CHAPTER I

MAIN RESULTS

1.1. Statement of the spectral problem

The principal objective of this book is the study of the spectral problem

$$Av = \lambda^{2m} v\,, \tag{1.1.1}$$

$$\left(B^{(j)}v\right)\Big|_{\partial M} = 0\,, \qquad j = 1, 2, \dots, m, \tag{1.1.2}$$

where $\lambda > 0$ is the spectral parameter and A is a positive definite self-adjoint elliptic linear differential operator of order $2m$ ($m \in \mathbb{N}$) acting on a compact n-dimensional ($n \geq 2$) manifold M with boundary ∂M or without boundary ($\partial M = \varnothing$). The $B^{(j)}$ are "boundary" linear differential operators describing boundary conditions in the case $\partial M \neq \varnothing$.

In this section we discuss the problem (1.1.1), (1.1.2) and introduce basic notation.

1. Spectral parameter. First of all we must explain the appearance of the $2m$ th power in the right-hand side of (1.1.1). It seems more natural to write

$$Av = \nu v \tag{1.1.1$'$}$$

with ν as spectral parameter, but this leads to fractional powers of ν in the resulting spectral asymptotics. So for the sake of convenience from the very beginning we have made a change of spectral parameter

$$\nu = \lambda^{2m}\,. \tag{1.1.3}$$

There is a discrete set of $\lambda = \lambda_k$, $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \lambda_k \leq \dots$, $\lambda_k \to \infty$, called *eigenvalues*, for which the problem (1.1.1), (1.1.2) has nontrivial solutions v. The number of linearly independent solutions v corresponding to a given λ_k is called the *multiplicity* of λ_k. We always enumerate our eigenvalues taking account of their multiplicities.

2. Coordinates. The compact manifold M and its $(n-1)$-dimensional boundary ∂M are assumed to be infinitely smooth. Without loss of generality M is assumed to be connected (otherwise the problem (1.1.1), (1.1.2) decomposes into separate subproblems). By $\overset{\circ}{M} := M \setminus \partial M$ we denote the interior of M.

By T^*M, $T^*\mathring{M}$, $T^*\partial M$ we denote the respective cotangent bundles, and by $T'M$, $T'\mathring{M}$, $T'\partial M$ the cotangent bundles T^*M, $T^*\mathring{M}$, $T^*\partial M$ with the zero section ($\xi = 0$ or $\xi' = 0$, see below) excluded.

Local coordinates (and points) on M, ∂M are denoted by $x = (x_1, x_2, \dots, x_n)$, $x' = (x_1, x_2, \dots, x_{n-1})$ respectively, their dual coordinates on the fibres T^*_xM, $T^*_x\partial M$ by $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, $\xi' = (\xi_1, \xi_2, \dots, \xi_{n-1})$. Thus, (x, ξ), (x', ξ') are local coordinates on (or points of) the cotangent bundles T^*M, $T^*\partial M$.

Near ∂M we always use special local coordinates $x = (x', x_n)$ such that $\partial M = \{x_n = 0\}$, and $x_n > 0$ for points in $\mathring{M}$; consequently $\xi = (\xi', \xi_n)$. We call x' the *boundary coordinates* (because when $x_n = 0$ the x' are indeed local coordinates on ∂M), ξ' the *coboundary coordinates*, x_n the *"normal" coordinate*, and ξ_n the *conormal coordinate* (note that the covector $\xi = (0, \xi_n)$ is conormal to ∂M).

Moreover, near ∂M we use only coordinates for which x_n is the same on the intersection of different coordinate patches. We fix once and for all the choice of the coordinate x_n near ∂M. One of the possible ways of defining invariantly such a coordinate x_n is with the help of Hamiltonian trajectories introduced in Section 1.3: trajectories originating from points (y, η) with $y \in \partial M$, $\eta' = 0$, can be used for this purpose, with the value of x_n being the time it takes for a trajectory to reach the point $x \in M$. Here $y' \in \partial M$ has to be chosen in such a way that the ray goes through x; this condition defines y' uniquely for x sufficiently close to ∂M and sufficiently small t.

In spectral problems of the type (1.1.1), (1.1.2) arising from mechanical applications there is always an obvious choice of the coordinate x_n near ∂M, which is invariant and convenient. Indeed, such problems are always associated with some natural Riemannian metric, not necessarily directly related to the differential operator A. This metric is either the Euclidean $\mathbb{R}^n$-metric (when $M \subset \mathbb{R}^n$), or an n-dimensional Riemannian metric generated by the Euclidean metric of the embedding $(n+1)$-dimensional space (when M is an n-dimensional surface in $\mathbb{R}^{n+1}$). Then the standard definition of the "normal" coordinate is $x_n := \operatorname{dist}(x, \partial M)$, where $\operatorname{dist}(\cdot, \cdot)$ is the geodesic distance.

If the Hamiltonian is associated with the given Riemannian metric (see formula (1.1.14) and Example 1.2.4 below) then the two definitions of the coordinate x_n given in the two previous paragraphs coincide.

Throughout the book we shall often speak of open sets in $\mathring{M}$ and in M (in fact we have already used the notion of an open set in M describing local coordinates in a neighbourhood of ∂M). This needs some explanation as the notions of an open set in $\mathring{M}$ and in M are different. Local coordinates establish a local diffeomorphism between points of $\mathring{M}$, M and of $\mathbb{R}^n$, $\mathbb{R}^n_+$ respectively. Consequently, the Euclidean notion of convergence in $\mathbb{R}^n$, $\mathbb{R}^n_+$ (considered as metric spaces) generates the notion of convergence in $\mathring{M}$, M with the subsequent standard definition of an open set. Note that under this definition a set which is open in M may contain points from ∂M.

