m \leq 2(k+1), \quad \ell \leq k+1.$$

These restrictions follow from the fact that it takes at least 4 derivatives of g_3 to make a non-zero contribution. The coefficients C_{mli} are universal constants.

4.3. A collection of formulae.

For future reference we assemble some notation and identities regarding the coefficients $A_{Lk\ell}$ and their relations to the wave invariants.

(4.3.1) Notation.

- (1) Define: $F_n(h_{-1}, \dots, h_{-n})$ by $[h(\xi) + \sum_{\ell=1}^{\infty} L^{\ell+1} \rho^{-\ell-1} h_{-\ell}(\xi)]^{-\frac{1}{2}}$
 $= h(\xi)^{-\frac{1}{2}} [1 + \sum_{n=2}^{\infty} L^n \rho^{-n} F_n(h^{-1}, h_{-1}, \dots, h_{-n})]$.
-

- (2) Define: $G_n(L, h_{-1}, \dots, h_{-n})$ by

$$e^{i \sum_{j=1}^{\infty} L^{j+1} \rho^{-j} h_{-j}} = 1 + \sum_{n=1}^{\infty} L^n \rho^{-n} G_n(L, h_{-1}, \dots, h_{-n}).$$

- (3) Define F_{nmj} by

$$F_n(h^{-1}, h_{-1}, \dots, h_{-n}) = \sum_{m=1}^n h^{-m} \left[\sum_{j=(j_1, \dots, j_m): |j|=n} F_{nmj} h_{-j_1} \dots h_{-j_m} \right].$$

- (4) Define G_{nmj} by

$$G_n(L, h_{-1}, \dots, h_{-n}) = \sum_{m=1}^n L^m \left[\sum_{j=(j_1, \dots, j_m): |j|=n} G_{nmj} h_{-j_1} \dots h_{-j_m} \right].$$

(4.3.2) Identities.

- (1) $A_{L0n} = L^n h^{-\frac{1}{2}} [F_n + G_n + \sum_{i+j=n} F_i G_j]$.
-

- (2) $A_{Lk} = [h(\xi) + \sum_{\ell=1}^{\infty} L^{\ell+1} \rho^{-\ell-1} h_{-\ell}(\xi)]^k A_{L0}$.
-

$$\begin{aligned}
 (3) \quad & A_{L0n} \\
 &= L^n h^{-\frac{1}{2}} \left[\sum_{m=1}^n \sum_{j=(j_1, \dots, j_m): |j|=n} \cdot \{F_{nmj} h^{-m} + G_{nmj} L^m\} h_{-j_1} \dots h_{-j_m} \right. \\
 &+ \left. \sum_{a+b=n} \sum_{m_1=1}^a \sum_{m_2=1}^b \sum_{i=(i_1, \dots, i_{m_1}): |i|=a} \cdot \sum_{j=(j_1, \dots, j_{m_2}): |j|=b} F_{am_1 i} G_{bm_2 j} h_{-i_1} \dots h_{-i_{m_1}} h_{-j_1} \dots h_{-j_{m_2}} \right].
 \end{aligned}$$

$$(4) \quad A_{Lk\ell} = h A_{L(k-1)\ell} + \sum_{ij: i+j+1=\ell, j \geq 1} L^{j+1} h_{-j} A_{Lki}.$$

The coefficients F_{nmj}, G_{nmj} are universal and hence up to the prefactor of $h^{-\frac{1}{2}}$, A_{Lkj} is a (non-homogeneous) polynomial of degree j in L , in h^{-1} , and in the h_{-j} 's with universal coefficients. In particular, the first few A_{Lkj} 's are given by:

$$\begin{aligned}
 A_{Lk0} &= h(\xi)^{-\frac{1}{2}+k}, & A_{Lk1} &= h(\xi)^{-\frac{1}{2}+k} L^2 h_{-1}, \\
 A_{Lk2} &= h(\xi)^{-\frac{1}{2}+k} \left[L^3 h_{-2} + L^4 h_{-1}^2 + \left(k - \frac{1}{2}\right) L^2 \frac{h_{-1}}{h} \right].
 \end{aligned}$$

It follows that the principal wave invariant is given by

$$a_{T_L, -\frac{3}{2}} := c_L = \frac{1}{\sqrt{2\pi i \alpha L}} e^{i(M_L, \mu)},$$

and that the higher wave invariants $a_{T_L, -\frac{1}{2}+k}$ are given by c_L times polynomials in L and in the derivatives of h, h^{-1}, h_{-1}, \dots at $\xi = 0$. For instance, the subprincipal wave invariant in dimension 2 is given in terms of universal coefficients C'_{ijk} by:

$$a_{T_L, -\frac{1}{2}} = c_L [C'_{004} \partial_\xi^4 g_3 A_{L00} + C'_{010} A_{L01} + C'_{002} \partial_\xi^2 A_{L00}],$$

which is easily seen to equal

$$c_L [C_{004} \partial_\xi^4 h(\xi)|_{\xi=0} + C_{010} L^2 h_{-1}(0) + C_{002} \partial_\xi^2 h(\xi)^{-\frac{1}{2}}|_{\xi=0}].$$

The above indices in the coefficients $C_{kmj_1 j_2 \dots}$ have the following meaning: k corresponds of course to the k index; m is the power of L and j_n are the jet orders of h_{-n} , with with exception of j_o which is the jet order of h .

(4.3.3) Proposition. *We have*

$$a_{L, -\frac{1}{2}+k} = c_L P_k(L, h^{(2)}(0), \dots, h^{(2k+4)}(0), h_{-1}(0), \dots, h_{-1}^{(2k)}(0), \dots, h_{-k}(0), h_{-k}^2(0), h_{-k-1}(0)),$$

where P_k is a polynomial with the following properties:

- (i) It involves only the first $2k + 4$ Taylor coefficients of h at 0, the first $2k$ of h_{-1}, \dots , the first $2k + 2 - 2n$ of $h_{-n} \dots$, the first 2 of h_{-k} and the 0th of h_{-k-1} .
- (ii) It is of degree 1 in the variables $h_{-k-1}(0), h_{-k}^{(2)}(0), \dots, h^{(2k+4)}(0)$, and each occurs in precisely one term.
- (iii) The L -order of the monomials containing these terms is respectively $L^{k+2}, L^{k+1}, \dots, L^0$.

Proof. To prove these claims we combine the formula for $a_{T_L, -k+\frac{1}{2}}$ in (4.2.4) with the formulae for the $A_{Lk\ell}$ given in (4.3.2). Since we may replace the factors of g_3 by h in (4.2.4) and since (4.3.2) expresses $A_{Lk\ell}$ as $h^{-\frac{1}{2}}$ times a polynomial in L and the h_{-j} 's, it is clear that $a_{T_L, -k+\frac{1}{2}}$ is given by a polynomial in the data stated above. It remains to prove that the polynomial has the properties claimed in (i)–(iii).

We prove these claims by proving the stronger statement that

$$a_{L, -\frac{1}{2}+k} = c_L Q_k(L, h(\xi)^{-\frac{1}{2}}, h(\xi), h^{(2)}(\xi), \dots, h^{(2k+4)}(\xi), h_{-1}(\xi), \dots, h_{-1}^{(2k)}(\xi), \dots, h_{-k}(\xi), h_{-k}^2(\xi), h_{-k-1}(\xi))|_{\xi=0},$$

where Q_k has the properties (i)–(iii) for variable ξ .

The proof is by induction on k . The properties are visibly true for the principal and subprincipal wave invariants. Assume then that they are correct for $k \leq N$ and consider how things change as $N \rightarrow N + 1$. First, the amplitude $A_{L(N+1)}$ is given by $[h + L^2 \rho^{-2} h_{-1} + \dots] A_{LN}$. Second, we are looking at the term of order ρ^{-N-2} rather than ρ^{-N-1} in the asymptotic series (4.2.4).

In going one further order into the asymptotic series, only two new things happen:

- The term h_{-N-2} appears for the first time, arising from the linear term in the Taylor expansion of the exponential in (4.2.2). The

linear term in the binomial expansion of the power in (4.2.2) does not contribute at this stage because its ρ^{-1} factor has one higher power.

- Two additional derivatives are allowed to fall on the previous h_j 's. A priori, there could be between two and six additional derivatives in a method of stationary phase expansion. However, the cases of three to six derivatives do not contribute any new data. Indeed, the cases of three-four derivatives occur when m goes up by one. One has to remove the extra factor of ρ^{-1} by applying at least one derivative to the $e^{i\rho g_3}$ factor. But in fact all four have to be applied to g_3 to make a non-zero contribution, and hence no derivatives are left to apply to the $h_{-\ell}$'s. In the case of five or six derivatives, where m goes up to two, one needs to remove two extra factors of ρ^{-1} by applying derivatives to the $e^{i\rho g_3}$ factor, and there is no non-zero way to do this.

Claim (i) follows immediately from these observations.

To prove (ii) we note that by the induction hypothesis,

$$h_{-N-1}(\xi), \dots, h^{(2N+4)}(\xi)$$

occur linearly in $P_N(\xi)$. Since it requires both of the two new derivatives to fall on these factors to produce $h_{-N-1}^{(2)}$ etc. in $P_{N+1}(\xi)$, these factors will also occur to order 1. As for h_{-N-2} , we observed above that it comes only from the linear term in the exponential in (4.2.2) and hence it too appears to order 1.

The proof of (iii) follows the proof of (ii). Since the terms in the exponent of the exponential factor in (4.2.2) have the form $L^{j+1}\rho^{-j}$, the new $h_{-N-2}(\xi)$ term has the coefficient L^{N+3} . Similarly, the other terms under discussion, e.g. $h_{-N-2+r}^{(2r)}(\xi)$, originated as h_{-N-2+r} at the $(N+1-r)$ th stage with the factor of L^{N+3-r} . We would like to show that they remain with just this factor of L as we move inductively up to the $(N+1)$ st stage. To this end, we note that in order to produce the term $h_{-N-2+r}^{(2r)}$ at the $(N+1)$ st stage, $2r$ derivatives must fall on the original h_{-N-2+r} . But there are exactly r stages intermediate between the $(N+1-r)$ th stage and the $(N+1)$ st stage and at each stage at most two new derivatives can fall on a factor. Hence, each of the two new derivatives at each stage must fall on the factor of concern.

Let us also consider what can be multiplied against the original h_{-N-2+r} in the course of producing $h_{-N-2+r}^{(2r)}(\xi)$. We observe that each new pair of derivatives is accompanied by a factor of ρ^{-1} . Since

$h_{-N-2+r}(\xi)$, at the $(N+1-r)$ th stage, comes with the factor ρ^{-N-2+r} , the $2r$ further derivatives will bring its ρ^{-1} -order up to ρ^{-N-2} . Hence no other factors of ρ^{-1} could have fallen on this factor. Therefore in the repeated multiplications by $[h + L^2\rho^{-2}h_{-1} + \dots]$, only the repeated choice of the term h , of order 0 in ρ^{-1} , can have given rise to the term $h_{-N-2+r}^{(2r)}$. It follows that it retains its original L -order, namely L^{N+3-r} .
q.e.d.

4.4. Completion of Proof of Main Lemma.

Our purpose is now to show that by using the joint $\rho, L \rightarrow \infty$ asymptotics one can recover the complete Taylor expansions of all the h_j 's from the wave invariants of the meridian torus and its iterates. That is, to complete the

Proof of the Main Lemma. We will prove by induction that from the wave invariants $a_{T_{pL, -\frac{1}{2}+k}}$ with $k \leq N$ and all $p \in \mathbf{N}$, we can determine the $2k+4$ -jet of h , the $2k$ -jet of h_{-1}, \dots , the $2k+2-2n$ -jet of h_{-n} at $\xi = 0$. (for $n \leq k+1$.)

We note that, unlike in the non-degenerate case, the principal wave invariant determines the 2-jet of h , and gives no new information under iteration. We therefore begin the induction with the subprincipal wave invariant $a_{T, -\frac{1}{2}}$.

From the explicit formula above for the subprincipal wave invariant $a_{L, -\frac{1}{2}}$ it is evident that the 2-jet term in h is old information, while the other two terms differ in the power of L . Hence they decouple under iteration $L \rightarrow pL$ and we can determine the 4-jet of h and the 0-jet of h_{-1} from the first two wave invariants. Hence the induction hypothesis is correct at the first stage.

Assume the induction hypothesis for the $(k-1)$ st stage. Then the only new information at the k th stage is that contained in the terms $h_{-k-1}(0), h_{-k}^{(2)}(0), \dots, h^{(2k+4)}(0)$. By the proposition above, they occur linearly in the monomials containing them and the monomials have also the factor of L to the powers L^{k+1}, L^k, \dots . Hence the terms containing the new data decouple as $p \rightarrow \infty$, and the new data can be determined as stated. This completes the inductive argument.

It follows that we can determine the full Taylor expansions of the h_j 's at $\xi = 0$. Since they are real analytic they are completely determined.

Then from the homogeneity of H_j , we can determine H_j from h_j and hence the entire function \tilde{H} is determined. q.e.d.

Remark. In the above argument, we are able to drop the terms involving only previously known data because of the universal nature of the polynomials P_k . This universality depends on the fact that we are only comparing wave data for quantum torus integrable Laplacians.

5. Proofs of Corollaries 1 and 2

We now show that the symplectic equivalence class of a metric in \mathcal{R}^* is spectrally determined among metrics in this class.

(5.1) Proof of Corollary 1. By the Main Lemma, the principal symbol $H_g(I_1, I_2)$ as a function of the action variables is spectrally determined for a metric g in \mathcal{R}^* . By Propositions (1.5.2-3), H_g determines the geodesic flow up to symplectic equivalence among metrics in \mathcal{R}^* . q.e.d.