An important example of an open set in M is the set $M_\delta := \{x \in M : x_n < \delta\}$, with $\delta > 0$ sufficiently small. The set M_δ is called the *tubular neighbourhood* of ∂M. This name is due to the fact that M_δ is diffeomorphic to the cylinder $\partial M \times [0, \delta)$. Note, however, that we do not fix this diffeomorphism, because near ∂M we allow changes of coordinates of the type $\widetilde{x}' = \widetilde{x}'(x)$, $\widetilde{x}_n = x_n$, and consequently

$\widetilde{x}'$ depends in the general case on x_n.

Sometimes it will be technically convenient for us to go outside the original manifold M. So by $\widehat{M}$ we shall denote some compact manifold without boundary which is a smooth extension of M, i.e., $M \subset \widehat{M}$; the existence of such an extension is established by a standard doubling argument (make two copies of M and smoothly connect their boundaries by cylinders). By $\widehat{A}$ we shall denote a differential operator on $\widehat{M}$ which is a smooth extension of A. We shall always assume that $\widehat{A}$ is chosen in such a way that it is elliptic and self-adjoint. The coordinates related to $\widehat{M}$ will also be marked by a "wide hat".

Running out of symbols we shall use later on (y,η) instead of (x,ξ) as coordinates on T^*M, and (y',η') instead of (x',ξ') as coordinates on $T^*\partial M$.

3. Symbols of differential operators. Throughout the book A and $B^{(j)}$ are linear differential operators of orders $2m$ and m_j respectively ($m \ge 1$, $m_j \ge 0$ are integers) the coefficients of which are complex-valued infinitely differentiable functions of x, x'. (The set of orders m_j may be different on different connected components of ∂M.) By $A_{2m}(x,\xi)$, $B^{(j)}_{m_j}(x',\xi)$ we denote the principal symbols of the operators A, $(B^{(j)}\,\cdot)\big|_{\partial M}$, i.e., homogeneous polynomials in ξ of degrees $2m$, m_j obtained by leaving only the leading (of orders $2m$, m_j) derivatives in A, $(B^{(j)}\,\cdot)\big|_{\partial M}$, and replacing each $D_{x_k} = -i\partial/\partial x_k$ by ξ_k, $k = 1,2,\dots,n$.

It is well known that under changes of coordinates x the principal symbol $A_{2m}(x,\xi)$ behaves as a function on $T'M$. This means that in new coordinates $\widetilde{x}$ it takes the form

$$\widetilde{A}_{2m}(\widetilde{x},\widetilde{\xi}) \;=\; A_{2m}\left(x(\widetilde{x})\,,\,\sum_{k=1}^{n}\widetilde{\xi}_k\,\frac{\partial\widetilde{x}_k}{\partial x}\bigg|_{x=x(\widetilde{x})}\right)\,.$$

Analogously, the principal symbols $B^{(j)}_{m_j}(x',\xi)$ behave as functions on $T'M|_{\partial M}$:

$$\widetilde{B}^{(j)}_{m_j}(\widetilde{x}',\widetilde{\xi}) \;=\; B^{(j)}_{m_j}\left(x'(\widetilde{x}',0)\,,\,\sum_{k=1}^{n}\widetilde{\xi}_k\,\frac{\partial\widetilde{x}_k}{\partial x}\bigg|_{x=(x'(\widetilde{x}',0),0)}\right)\,.$$

Recall that near ∂M the coordinate x_n is assumed to be fixed once and for all (see subsection 2 above), so $\partial\widetilde{x}_n/\partial x_n \equiv 1$.

Later on we will have to deal with the subprincipal symbol $A_{\mathrm{sub}}(x,\xi)$ of the operator A, which is also a function on $T'M$. The subprincipal symbol is a polynomial in ξ of degree $2m-1$, and it plays the role of the "second" symbol of A. The rigorous definition of $A_{\mathrm{sub}}(x,\xi)$ will be given in subsection 2.1.3.

4. Ellipticity. The problem (1.1.1), (1.1.2) is assumed to be regular elliptic [LioMag, Chap. 2, Sect. 1.4], [RoShSo, Sect. 2.4]. This means that the following four conditions are fulfilled.

Condition 1.1.1. The operator A is elliptic, i.e.

$$A_{2m}(x,\xi) \;\ne\; 0\,, \qquad \forall (x,\xi) \in T'M\,. \tag{1.1.4}$$

CONDITION 1.1.2. The orders of the operators $B^{(j)}$ are different and lower than the order of the operator A, i.e.

$$0 \le m_1 < m_2 < \ldots < m_m \le 2m-1\,. \tag{1.1.5}$$

CONDITION 1.1.3. The operators $B^{(j)}$ can be resolved with respect to their leading conormal derivatives, i.e.

$$B^{(j)}_{m_j}(x',0,\xi_n) \ne 0\,, \qquad j=1,2,\ldots,m\,, \quad \forall x' \in \partial M\,, \quad \forall \xi_n \ne 0\,. \tag{1.1.6}$$

CONDITION 1.1.4 (THE SHAPIRO–LOPATINSKII CONDITION). For all $(x',\xi') \in T'\partial M$ the equation $A_{2m}(x',0,\xi',\xi_n)=0$ has exactly m ξ_n-roots with $\operatorname{Im}\xi_n > 0$ (for $n \ge 3$ this automatically follows from Condition 1.1.1) and $\nu = 0$ is not an eigenvalue (see Definition 1.1.5 below) of the auxiliary one-dimensional spectral problem

$$A_{2m}(x',0,\xi',D_{x_n})v \,=\, \nu v\,, \tag{1.1.7}$$

$$\bigl(B^{(j)}_{m_j}(x',\xi',D_{x_n})v\bigr)\Big|_{x_n=0} = 0\,, \qquad j=1,2,\ldots,m, \tag{1.1.8}$$

on the half-line $x_n \in \mathbb{R}_+ = [0,+\infty)$.

Note that (1.1.7) is an ordinary differential equation with constant coefficients and (x',ξ') comes into the spectral problem (1.1.7), (1.1.8) simply as a parameter.

DEFINITION 1.1.5. We call the number ν an eigenvalue of the problem (1.1.7), (1.1.8) if for this value of ν it has a solution $v(x_n) \not\equiv 0$ which vanishes as $x_n \to +\infty$.

5. Self-adjointness. The problem (1.1.1), (1.1.2) is assumed to be formally self-adjoint with respect to some inner product $(\cdot\,,\cdot)$. Formal self-adjointness means that $(Av,w)=(v,Aw)$ for any $v,w \in C^\infty(M)$ satisfying the boundary conditions (1.1.2).

Usually it is assumed that v, w are functions and the inner product is defined by the formula

$$(v,w) \,=\, \int\limits_M v(x)\overline{w(x)}\mu(x)\,dx \tag{1.1.9}$$

where $\mu \in C^\infty(M)$ is some positive density. We say that v is a *function* if it does not depend on the choice of local coordinates, i.e., $v(x)=\widetilde{v}(\widetilde{x}(x))$ where $\widetilde{x}=\widetilde{x}(x)$ are new local coordinates and $\widetilde{v}$ is the representation of v in the coordinates $\widetilde{x}$. We say that μ is a *density* if $\mu(x)=J(x)\widetilde{\mu}(\widetilde{x}(x))$ where $\widetilde{\mu}$ is the representation of μ in the coordinates $\widetilde{x}$ and $J=|\det \partial\widetilde{x}/\partial x|$. It is easy to see that the integrand in (1.1.9) is independent of the choice of local coordinates and so the inner product is well defined.

However, under such an approach the definition of the inner product and, consequently, the notion of self-adjointness depend on the choice of the density μ. This

is somewhat inconvenient because it subsequently leads to a parasitic dependence of some quantities on μ (see, for example, the definition of the subprincipal symbol in subsection 2.1.3). The following technical device allows us to avoid these inconveniences. Instead of functions and operators acting in spaces of functions one can consider half-densities and operators acting in spaces of half-densities.

We say that v is a *half-density* if $v(x) = J^{1/2}(x)\widetilde{v}(\widetilde{x}(x))$ where $\widetilde{v}$ is the representation of v in the coordinates $\widetilde{x}$. For half-densities the inner product is defined by the formula

$$(v, w) = \int\limits_M v(x)\overline{w(x)}\,dx \tag{1.1.10}$$

(cf. (1.1.9)). Thus, the Hilbert space $L_2(M)$ and, consequently, the concept of self-adjointness are invariantly defined for half-densities without any auxiliary constructions.

The elementary substitution $v \to \mu^{1/2}v$, $A \to \mu^{1/2}A\mu^{-1/2}$, $B^{(j)} \to B^{(j)}\mu^{-1/2}$ transforms the spectral problem (1.1.1), (1.1.2) in functions into a spectral problem in half-densities. So, without loss of generality, we shall assume from now on that v, w are half-densities, A acts in the space of half-densities, $(B^{(j)}\,\cdot\,)\big|_{\partial M}$ are operators acting on half-densities, and the problem (1.1.1), (1.1.2) is formally self-adjoint with respect to the inner product (1.1.10).

Let v be a half-density on M. According to our convention Av is also a half-density on M. A separate question is an invariant interpretation of the expressions $(B^{(j)}v)\big|_{\partial M}$. An invariant interpretation of the expressions $(B^{(j)}v)\big|_{\partial M}$ is not really essential for the formulation of the eigenvalue problem (1.1.1), (1.1.2) because the boundary conditions are homogeneous. It is sufficient to have the boundary conditions (1.1.2) written in local coordinates on a finite number of coordinate patches, with them being equivalent on the intersections of different patches. (Here equivalence means that on intersections of coordinate patches the conditions (1.1.2) in different coordinates define the same linear relationships between $\big(\partial^k v/\partial x_n^k\big)\big|_{x_n=0}$, $k = 0, 1, 2, \dots, 2m-1$, subject to the standard rules of transformation of the half-density v and of its partial derivatives under changes of coordinates.)

However, at some stage (in the course of the effective construction of oscillatory integrals related to our eigenvalue problem) we will be forced to consider various half-densities on M which, taken separately, do not necessarily satisfy the boundary conditions (1.1.2). So it is convenient to assign an invariant meaning to the expressions $(B^{(j)}v)\big|_{\partial M}$. Further on we shall assume that they are half-densities on ∂M. This can always be achieved by an adequate renormalization of the operators $B^{(j)}$ in some local coordinates, with subsequent transformation of these operators under changes of coordinates x' in the appropriate way. The identity operator is an example of an operator $B^{(j)}$ satisfying the required invariance condition: the expression $v|_{\partial M}$ is a half-density on ∂M because our "normal" coordinate x_n is specified once and for all (see subsection 2).