Remark. The Corollary could also be proved by noting that the first return time τ_E is spectrally determined. But geodesic flows in the class \mathcal{R}^* are symplectically equivalent if and only if they have the same first return times τ_E . (See also [12] for an equivalent statement.) Also, it should be noted that the first return time could be determined from the wave invariants at the equator; hence Corollary 1 would also follow from Guillemin's inverse result for non-degenerate elliptic closed geodesics.

(5.2) Proof of Corollary 2. Since $\tau_E(I)$ is spectrally determined, so is the function $s(i(I))$ of Proposition 1.3.6. It is given by

$$s(i(I)) = \int_{i(I)}^{\pi-i(I)} f(\cos u) \sin(u) (\sin^2 u - \sin^2(i(I)))_+^{-\frac{1}{2}} du,$$

or, putting $x = \cos u$ for $u \in [0, \frac{\pi}{2}]$ and $-x = \cos u$ for $u \in [\frac{\pi}{2}, \pi]$, by

$$s(i(I)) = \int_0^{\cos(i(I))} [f(x) + f(-x)] (\cos(i(I))^2 - x^2)_+^{-\frac{1}{2}} dx.$$

It therefore suffices to show that

$$Tf(u) = \int_0^u [f(x) + f(-x)] (u^2 - x^2)_+^{-\frac{1}{2}} dx$$

determines $[f(x) + f(-x)]$. But $[f(x) + f(-x)]$ is smooth and even so it may be written as $g(x^2)$ for a smooth g ; changing variables $y = x^2$, and $v = u^2$ we get

$$Tf(v) = \int_0^1 g(y)y^{-\frac{1}{2}}(y-v)_+^{-\frac{1}{2}} dy.$$

Thus T is a standard Abel transform, which is well-known to be invertible. It follows that $g(y)y^{-\frac{1}{2}}$ is spectrally determined and hence is the even part of f . q.e.d.

6. Proof of Final Lemma

To complete the proof of the Theorem, we need to show that \hat{H} determines a metric in \mathcal{R} . It is plausible that this can be done for the following reason: The spectrum of Δ is the set of values $\{\hat{H}(n, k + \frac{1}{2})^2\}$ of \hat{H}^2 on the integer points of the action cone Γ . On the other hand since $\hat{I}_1 = \frac{\partial}{\partial \theta}$, the set of values $\hat{H}(n, k + \frac{1}{2})$ for fixed n , is just the spectrum $\{E_{nk}\}$ of the the singular Sturm-Liouville operator

$$L_n = -\left(\frac{d}{dr}\right)^2 + q_n(r), \quad q_n(r) := q(r) + \frac{n^2}{a(r)^2},$$

$$q(r) = -\frac{2a(r)a''(r) - (a'(r))^2}{a(r)^2}$$

obtained by separating variables in Δ , fixing the eigenvalue of $\frac{\partial}{i\partial \theta}$ equal to n , and putting the radial part in normal form. Hence, from the coefficients of \hat{H} we can determine $Spec(L_n)$ for each n . That is, from $Spec(\Delta)$ we have determined the *joint* spectrum of $(\Delta, \frac{\partial}{\partial \theta})$. It would thus remain to show that the metric can be determined from this joint spectrum, a much more elementary inverse result which has been stated several times in the literature ([21], [3], [20]). Since these prior discussions seem to us somewhat sketchy and incomplete, we give a self-contained proof below which was found before we were aware of these references.

Proof of Final Lemma.

The proof is basically to write down explicit expressions for H and H_{-1} in terms of the metric (i.e., in terms of $a(r)$) and then to invert the

expressions to determine $a(r)$. The first step is therefore to construct an initial part of the quantum normal form explicitly from the metric. Up to now, we only know that a polyhomogeneous normal form exists.

To be sure, the principal term H has already been implicitly expressed in terms of the metric: Knowledge of H is equivalent to knowledge of the level set $\{H = 1\}$ and hence to knowledge of the function $F(I_1)$ in (1.3.1). As we have seen in Corollary (5.2), H only determines the ‘even part’ of the metric. Hence we need to determine at least one of the subprincipal terms H_{-j} . It turns out that only H_{-1} is needed in addition to H to determine g . Since the calculation of H_{-1} requires a new calculation of H , we start calculating both from scratch.

To determine H and H_{-1} in terms of g , we are going to study the spectral asymptotics of $\sqrt{\Delta} = \hat{H}(\hat{I}_1, \hat{I}_2)$ along ‘rays of representations’ of the quantum torus action, i.e., along multiples of a given lattice point (n_o, k_o) . Such rays are the quantum analogue of rays $\mathbb{R}^+ T_I \subset T^*S^2 - 0$ thru invariant tori and are basic to homogeneous quantization theory [19]. The basic idea is that the lattice points $(n_o, k_o + \frac{1}{2})$ parametrize tori T_{n_o, k_o} satisfying the Bohr-Sommerfeld quantization condition. To each such quantizable torus one can construct a joint eigenfunction ϕ_{n_o, k_o} of (\hat{I}_1, \hat{I}_2) by the WKB method. The $\{\phi_{n, k}\}$ are eigenfunctions of Δ with complete asymptotic expansions along rays. By studying the eigenvalue problem as $|(n, k)| \rightarrow \infty$ we can determine the H_{-j} ’s.

The WKB method which we employ is closely related to the classical WKB method for constructing quasi-modes (cf. [11] and the Appendix), except that we have an internal rather than an external Planck constant. Let us recall the relevant terminology and notation from the latter case. For each torus T_I , we denote by $\mathcal{O}^\mu(T_I, A)$ the space of oscillatory integrals (semi-classical Lagrangean distributions) associated to T_I , with semi-classical parameters $\{k_m\}$ in a set $A \subset \mathbb{R}^+$ to be specified by the quantization condition. Such oscillatory integrals have the form

$$(6.1) \quad u(x, k) = k^{\frac{n}{2} + \mu} \sum_{\ell \in L} \int_{\mathcal{V}_\ell} e^{ik\psi_\ell(x, \xi)} \alpha(x, \xi, k) d\xi,$$

where the projection of T_I is covered by open sets \mathcal{U}_ℓ , $T^*(\mathcal{U}_\ell) \simeq \mathcal{U}_\ell \times \mathcal{V}_\ell$, and the phase $\phi_\ell(x, \xi)$, with $(x, \xi) \in \mathcal{U}_\ell \times \mathcal{V}_\ell$, parametrizes a part of T_I . The amplitude is a classical symbol in k of order μ . For further details, we refer to [11, §8].