In mechanical applications assigning an invariant interpretation to the expressions $(B^{(j)}v)\big|_{\partial M}$ is usually easy because the boundary conditions have a clear mechanical meaning: they state that some mechanical quantity (say, normal displacement, angle of rotation, normal stress, flexural moment, etc.) is zero. These quantities are either functions or densities on ∂M, so one has only to make an

elementary substitution of the form $B^{(j)} \to (\mu')^{\pm 1/2} B^{(j)}$, where $\mu'(x')$ is some smooth positive density on ∂M.

The principal symbol of the formal adjoint of a differential operator is the complex conjugate of the initial principal symbol. Therefore the principal symbol $A_{2m}(x,\xi)$ of the formally self-adjoint operator A is real.

6. Positiveness. The problem (1.1.1), (1.1.2) is assumed to be positive definite.

DEFINITION 1.1.6. We call the formally self-adjoint problem (1.1.1), (1.1.2) semi-bounded from below if there exists a real constant c such that

$$(Av, v) \ \geq \ c\,(v, v) \tag{1.1.11}$$

for all half-densities $v \in C^\infty(M)$ satisfying the boundary conditions (1.1.2). If (1.1.11) holds with $c > 0$ we call the problem (1.1.1), (1.1.2) positive definite.

Though it is not always easy to check whether a spectral problem is positive definite, one can usually check effectively whether it is semi-bounded from below (i.e., establish a somewhat weaker property). Namely, a regular elliptic formally self-adjoint spectral problem (1.1.1), (1.1.2) is semi-bounded from below if and only if the following two conditions are fulfilled.

CONDITION 1.1.1′. The principal symbol of the operator A is positive, i.e.,

$$A_{2m}(x,\xi) \ > \ 0\,, \qquad \forall (x,\xi) \in T'M \tag{1.1.12}$$

(cf. (1.1.4)).

CONDITION 1.1.4′. The auxiliary one-dimensional spectral problem (1.1.7), (1.1.8) on the half-line $0 \leq x_n < +\infty$ is positive definite, i.e., for all $(x',\xi') \in T'\partial M$ it does not have eigenvalues $\nu \leq 0$.

Indeed, the necessity of Conditions 1.1.1′, 1.1.4′ follows from the fact that if at least one of them does not hold then we can effectively construct (in the form of a linear combination of up to m oscillating or decaying exponential functions modulated by smooth amplitudes) a sequence of half-densities $w_k(x) \in C^\infty(M)$, $k = 1, 2, \dots$, satisfying boundary conditions (1.1.2) and such that $(Aw_k, w_k)/(w_k, w_k) \to -\infty$. Sufficiency is proved, for example, in [AgrVi].

It is easy to see that Conditions 1.1.1, 1.1.4 follow from Conditions 1.1.1′, 1.1.4′. This means that formulating the notion of a semi-bounded from below self-adjoint elliptic boundary value problem we can replace Conditions 1.1.1, 1.1.4 by Conditions 1.1.1′, 1.1.4′.

REMARK 1.1.7. In practice it is quite enough to establish whether the spectral problem under consideration is semi-bounded from below. Indeed, any problem semi-bounded from below can be turned into a positive definite one by the elementary substitution

$$\widetilde{A} = A - cI\,, \qquad \widetilde{\nu} = \nu - c = \widetilde{\lambda}^{2m} \tag{1.1.13}$$

(cf. (1.1.11)); here I is the identity operator. Moreover, writing the resulting spectral asymptotics (see Sections 1.2, 1.4, 1.6, 1.7, 1.8) in terms of the spectral parameter $\widetilde{\nu} \to +\infty$ one may notice that all these asymptotics admit a formal interchange $\widetilde{\nu} \leftrightarrow \nu$. This is due to the fact that replacing $\widetilde{\nu}$ by ν we introduce an additional relative error of $o(\widetilde{\nu}^{-1})$, whereas the inherent relative error of these asymptotics is at least $o(\widetilde{\lambda}^{-1}) \equiv o(\widetilde{\nu}^{-1/2m})$ which is obviously greater than $o(\widetilde{\nu}^{-1})$.

7. Statement of our problem in the framework of operator theory. Let $H^{2m}(M)$ denote the Sobolev space of half-densities which belong to $L_2(M)$ together with all their partial derivatives of order $\le 2m$. Under the conditions described above the differential operator A initially defined on

$$\{v \in C^\infty(M) \,:\, \bigl(B^{(j)}v\bigr)\Big|_{\partial M} = 0\,, \quad j = 1, 2, \dots, m\}$$

is essentially self-adjoint and admits a self-adjoint closure $\mathcal{A}$ in $L_2(M)$ with domain of definition

$$D(\mathcal{A}) = \{v \in H^{2m}(M) \,:\, \bigl(B^{(j)}v\bigr)\Big|_{\partial M} = 0\,, \quad j = 1, 2, \dots, m\}\,.$$

It is well known [RoShSo, Sect. 5.1] that the operator $\mathcal{A}$ has a positive discrete spectrum $0 < \nu_1 \le \nu_2 \le \dots$ accumulating to $+\infty$ (we enumerate the eigenvalues taking into account their multiplicities). The numbers $\lambda_k = \nu_k^{1/2m}$, $k = 1, 2, \dots$ (see (1.1.1′), (1.1.3)) may be interpreted as the eigenvalues of the operator $\mathcal{A}^{1/(2m)}$. It is also well known that the respective eigenfunctions (more precisely, half-densities) v_k are infinitely smooth on M, satisfy (1.1.1), (1.1.2) and form an orthonormal basis in $L_2(M)$.

Remark 1.1.8. Throughout the book we denote by A the differential expression from the left-hand side of (1.1.1), i.e., we do not normally assign A any particular domain of definition in any particular space. However, for the sake of simplicity we refer to A as to an operator. This should not cause confusion because the real operator $\mathcal{A}$ is distinguished by different script.

Example 1.1.9. Let A be a formally self-adjoint second order differential operator (more precisely, differential expression) with $A_2(x, \xi) > 0$. Then the Dirichlet and Neumann boundary value problems for A satisfy Conditions 1.1.1′, 1.1.2, 1.1.3, 1.1.4′ and generate self-adjoint operators $\mathcal{A}$.

Example 1.1.10. Let A be a formally self-adjoint differential operator of order $2m$ with $A_{2m}(x, \xi) > 0$. Then the Dirichlet boundary value problem ($B^{(j)} = \partial^{j-1}/\partial x_n^{j-1}$, $j = 1, 2, \dots, m$) for A satisfies Conditions 1.1.1′, 1.1.2, 1.1.3, 1.1.4′ and generates a self-adjoint operator $\mathcal{A}$.

8. Pseudodifferential case. For manifolds without boundary ($\partial M = \varnothing$) one can also consider a somewhat more general case when A is a positive definite self-adjoint elliptic pseudodifferential operator (see Section 2.1) of positive order $2m$; here m is an arbitrary positive real number. In this case the spectrum of the problem (1.1.1) remains discrete, positive and accumulating to $+\infty$. Subsequent asymptotic analysis differs insignificantly from the case of a differential operator; only the reasoning from Remark 1.1.7 suffers for $2m \le 1$. Note that spectral problems for pseudodifferential operators are a normal occurrence in applications, see Sections 6.4, 6.5.

9. Mechanical interpretation. In mechanical and physical applications the spectral problem (1.1.1) (if $\partial M = \varnothing$) or (1.1.1), (1.1.2) (if $\partial M \ne \varnothing$) usually describes free harmonic oscillations of some system (elastic body, resonator, etc.) with ν being the frequency parameter proportional to some integer power of the natural frequency. Self-adjointness of the spectral problem means conservation of the full energy in the oscillating system, and positiveness means stability (i.e., absence of movements with amplitude exponentially growing in time). Some examples arising from applications will be considered in Chapter 6.

10. Canonical differential forms and measures. In order to formulate our results we will have to deal with some invariant differential forms and measures (see [Ar2], [Tr, vol. 2], [Hö3, vol. 3] for details).

The differential forms on T^*M, $T^*\partial M$ defined in local coordinates by

$$\langle \xi, dx \rangle \,=\, \xi_1 dx_1 \,+\, \xi_2 dx_2 \,+\, \ldots \,+\, \xi_n dx_n\,,$$

$$\langle \xi', dx' \rangle \,=\, \xi_1 dx_1 \,+\, \xi_2 dx_2 \,+\, \ldots \,+\, \xi_{n-1} dx_{n-1}$$

respectively, are called symplectic 1-forms. One can easily check that these forms are independent of the choice of local coordinates. Their differentials are the canonical 2-forms

$$d\xi \wedge dx \;=\; d\xi_1 \wedge dx_1 \,+\, d\xi_2 \wedge dx_2 \,+\, \ldots \,+\, d\xi_n \wedge dx_n\,,$$

$$d\xi' \wedge dx' \,=\, d\xi_1 \wedge dx_1 \,+\, d\xi_2 \wedge dx_2 \,+\, \ldots \,+\, d\xi_{n-1} \wedge dx_{n-1}\,.$$

As usual we denote

$$dx \,=\, dx_1\, dx_2\, \ldots\, dx_n\,, \qquad d\xi \,=\, d\xi_1\, d\xi_2\, \ldots\, d\xi_n\,,$$

$$dx' \,=\, dx_1\, dx_2\, \ldots\, dx_{n-1}\,, \qquad d\xi' \,=\, d\xi_1\, d\xi_2\, \ldots\, d\xi_{n-1}\,.$$

Then the elements of symplectic volumes $\mathrm{vol}(\cdot)$, $\mathrm{vol}'(\cdot)$ on the cotangent bundles T^*M, $T^*\partial M$ are defined in local coordinates as $dx\, d\xi$, $dx'\, d\xi'$. Obviously, these expressions are also independent of the choice of coordinates.

We shall also use the notation

$$đ\xi = (2\pi)^{-n}\, d\xi, \qquad dx\, đ\xi = (2\pi)^{-n}\, dx\, d\xi,$$

$$đ\xi' = (2\pi)^{1-n}\, d\xi', \qquad dx'\, đ\xi' = (2\pi)^{1-n}\, dx'\, d\xi'\,.$$

Throughout the book we shall extensively use the function

$$h(x,\xi) \,=\, (A_{2m}(x,\xi))^{1/(2m)} > 0 \tag{1.1.14}$$

which we shall call the *Hamiltonian*. The set

$$S^*M \,=\, \{(x,\xi) \in T'M \,:\, h(x,\xi) \,=\, 1\}$$

is said to be the *unit cosphere bundle*. Having fixed $x \in M$, we shall also deal with the *unit cosphere*

$$S^*_x M \,=\, \{\xi \in T'_y M \,:\, h(x,\xi) \,=\, 1\}$$

(the fibre of S^*M over the point x).