As discussed above, the torii T_I project $2 - 1$ to the annuli $r_+(I) \leq r \leq r_-(I)$ in S^2 and have fold singularities along the extremal parallels. Away from the parallels, an associated quasi-mode is given by a sum of two simple WKB functions $\alpha_{\pm}(r, \theta)e^{ik\psi_{\pm}(r, \theta)}$. Since the actual Δ -eigenfunctions ϕ_{n_o, k_o} are quasi-modes attached to the Bohr-Sommerfeld torii T_{n_o, k_o} , they have such a form modulo $|(n_o, k_o)|^{-\infty}$. And, since ϕ_{n_o, k_o} is an exact $\frac{\partial}{\partial \theta}$ -eigenfunction, its phases must take the form $\phi_{\ell}(r, \theta) = n_o\theta + S_{n_o, k_o}(r)$ with θ -independent amplitudes in polar coordinates. It follows that $\phi_{n_o, k_o}(r, \theta)$ has the form $e^{in_o\theta} f_{(n_o, k_o)}(r)$, where $f_{(n_o, k_o)}(r)$ is an oscillatory integral in the r -variable. It is of course associated to the pushed-forward Lagrangean $\Lambda_{n_o, k_o} := p_* T_{n_o, k_o}$ where $p : T^*S^2 \rightarrow T^*[0, L]$ is the projection induced from the map $(r, \theta) \rightarrow r$. In the case of the meridian torii T_{0, k_o} , the pushforward is just a horizontal line $p_r = C$ in $T^*[0, L]$. In the other cases, the Λ_{n_o, k_o} is a closed curve projecting to an interval $[r_+(n_o, k_o), r_-(n_o, k_o)]$ with fold singularities at the turning points (endpoints). The curve is given by an equation of the form $H_{n_o}(r, p_r) = E$ where $H_{n_o}(r) = p_r^2 - \frac{n_o^2}{a(r)^2}$ is the radial Hamiltonian, and E was the level of the torus.

The radial part $f_{n_o, k_o}(r)$ of ϕ_{n_o, k_o} , is an eigenfunction of the radial operator $D_{n_o} = -(\frac{d}{dr})^2 - \frac{a'}{a} \frac{d}{dr} + \frac{n_o^2}{a(r)^2}$ arising from separating variables in Δ . Before proceeding, we simplify by conjugating D_{n_o} to the 1/2-density radial Laplacian

$$\hat{D}_{n_o} := a(r)^{\frac{1}{2}} D_{n_o} a(r)^{-\frac{1}{2}} = D_r^2 + \frac{n_o^2}{a(r)^2} + W,$$

where $W = a(r)^{\frac{1}{2}} [-(\frac{d}{dr})^2 - \frac{a'}{a} \frac{d}{dr}] a(r)^{-\frac{1}{2}}$. (Note that the volume form $dv_g = a(r) dr d\theta$ of (S^2, g) projects to $a(r) dr$.) Thus, we view the radial eigenfunction as having the form $g_{n_o, k_o}(r) \sqrt{adr}$ and apply the WKB method to the eigenvalue problem

$$(6.2) \quad \hat{D}_{n_o} g_{n_o, k_o}(r) \sim (H(n_o, k_o + \frac{1}{2}) + H_{-1}(n_o, k_o + \frac{1}{2}) + \dots)^2 g_{n_o, k_o}(r).$$

Although the coefficients become singular at $r = 0, L$, the standard WKB theory applies in the interior because it applies to quasi-modes of Δ on S^2 . Away from the turning points, the radial part of the 1/2-density eigenfunction therefore has the form

$$(6.3) \quad g_{n_o, k_o} = \sum_{\pm} [e^{\pm i S_{n_o, k_o}(r)} \sum_{m=0}^{\infty} \alpha_{n_o, k_o; m}(r)],$$

where the phase S_{n_o, k_o} is homogeneous of degree 1, and the amplitude $\alpha_{n_o, k_o; m}$ is homogeneous of degree $-m$ in $(n_o, k_o + \frac{1}{2})$. These homogeneities replace the powers of k in the non-homogeneous theory described above.

The Bohr-Sommerfeld quantization condition on T_I is that:

$$\frac{1}{2\pi}I = (n_o, k_o) + \frac{1}{4}\mu_o, \quad (n, k) \in \Gamma \cap \mathbb{Z}^2,$$

where $\mu_o = (0, 2)$ is the Maslov index (cf. [11, §4]). It is satisfied by T_{n_o, k_o} and hence by the radial Lagrangean Λ_{n_o, k_o} . The local WKB ansatz (6.3) therefore extends to a quasi-mode of infinite order associated to the global Λ_{n_o, k_o} . Our purpose now is to write down and solve the first two transport equations, which are needed to determine H and H_{-1} . For background on the relevant aspects of the WKB method, we refer to the Appendix.

The transport equations are obtained by separating out terms of like order in the asymptotic eigenvalue problem:

$$\begin{aligned} & \hat{D}_{n_o} \sum_{\pm} [e^{\pm S_{n_o k_o}(r)} \sum_{m=0}^{\infty} \alpha_{n_o, k_o; m}(r)] \\ & \sim (H(n_o, k_o + \frac{1}{2}) + H_{-1}(n_o, k_o + \frac{1}{2}) + \dots)^2 \\ & \cdot \sum_{\pm} [e^{\pm S_{n_o k_o}(r)} \sum_{m=0}^{\infty} \alpha_{n_o, k_o; m}(r)]. \end{aligned}$$

The leading term, of order 2, is the eikonal equation

$$(6.4a) \quad |S'_{n_o k_o}(r)|^2 + \frac{n_o^2}{a(r)^2} = H(n_o, k_o + \frac{1}{2})^2,$$

whose solution is the first order phase function

$$(6.4b) \quad S_{n_o k_o}(r) := \int \sqrt{H(n_o, k_o)^2 - \frac{n_o^2}{a(r)^2}} dr.$$

The Bohr-Sommerfeld quantization condition on $\Lambda_{(n_o, k_o)}$ thus reads:

$$I_{n_o}(E) := \text{Area} \{H_{n_o} \leq E\} = 2\pi(k_o + \frac{1}{2}) \quad \text{with} \quad E = H(n_o, k_o + \frac{1}{2}).$$

The first transport equation, $2\alpha'_{n_o, k_o; 0} S' + \alpha S' = 0$, is solved by

$$(6.5) \quad \alpha_{(n_o, k_o; 0)}(r) = [E_{n_o, k_o} - \frac{1}{a(r)^2}]^{-\frac{1}{4}}$$

away from the turning points. Here, $E_{n_o, k_o} = \frac{H(n_o, k_o + \frac{1}{2})^2}{n_o^2}$. The solution is of course determined only up to a constant, and we have normalized it so that $\alpha_{(n_o, k_o; 0)}$ is homogeneous of order 1.