There exists a natural measure meas_x on the unit cosphere $S^*_x M$ with element $d\widetilde{\xi}$ defined by the equality $d\xi = d\widetilde{\xi}\, dh$. Note that the Euclidean measure dS_x on $S^*_x M$ is defined by $d\xi = dS_x\, d\rho$, where $\rho(\xi)$ is the Euclidean distance from ξ to $S^*_x M$, so $d\widetilde{\xi} = |\nabla_\xi h|^{-1}\, dS_x$. Equivalently, one can define $d\widetilde{\xi}$ by the condition that for any function $f(\xi)$ positively homogeneous in ξ of degree $d > -n$

$$(n+d) \int\limits_{h(x,\xi)\le 1} f\, d\xi \,=\, \int\limits_{h(x,\xi)=1} f\, d\widetilde{\xi}\,. \tag{1.1.15}$$

Clearly, for all measurable sets $\Omega \subset S^*_x M$ and functions f defined on $S^*_x M$, the integrals $\int_\Omega f \, d\widetilde{\xi}$ (in particular, $\operatorname{meas}_x \Omega_x = \int_\Omega d\widetilde{\xi}$) depend on the choice of the coordinates x and behave as densities in x under change of coordinates. This means that the measure meas_x takes its values in the space of densities on M.

The natural measure meas on the unit cosphere bundle S^*M is defined as $dx\, d\widetilde{\xi}$. We shall denote

$$đ\widetilde{\xi} \;=\; (2\pi)^{-n}\, d\widetilde{\xi}\,, \qquad dx\, đ\widetilde{\xi} = (2\pi)^{-n}\, dx\, d\widetilde{\xi}\,.$$

The element $d\widetilde{\xi}$ of the measure meas_x can be written in local coordinates $\widetilde{\xi}$ on $S^*_x M$. In particular, on a conic subset of $T'_x M$ on which $h_{\xi_n} \neq 0$ one can choose as local coordinates $(\widetilde{\xi}, h)$, where

$$\widetilde{\xi} \;=\; \bigl(\widetilde{\xi}_1\,, \widetilde{\xi}_2\,, \ldots\,, \widetilde{\xi}_{n-1}\bigr) \;=\; \bigl(\xi_1/h\,, \xi_2/h\,, \ldots\,, \xi_{n-1}/h\bigr)\,.$$

Then $\widetilde{\xi}$ are local coordinates on $S^*_x M$ and

$$d\widetilde{\xi} \;=\; (h_{\xi_n})^{-1} d\widetilde{\xi}_1 d\widetilde{\xi}_2 \ldots d\widetilde{\xi}_{n-1}\,.$$

1.2. One-term asymptotic formula for $N(\lambda)$

1. Statement of the result. Let us introduce the *counting function* $N(\lambda)$ which is defined as the number of eigenvalues λ_k of the problem (1.1.1), (1.1.2) less than a given λ:

$$N(\lambda) \;=\; \#\{k : \lambda_k < \lambda\}\,.$$

Our final aim is to describe the asymptotic behaviour of $N(\lambda)$ as $\lambda \to +\infty$. Of course one can afterwards invert such formulae (see Remarks 1.2.2 and 1.6.2) and obtain asymptotic formulae for λ_k as $k \to +\infty$.

The following theorem gives a one-term asymptotic formula for $N(\lambda)$. Under additional restrictions it was established in [Hö1] ($\partial M = \varnothing$) and [Se1], [Se2] ($m = 1$). The final result appeared in [Va3], [Va4], [Va7].

THEOREM 1.2.1.

$$N(\lambda) \;=\; c_0 \lambda^n \,+\, O(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \tag{1.2.1}$$

where

$$c_0 \;=\; (2\pi)^{-n}\, \operatorname{vol}\{(x,\xi) : A_{2m}(x,\xi) \le 1\} \;=\; \int\limits_{A_{2m} \le 1} dx\, đ\xi\,. \tag{1.2.2}$$

By (1.1.15) the Weyl constant (1.2.2) can also be written as

$$c_0 \;=\; \frac{1}{n\,(2\pi)^n}\, \operatorname{meas} S^*M \;=\; \frac{1}{n} \int\limits_{S^*M} dy\, đ\widetilde{\eta}\,. \tag{1.2.2$'$}$$

REMARK 1.2.2. Formula (1.2.1) can be written down in the equivalent form

$$\lambda_k \;=\; (c_0)^{-1/n} k^{1/n} \,+\, O(1)\,, \qquad k \to \infty\,. \tag{1.2.1$'$}$$

EXAMPLE 1.2.3. Let M be a region in the Euclidean space $\mathbb{R}^n$ and $A = (-\Delta)^m$, where

$$\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \ldots + \partial^2/\partial x_n^2$$

is the Laplacian in Cartesian coordinates. In this case $A_{2m} = |\xi|^{2m}$ and

$$c_0 = (2\pi)^{-n}\omega_n \operatorname{Vol} M\,, \tag{1.2.3}$$

where $\operatorname{Vol} M$ is the n-dimensional volume of M and ω_n is the volume of the unit ball in $\mathbb{R}^n$. In particular, for $n = 2$ we have $c_0 = S/4\pi$ where S is the surface area of M, and for $n = 3$ we have $c_0 = V/6\pi^2$ where V is the three-dimensional volume of M.

EXAMPLE 1.2.4 (GENERALIZATION OF EXAMPLE 1.2.3). Let M be a Riemannian manifold with metric tensor $\{g_{ij}\}$, $1 \le i, j \le n$, and contravariant metric tensor $\{g^{ij}\} = \{g_{ij}\}^{-1}$. Let $A = (-\Delta)^m$ where

$$\Delta = g^{-1/4}\left(\sum_i \frac{\partial}{\partial x_i}\left(\sum_j g^{ij}\sqrt{g}\,\frac{\partial}{\partial x_j}\right)\right) g^{-1/4}$$

is the Laplacian on half-densities; here $g := \det\{g_{ij}\}$, and $\sqrt{g}$ is the standard Riemannian density. In this case $A_{2m}(x,\xi) = |\xi|_x^{2m}$ where $|\xi|_x = \left(\sum_{i,j} g^{ij}\xi_i\xi_j\right)^{1/2}$, and formula (1.2.3) holds with $\operatorname{Vol} M = \int\limits_M \sqrt{g}\,dx$.

2. Discussion of the result. Theorem 1.2.1 is remarkable in two ways. First of all, the one-term asymptotic formula (1.2.1) does not depend on the boundary conditions (1.1.2). This fact suggests that in practice the asymptotic formula (1.2.1) is not very accurate and that it is natural to search for a refined two-term asymptotic formula with the second asymptotic term describing boundary phenomena.

The second remarkable fact about Theorem 1.2.1 is that the remainder estimate in (1.2.1) is sharp, that is, in (1.2.1) one can not replace $O(\lambda^{n-1})$ by $o(\lambda^{n-1})$. This will become evident after we single out the second asymptotic term in Sections 1.6, 1.7; as we shall see, the second term is of the order of λ^{n-1} and in the general situation is not identically zero. However, the simplest way of checking that the remainder estimate in (1.2.1) is sharp is by analyzing the following simple example.

EXAMPLE 1.2.5. Let M be a unit n-dimensional sphere,

$$M = \mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}\,,$$

and $A = (-\Delta + (n-1)^2/4)^m$, where Δ is the Laplacian (see Example 1.2.4). In this case the spectrum of the operator $\mathcal{A}^{1/2m}$ consists of the eigenvalues

$$\Lambda_j = j + (n-1)/2, \qquad j = 0, 1, 2, \ldots,$$

with multiplicities

$$\frac{(n+j-2)!\,(n+2j-1)}{(n-1)!\,j!}$$

(see, e.g., [Sh, Sect. 22]). Therefore

$$\lambda_k = j + (n-1)/2, \qquad k = 1, 2, \dots,$$

where $j = j(k)$ is the maximal integer satisfying the inequality

$$\sum_{i=0}^{j} \frac{(n+i-2)!\,(n+2i-1)}{(n-1)!\,i!} \le k\,.$$

Elementary calculations show that

$$N(\Lambda_j) = \frac{2}{n!}\Lambda_j^n - \frac{n^2-3n+3}{(n-1)!}\Lambda_j^{n-1} + O(\Lambda_j^{n-2})\,, \tag{1.2.4}$$

$$N(\Lambda_j + 0) = \frac{2}{n!}\Lambda_j^n + \frac{n^2-3n+3}{(n-1)!}\Lambda_j^{n-1} + O(\Lambda_j^{n-2})\,, \tag{1.2.5}$$

(here $c_0 = 2/n!$). Formulae (1.2.4), (1.2.5) demonstrate that in the general case the order of the remainder estimate in (1.2.1) can not be decreased.

3. Pseudodifferential case. Theorem 1.2.1 remains true in the case of a pseudodifferential operator acting on a manifold without boundary [Hö1], [DuiGuiHö].

1.3. Hamiltonian billiards I: basic definitions and results

Singling out the second term in the asymptotic formula for $N(\lambda)$ requires the study of some global geometric characteristics of the problem. These characteristics are formulated in terms of billiard (or Hamiltonian) trajectories generated by the principal symbol of the differential operator A. In this section we introduce the concept of branching Hamiltonian billiards and describe their basic properties. More detailed discussions and proofs are contained in Appendix D, and, partially, in various sections of the main text.

The notions introduced in this section are necessary for the formulation of all our results on two-term spectral asymptotics (Sections 1.6–1.8).

The concept of branching Hamiltonian billiards is a generalization of the classical concepts of Hamiltonian flows and geodesic billiards. Therefore we shall first consider Hamiltonian flows (subsection 1) and geodesic billiards (subsection 2); branching Hamiltonian billiards as such will appear in subsection 3.

1. Hamiltonian flow. Let us consider the case when M is a manifold without boundary.

For $t \in \mathbb{R}$ we shall denote by

$$(x^*(t; y, \eta), \xi^*(t; y, \eta)) \tag{1.3.1}$$

the integral curve of the Hamiltonian vector field generated by the Hamiltonian (1.1.14). The curve (1.3.1) lies in $T'M$ and (y, η) denotes its starting point, i.e.,

$$(x^*(0; y, \eta), \xi^*(0; y, \eta)) = (y, \eta) \in T'M\,. \tag{1.3.2}$$

In other words, (1.3.1) is the solution of the Hamiltonian system of equations

$$\dot{x}^* = h_\xi(x^*, \xi^*), \qquad \dot{\xi}^* = -h_x(x^*, \xi^*) \tag{1.3.3}$$

with initial condition (1.3.2) at $t = 0$. Here and further on the dot stands for the time derivative, and subscripts are used for denoting respective partial derivatives.