In the usual way (see the Appendix), we will interpret $\alpha_{n_o, k_o; 0}$ as the coefficient of the $\Xi_{H_{n_o}}$ -invariant 1/2-density

$$\nu_0 = [E_{n_o, k_o} - \frac{1}{a(r)^2}]^{-\frac{1}{4}} \sqrt{dr}$$

on $\Lambda_{n_o k_o} = \{H_{n_o}(r, p_r) = H(n_o, k_o + \frac{1}{2})\}$, where $\Xi_{H_{n_o}}$ is the Hamilton vector field of H_{n_o} . We then re-write the higher coefficients $\alpha_{n_o, k_o; m}(r)$ in the form $\alpha_{n_o, k_o; m}(r)\nu_0$.

The second transport equation (of order zero) then has the form:

$$(6.6) \quad \begin{aligned} \Xi_{H_{n_o}} \alpha_{n_o, k_o, -1} &= -\alpha_{n_o, k_o, 0}^{-1} \left(\frac{d}{dr}\right)^2 \alpha_{n_o, k_o, 0} + W \\ &+ 2H(n_o, k_o + \frac{1}{2}) H_{-1}(n_o, k_o + \frac{1}{2}), \end{aligned}$$

where r denotes the local coordinate on $\Lambda_{n_o k_o}$ obtained by pulling back the base coordinate under the projection. The integral of the left-hand side over the closed curve $\Lambda_{n_o k_o} = \{H_{n_o} = H(n_o, k_o + \frac{1}{2})\}$ with respect to the $\Xi_{H_{n_o}}$ -invariant density

$$\alpha_{n_o, k_o, 0}^2 dr = \frac{1}{\sqrt{E_{n_o, k_o} - \frac{1}{a(r)^2}}} dr$$

must equal zero. This gives a formula for the first correction to the Bohr-Sommerfeld eigenvalue $H(k_o, n_o)$:

$$(6.7) \quad \begin{aligned} &H_{-1}(n_o, k_o + \frac{1}{2}) \\ &= \frac{1}{2T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \cdot \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} [\alpha_{n_o, k_o, 0} \left(\frac{d}{dr}\right)^2 \alpha_{n_o, k_o, 0} \\ &\quad - \frac{1}{2} W \alpha_{n_o, k_o, 0}^2] dr. \end{aligned}$$

Here, $T(n_o, k_o)$ denotes the period of the $\Xi_{H_{n_o}}$ flow on $\Lambda_{n_o k_o}$, and $r_{\pm}(n_o, k_o)$ are the turning points. Plugging in (6.5) we get, formally, the expression

$$\begin{aligned}
 & H_{-1}(n_o, k_o + \frac{1}{2}) \\
 &= \frac{1}{2T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \\
 (6.8) \quad & \cdot \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} (E_{n_o, k_o} - \frac{1}{a(r)^2})^{-1/4} (\frac{d}{dr})^2 (E_{n_o, k_o} - \frac{1}{a(r)^2})^{-1/4} dr \\
 & - \frac{1}{2T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \frac{W}{\sqrt{E_{n_o, k_o} - \frac{1}{a(r)^2}}} dr.
 \end{aligned}$$

Actually, the integral in (6.8) diverges at the turning points, and the correct formula is the regularization obtained (for instance) by the method of the Maslov canonical operator. For the sake of completeness, we provide an exposition of this method in the appendix. Roughly speaking, it regularizes (6.8) by formally integrating the $\frac{d}{dr}$ derivatives by parts and by moving them outside the integral (in the appropriate way) as derivatives in the energy level E . A crucial consequence of the regularization is that only first derivatives of a appear in the formulae for $H_{-1}(n_o, k_o + \frac{1}{2})$.

Thus the first term of (6.8) is regularized by

$$\begin{aligned}
 (6.8.1 \text{ reg}) \quad & \frac{C_1}{T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \\
 & \cdot \partial_E^2 \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \frac{a'(r)^2}{a(r)^6} (E - \frac{1}{a(r)^2})^{-\frac{1}{2}} dr |_{E=E(n_o, k_o)}.
 \end{aligned}$$

Here and below C_i denote (non-zero) constants which we will not need to determine. For the second term of (6.8), we note that

$$\begin{aligned}
 \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} W \alpha_{n_o, k_o; 0}^2 &= \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \frac{d}{dr} (\alpha_{n_o, k_o; 0}^2 a^{\frac{1}{2}}) \frac{d}{dr} a^{-\frac{1}{2}} dr \\
 &+ \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \alpha_{n_o, k_o; 0}^2 a^{\frac{1}{2}} \frac{a'}{a} \frac{d}{dr} a^{-\frac{1}{2}} dr.
 \end{aligned}$$

After some simplification, this regularizes to:

$$\begin{aligned}
 & \frac{C_2}{T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \\
 & \cdot \partial_E \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \frac{a'(r)^2}{a(r)^4} \left(E - \frac{1}{a(r)^2}\right)^{-\frac{1}{2}} dr \Big|_{E=E(n_o, k_o)} \\
 (6.8.2 \text{ reg}) \quad & + \frac{C_3}{T(n_o, k_o)H(n_o, k_o + \frac{1}{2})} \\
 & \cdot \int_{r_-(n_o, k_o)}^{r_+(n_o, k_o)} \frac{a'(r)^2}{a(r)^2} \left(E - \frac{1}{a(r)^2}\right)^{-\frac{1}{2}} dr \Big|_{E=E(n_o, k_o)}.
 \end{aligned}$$

Now let us return to the inverse problem. We begin from the fact that $\hat{H}(n_o, \hat{I}_2)$ is a known function of the variable $I := \hat{I}_2$ for each n_o . Its principal symbol $H_{n_o}(I) := H_1(n_o, I)$ is then a known function and hence the inverse function

$$I_{n_o}(E) = \int_{r_-(E)}^{r_+(E)} \sqrt{E - \frac{1}{a(r)^2}} dr$$

satisfying $H_{n_o}(I_{n_o}(E)) = E$ is a known of function E . This of course presupposes that $\partial_I H_{n_o}(I) \neq 0$, which follows from the non-degeneracy assumption (1.1.6). We may write the integral in the form

$$\int_{\mathbb{R}} (E - x)_{+}^{\frac{1}{2}} d\mu(x),$$

where μ is the distribution function of $\frac{1}{a^2}$, i.e.,

$$\mu(x) := |\{r : \frac{1}{a(r)^2} \leq x\}|$$

with $|\cdot|$ the Lebesgue measure. The above integral is an Abel transform and as mentioned above it is invertible. Hence

$$d\mu(x) = \sum_{r: \frac{1}{a(r)^2} = x} \left| \frac{d}{dr} \frac{1}{a(r)^2} \right|^{-1} dx$$

is a spectral invariant. Some simplification leads to the conclusion that the function

$$J(x) := \sum_{r: a(r)=x} \frac{1}{|a'(r)|}$$

is known from the spectrum. By the simplicity assumption on a , there are just two solutions of $a(r) = x$; the smaller will be written $r_-(x)$ and the larger, $r_+(x)$. Thus, the function

$$(6.9) \quad J(x) = \frac{1}{|a'(r_-(x))|} + \frac{1}{|a'(r_+(x))|}$$

is a spectral invariant.

Now let us turn to the H_{-1} expression. Since $H(I_1, I_2)$ is a spectral invariant, the factors $H(n_o, k_o + \frac{1}{2})$ and $T(n_o, k_o)$ are spectral invariants. Hence we may remove them from the expression for H_{-1} and still get a spectral invariant. For various universal constants C_1, C_2, C_3 it takes the form

$$(6.10) \quad [C_1 \partial_E^2 \int \frac{(a')^2}{a^6} (E - \frac{1}{a^2})_+^{-\frac{1}{2}} dr + C_2 \partial_E \int \frac{(a')^2}{a^4} (E - \frac{1}{a^2})_+^{-\frac{1}{2}} dr + C_3 \int \frac{(a')^2}{a^2} (E - \frac{1}{a^2})_+^{-\frac{1}{2}} dr] |_{E = \frac{H(I_1, I_2)}{I_1^2}}$$

By a change of variables, we may rewrite (6.10) in the form

$$(6.11) \quad [C_1 \partial_E^2 \int K(x) x^{\frac{3}{2}} (E - x)_+^{-\frac{1}{2}} dx + C_2 \partial_E \int K(x) x^{\frac{1}{2}} (E - x)_+^{-\frac{1}{2}} dx + C_3 \int K(x) x^{-\frac{1}{2}} (E - x)_+^{-\frac{1}{2}} dx] |_{E = \frac{H(I_1, I_2)}{I_1^2}},$$

where

$$(6.12) \quad K(x) = |a'(r_-(x))| + |a'(r_+(x))|.$$

All values of E which occur as ratios $\frac{H(I_1, I_2)}{I_1^2}$ give spectral invariants, so (1.10) (as a function of the variable E) is a spectral invariant.

We now claim that K itself is a spectral invariant. To determine it from (6.11) we rewrite (6.11) in terms of the fractional integral operators (cf. [16, Ch.1 §5.5])

$$I_\alpha f(E) = f * \frac{x_+^{\alpha-1}}{\Gamma(\alpha)}(E) = \frac{1}{\Gamma(\alpha)} \int_0^E f(y) (E - y)^{\alpha-1} dy$$

on the half-line $[0, \infty]$. These operators satisfy

$$I_\alpha \circ I_\beta = I_{\alpha+\beta}, \quad I_{-k} = \left(\frac{d}{dx}\right)^k.$$

Thus (6.11) equals $\mathcal{L}(K)$ where \mathcal{L} is the fractional integral operator

$$(6.13) \quad \mathcal{L} := C_1 I_{-3/2} x^{\frac{3}{2}} + C_2 I_{-\frac{1}{2}} x^{\frac{1}{2}} + C_3 I_{\frac{1}{2}} x^{-\frac{1}{2}}.$$

To solve for K we apply $I_{-\frac{1}{2}}$ to $\mathcal{L}K$ to get

$$(6.14) \quad C_1' \frac{d^2}{dx} (x^{\frac{3}{2}} K(x)) + C_2' \frac{d}{dx} (x^{\frac{1}{2}} K) + C_3' x^{-\frac{1}{2}} K = I_{-\frac{1}{2}} \mathcal{L}K.$$

The left-hand side determines K up to a solution f of the associated homogeneous equation, essentially an Euler equation. But also $K = 0$ on $[0, a(r_o)^{-2}]$. Since no homogeneous solution can have this property, K is uniquely determined by this boundary condition.

It follows that both (6.9) and (6.12) are spectral invariants. But from $a + b$ and $\frac{1}{a} + \frac{1}{b}$ one can determine the pair (a, b) . Hence $a'(r)$ is determined from the spectrum. Since $a(0) = 0$, this determines a and hence the surface. q.e.d.

7. Appendix

The purpose of this appendix is to give an algorithm for calculating the higher order terms in the quasi-classical approximation of eigenvalues for 1 D Schrodinger operators $-\frac{\hbar^2}{2} \frac{d^2}{dx^2} + V$ with confining potentials. In particular, we carry out the calculation of the $\hbar^2 E_n^{(2)}$ term, which was used in the proof of the Final Lemma.

The algorithm is based on the Maslov method of canonical operators. Expositions and refinements of this method can be found, among other places, in Maslov's book [23], in the article of Colin de Verdiere [11] (and in its references), and in the recent book of Bates- Weinstein [2]. Although these references explain the construction of the canonical operator and prove the existence of complete quasi-classical eigenvalue expansions, they do not generally go on to describe the calculation of the terms. An exception is the original book of Maslov [23], which does calculate the first two or three terms; but the method of canonical operators is abandoned at this point in favor of some methods of special functions. As we will show, the canonical operator method gives the required corrections quite efficiently.

The set-up

We are concerned with the semi-classical eigenvalue problem:

$$\begin{cases} [-\frac{\hbar^2}{2} \frac{d^2}{dx^2} + V]\psi_n(x, \hbar) = E_n(\hbar)\psi_n(x, \hbar), \\ \langle \psi_n, \psi_m \rangle = \delta_{mn} + O(\hbar^\infty), \\ E_n(\hbar) = E_n^{(1)}(\hbar) + \hbar^2 E_n^{(2)} + \hbar^3 E_n^{(3)} + \dots \end{cases}$$

The unknown function $\psi_n(x, \hbar)$ is an oscillatory associated to a Lagrangean of the form $\Lambda_n := \{H = E_n^{(1)}(\hbar)\}$ where $E_n^{(1)}(\hbar)$ is determined by the Bohr-Sommerfeld-Maslov quantization condition:

$$\frac{1}{2\pi\hbar} \int_{\Lambda_n} \xi \cdot dx = n + \frac{1}{4}\mu.$$

Here, μ is the Maslov index of Λ_n ; it equals 2 for connected level sets of Hamiltonians of the form $H(x, \xi) = \xi^2 + V(x)$.

To find the higher order corrections $E_n^{(k)}$ we consider the Maslov canonical operator

$$\mathcal{U}_\hbar : S^m(\Lambda_n, \Omega_{\frac{1}{2}} \otimes \mathcal{M}) \rightarrow \mathcal{O}^m(\mathbb{R}, \Lambda_n).$$

We follow here the notation and terminology of [2], [11]: $S^m(\Lambda_n, \Omega_{\frac{1}{2}} \otimes \mathcal{M})$ is the space of symbolic sections of the bundle of 1/2-densities times Maslov factors, and $\mathcal{O}^m(\mathbb{R}, \Lambda_n)$ is the space of oscillatory integrals associated to Λ_n . There is a natural symbol map in the reverse direction; any inverse modulo $O(\hbar)$ is a quantization or canonical operator. Its existence is equivalent to the condition that Λ_n satisfy the BSM quantization condition. For the background we again refer to [11], [2].

We also denote by Ξ_H the Hamilton vector field of H , by \mathcal{L}_{Ξ_H} the Lie derivative on any bundle, by s a nowhere vanishing section of \mathcal{M} , and by ρ a \mathcal{L}_{Ξ_H} -invariant density on Λ_n (for a fixed n). By surjectivity of \mathcal{U}_\hbar , we can write an oscillatory integral associated to Λ_n in the form

$$\psi_n(x, \hbar)|dx|^{\frac{1}{2}} \sim \mathcal{U}_\hbar[e^{\frac{i}{\hbar}\phi} \cdot \sum_{j=0}^{\infty} \hbar^j f_j \rho^{\frac{1}{2}} \otimes s].$$

Our aim is to determine the quasi-classical series $\sum E_n^{(j)} h^j$ and coefficient functions f_j for which the asymptotic eigenvalue problem is solvable.

We begin by constructing local solutions. Thus we first consider x -projectible pieces of $\Lambda_n \subset T^*\mathbb{R}$: pieces which project regularly from a neighborhood of $\lambda \in \Lambda_n$ to a neighborhood of $x \in \mathbb{R}$. Restricted to densities supported on such pieces, the Maslov canonical operator is truly canonical: if $S(x)$ is a phase locally parametrizing Λ_n , then

$$\mathcal{U}_h[e^{\frac{i}{h}\phi} \cdot \sum_{j=0}^{\infty} h^j f_j \rho^{\frac{1}{2}} \otimes s] = e^{\frac{i}{h}S} \cdot \sum_{j=0}^{\infty} h^j a_j$$

for some smooth coefficients a_j . We may then substitute this expression into the eigenvalue problem and obtain eikonal and transport equations. The eikonal equation $(S')^2 + V(x) = E_n^{(1)}(h) = 0$ has been solved by our choice of phase, so the transport equations become:

$$\begin{cases} a_0 \frac{d^2}{dx^2} S + 2\nabla a_0 \nabla S = 0, \\ a_1 \frac{d^2}{dx^2} S + 2\nabla a_1 \cdot \nabla S - i \frac{d^2}{dx^2} a_0 = 2i E_n^{(2)} a_0, \\ a_2 \frac{d^2}{dx^2} S + 2\nabla a_2 \cdot \nabla S - i \frac{d^2}{dx^2} a_1 = 2i [E_n^{(3)} a_0 + E_n^{(2)} a_1], \end{cases}$$

and so on. As is well-known (cf. e.g. [2]), these equations may be put into geometric form by observing that $\nabla S \cdot \nabla$ is the projection to \mathbb{R} of \mathcal{L}_{Ξ_H} and that $[a \frac{d^2}{dx^2} S + 2\nabla \cdot S] |dx| = \text{div}(a^2 \nabla S) |dx| = \mathcal{L}_{\Xi_H}(a^2 |dx|)$. Hence the transport equations become

$$\begin{cases} \mathcal{L}_{\Xi_H}(a_0 |dx|^{\frac{1}{2}}) = 0, \\ \mathcal{L}_{\Xi_H}(a_1 |dx|^{\frac{1}{2}}) = (2i E_n^{(2)} a_0 + i \frac{d^2}{dx^2} a_0) |dx|^{\frac{1}{2}}, \\ \mathcal{L}_{\Xi_H}(a_2 |dx|^{\frac{1}{2}}) = 2i [(E_n^{(3)} a_0 + E_n^{(2)} a_1) + i \frac{d^2}{dx^2} a_1] |dx|^{\frac{1}{2}}. \end{cases}$$

The invariant 1/2-density is given by the well-known formula $a_0 = (E - V)^{-\frac{1}{4}}$, or equivalently by $\frac{|dx|^{\frac{1}{2}}}{p^{\frac{1}{2}}}$ on Λ_n . If we write $a_j |dx|^{\frac{1}{2}} = f_j \rho^{\frac{1}{2}}$, then

$f_0 = 1, f_j = \frac{a_j}{a_o}$ and the transport equations become

$$\begin{cases} \Xi_H f_1 = (2iE_n^{(2)} + ia_o^{-1} \frac{d^2}{dx^2} a_o), \\ \Xi_H f_2 = 2i[(E_n^{(3)} + E_n^{(2)} \frac{a_1}{a_o}) + ia_o^{-1} \frac{d^2}{dx^2} a_1]. \end{cases}$$

Here, the expressions in a_o, a_1, \dots are understood to have been lifted up to Λ_n .

The eigenvalue corrections $E_n^{(k)}$ are determined by integrating both sides of the transport equations over the level $\{H = E_n^{(1)}\}$. Since the equations are solvable and the left-hand sides will integrate to zero, we get

$$\begin{cases} E_n^{(2)} = \frac{1}{2iT(E_n^{(1)})} \int_{\{H=E_n^{(1)}\}} [a_o^{-1} \frac{d^2}{dx^2} a_o] \rho, \\ E_n^{(3)} = \frac{1}{2iT(E_n^{(1)})} \int_{\{H=E_n^{(1)}\}} [E_n^{(2)} \frac{a_1}{a_o} + ia_o^{-1} \frac{d^2}{dx^2} a_1] \rho. \end{cases}$$

Here, $T(E)$ is the period of Ξ_H at level E . Parametrizing $\{H = E_n^{(1)}\}$ as a graph over the x -axis away from the turning points, the invariant density takes the form $\frac{dx}{\sqrt{E-V}}$ with $E = E_1^{(1)}$. Hence at least formally the eigenvalue corrections are given by

$$\begin{aligned} E_n^{(2)} &= \frac{1}{2iT(E)} \int_{x_-(E)}^{x_+(E)} [a_o^{-1} \frac{d^2}{dx^2} a_o] \frac{dx}{\sqrt{E-V}} \\ &= \frac{1}{2iT(E)} \int_{x_-(E)}^{x_+(E)} (E - V(x))^{-\frac{1}{4}} \frac{d^2}{dx^2} (E - V(x))^{-\frac{1}{4}} dx. \end{aligned}$$

Unfortunately the integral is ill-defined due to the singularities at the turning points. The problem is that the Maslov operator cannot be defined near these points as a simple pull-back operator. Rather it should be defined as the composition of the Fourier transform with the pull-back operator defined over the ξ -projection. This problem and its solution constitute a key aspect of the Maslov method (in one dimension); we refer to [11], [2] for extended discussions

The point which is important for us is that the Maslov method gives a regularization of the divergent integral; it works in the following way: we introduce a cut-off ψ_δ supported away from a δ -neighborhood of the turning points $x_\pm(E)$. More precisely we define ψ_δ^\pm on Λ_n , equal to one

on $(2\delta, \frac{1}{2}T(E) - 2\delta)$ resp. $(\frac{1}{2}T(E) + 2\delta, T(E) - 2\delta)$ and equal to zero on $(T(E) - \delta, \delta)$ resp. $(\frac{1}{2}T(E) - \delta, \frac{1}{2}T(E) + \delta)$. We then put

$$\mathcal{U}_h(f\rho \otimes se^{\frac{i}{h}\phi}) := I_h(\psi_\delta f\rho \otimes se^{\frac{i}{h}\phi}) + J_h((1 - \psi_\delta)f\rho \otimes se^{\frac{i}{h}\phi}),$$

where I_h is the pull-back to \mathbb{R} under the phase parametrization by $\xi = S'(x)$, and J_h is the Fourier transform of the ξ -parametrization. The notation ψ_δ stands for ψ_δ^\pm . For details on I_h, J_h , see [2].

Returning to the previous calculation of eigenvalue corrections, we see that what is missing is the cut-off ψ_δ in the integrals and the contributions from J_h . We wish to avoid confronting the latter. Fortunately, it is not necessary to do so: the fact that the eigenvalues are independent of δ allows us to determine the J_h (i.e., the turning point) contribution indirectly. To see this, we substitute the cut-off into the formula for $E_n^{(2)}$ to get

$$E_n^{(2)} = \frac{1}{2iT(E)} \int_{x_-(E)}^{x_+(E)} (E - V(x))^{-\frac{1}{4}} \frac{d^2}{dx^2} [(E - V(x))^{-\frac{1}{4}} \psi_\delta] dx + II_\delta$$

with II_δ the contribution from J_h . Since the integral is now nicely convergent, we can integrate by parts and simplify to the form

$$\frac{1}{16T(E)} \int_{x_-(E)}^{x_+(E)} \frac{V'(x)^2}{(E - V)^{\frac{5}{2}}} \psi_\delta dx - \frac{1}{T(E)} \int_{x_-(E)}^{x_+(E)} \frac{V'(x)}{(E - V)^{\frac{3}{2}}} \psi'_\delta(x) dx.$$

The first term tends to

$$\frac{1}{12T(E)} \frac{d^2}{dE^2} \int_{x_-(E)}^{x_+(E)} \frac{V'(x)^2}{(E - V)^{\frac{1}{2}}} \psi_\delta dx$$

as $\delta \rightarrow 0$. The second expression only involves the Taylor expansion of V near the turning points. Its singular part must be cancelled by the singular part of II_δ , leaving a possible ‘residue’ as $\delta \rightarrow 0$. We claim that this residue is zero; in fact, this is known to be the case since the first term is well-defined, independent of δ , and agrees with the formula given in [23]. To give an independent proof that it vanishes, without analysing the J_h -term in detail, we observe that the limit contribution involves only the 2-jet of V at the turning points. Hence it must agree with the corresponding expression for a harmonic oscillator at its turning points. But no such correction occurs.

References

- [1] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer, New York, 1977.
- [2] S. Bates & A. Weinstein, *Lectures on the geometry of quantization*, Berkeley lecture notes, 1996.
- [3] P. Berard, *Quelques remarques sur les surfaces de revolution dans \mathbb{R}^3* , C. R. Acad. Sci. Paris **282** (1976) 159–161.
- [4] A. Besse, *On manifolds all of whose geodesics are closed*, *Ergeb. Math. Grenoble* (3) **93**, Springer, New York, 1978.
- [5] P. Bleher, *Distribution of energy levels of a quantum free particle on a surface of revolution*, *Duke Math. J.* **74** (1994) 45–93.
- [6] L. Boutet de Monvel & V. Guillemin, *The Spectral Theory of Toeplitz Operators*, *Ann. of Math. Stud.* **99** Princeton U. Press (1981).
- [7] J. Brüning & E. Heintze, *Spektrale Starrheit gewisser Drehflächen*, *Math. Ann.* **269** (1984) 95–101.
- [8] G. Darboux, *Theorie des surfaces. III*, Gauthier-Villars, Paris 1894.
- [9] Y. Colin de Verdière, *Spectre conjoint d'opérateurs pseudo-différentiels qui commutent I: Le cas non-intégrable*, *Duke Math. J.* **46** (1979) 169–182.
- [10] ———, *Sur les longueurs des trajectoires périodiques d'un billard*, (P. Dazord and N. Desolneux-Moulis eds.), *Geometrie Symplectique et de Contact: Autour du Theoreme de Poincaré-Birkhoff*. Travaux en Cours, Sem. Sud-Rhodanien de Geometrie III Pairs: Herman, 1984, 122-139.
- [11] ———, *Quasi-modes sur les varietes Riemanniennes*, *Invent. Math.* **43** (1977) 15–52.
- [12] C. Croke & B. Kleiner, *Conjugacy and rigidity for manifolds with a parallel vector field*, *J. Differential Geom.* **39** (1994) 659–680.
- [13] J. J. Duistermaat, *On the existence of global action-angle variables*, *Comm. Pure. Appl. Math.* **33** (1980) 687–706.
- [14] J. J. Duistermaat & V. Guillemin, *The spectrum of positive elliptic operators and periodic bicharacteristics*, *Invent. Math.* **29** (1975) 29–79.
- [15] J. P. Francoise & V. Guillemin, *On the period spectrum of a symplectic mapping*, *J. Funct. Anal.* **100** (1991) 317–358.
- [16] I. M. Gel'fand & G. E. Shilov, *Generalized Functions*, Vol. 1, Academic Press, New York, 1964.
- [17] V. Guillemin, *Wave trace invariants*, *Duke Math. J.* **83** (1996) 287–52.

- [18] V. Guillemin & S. Sternberg, *Symplectic techniques in Physics*. Cambridge Univ. Press, New York, 1985.
- [19] ———, *Homogeneous quantization and multiplicities of group representations*, J. Funct. Anal. **47** (1982) 344–380.
- [20] D. Gurarie, *Semiclassical eigenvalues and shape problems on surfaces of revolution*, J. Math. Phys. **36** (1995) 1934–1944.
- [21] M. Kac, *On applying mathematics: reflections and examples*, Mark Kac: Probability, Number Theory, and Statistical Physics, (K. Baclawski and M. D. Donskder eds.), MIT Press, Cambridge, 1979.
- [22] W. Klingenberg, *Lectures on Closed Geodesics*, Grundlehren Math. Wiss. Vol. 230, Springer, 1978.
- [23] V. P. Maslov, *Theorie des Perturbations et Methods Asymptotiques*, Dunod, Paris, 1972.
- [24] G. Popov, *Length spectrum invariants of Riemannian manifolds*, Math. Z. **213** (1993) 311–351.
- [25] S. Zelditch, *Wave invariants at elliptic closed geodesics*, Geom. Funct. Anal. **7** (1997) 145–213.
- [26] ———, *Wave invariants at non-degenerate closed geodesics*, Geom. Funct. Anal. **8** (1998) 179–217.
- [27] ———, *Isospectrality in the FIO category*, J. Differential Geom. **35** (1992) 689–710.
- [28] ———, *Kuznecov sum formulae and Szegő limit formulae on manifolds*, Comm. Partial Differential Equations **17** (1992) 221–260.

JOHNS HOPKINS UNIVERSITY