In our notation for Hamiltonian trajectories we shall sometimes omit (for the sake of brevity) the variables (y, η), leaving only the variable t; i.e., we shall sometimes write $(x^*(t), \xi^*(t))$ instead of (1.3.1).

Denote by Φ^t the group of shifts along the trajectories (1.3.1). This group is called the *Hamiltonian flow*. The group Φ^t preserves the symplectic differential 1-form and, consequently, the symplectic differential 2-form and the symplectic volume vol (see Section 2.3).

We obviously have

$$\frac{d}{dt}\, h(x^*, \xi^*) \;=\; \langle h_x, \dot{x}\rangle \,+\, \langle h_\xi, \dot{\xi}\rangle \;=\; 0\,,$$

so

$$h(x^*(t; y, \eta), \xi^*(t; y, \eta)) \;=\; h(y, \eta)\,,$$

i.e., the Hamiltonian h is preserved by the Hamiltonian flow Φ^t. Since the Hamiltonian h is positively homogeneous in ξ of degree 1, the Hamiltonian flow is also homogeneous:

$$(x^*(t; y, \mu\eta), \xi^*(t; y, \mu\eta)) \;=\; (x^*(t; y, \eta), \mu\xi^*(t; y, \eta))\,, \qquad \forall \mu > 0\,. \tag{1.3.4}$$

Therefore it is convenient to consider the restriction of the Hamiltonian flow Φ^t to the unit cosphere bundle S^*M. We retain for this restriction the original notation Φ^t. Since $dy\, d\eta = dy\, d\widetilde{\eta}\, dh$, the flow Φ^t preserves the canonical measure meas on S^*M.

We shall call the curves $x^*(t)$ *rays* (by analogy with geometric optics). Note that $\dot{x}^* \neq 0$, i.e., the ray cannot stop. This fact follows from Euler's formula $\langle \xi, h_\xi\rangle = h \neq 0$ and the conservation of the Hamiltonian.

In the special case when M is a Riemannian manifold and $h(x, \xi) = |\xi|_x$ (see Example 1.2.4) the Hamiltonian flow is called *geodesic* flow. In this situation the rays x^* are geodesics and $\xi^*_i = \sum_j g_{ij}(x^*)\, \dot{x}^*_j$.

Having defined the Hamiltonian flow Φ^t we can now introduce the notions of periodicity and absolute periodicity which are decisive in obtaining refined spectral asymptotics.

Definition 1.3.1. The point $(y, \eta) \in T'M$ and the trajectory (1.3.1) originating from this point are called *T-periodic*, $T > 0$, if

$$(x^*(T; y, \eta), \xi^*(T; y, \eta)) \;=\; (y, \eta)\,. \tag{1.3.5}$$

The point (y, η) and the trajectory (1.3.1) are called *periodic* if they are T-periodic for some $T > 0$.

Definition 1.3.2. The point $(y_0, \eta_0) \in T'M$ and the trajectory (1.3.1) originating from this point are called *absolutely T-periodic*, $T > 0$, if the function

$$|x^*(T; y, \eta) - y|^2 \,+\, |\xi^*(T; y, \eta) - \eta|^2 \tag{1.3.6}$$

of the variables (y, η) has an infinite order zero at (y_0, η_0) (clearly, this property does not depend on the choice of coordinates). The point (y_0, η_0) and the trajectory

(1.3.1) originating from this point are called *absolutely periodic* if they are absolutely T-periodic for some $T > 0$.

REMARK 1.3.3. In the analytic case (when M is a real-analytic manifold and h is a real-analytic function) the function (1.3.6) is analytic as well. Therefore in the analytic case the existence of one absolutely T-periodic trajectory implies that all the trajectories are absolutely T-periodic with the same T. (Recall that the manifold M is assumed to be connected.)

It is important to know how rich are the sets of periodic and absolutely periodic trajectories, because this determines the structure of the two-term spectral asymptotics.

Denote by Π_T, Π_T^a, Π, and Π^a the sets of T-periodic, absolutely T-periodic, periodic, and absolutely periodic points in S^*M, respectively. Of course,

$$\Pi = \bigcup_{T>0} \Pi_T\,, \qquad \Pi^a = \bigcup_{T>0} \Pi_T^a\,, \qquad \Pi_T^a \subset \Pi_T\,, \qquad \Pi^a \subset \Pi\,.$$

LEMMA 1.3.4. *The set $\Pi \subset S^*M$ is measurable. Moreover, for all $T_+ > 0$ the sets $\bigcup_{0<T\le T_+} \Pi_T$ are also measurable.*

As a rule, the set Π has a very complicated structure whereas the set Π_a is organised essentially more simply (see Remark 1.3.3). Moreover, in realistic situations the set of absolutely periodic points is usually empty. Nevertheless, the following lemma shows that the set Π is insignificantly richer than Π^a.

LEMMA 1.3.5.

$$\operatorname{meas}\left(\bigcup_{T>0}(\Pi_T \setminus \Pi_T^a)\right) = 0\,. \tag{1.3.7}$$

From (1.3.7) we immediately obtain

COROLLARY 1.3.6. *Almost all periodic points are absolutely periodic, i.e.,*

$$\operatorname{meas}(\Pi \setminus \Pi^a) = 0\,. \tag{1.3.8}$$

Now we are prepared to introduce the notion of nonperiodicity which will play a key role in the study of refined spectral asymptotics.

DEFINITION 1.3.7. We will say that the *nonperiodicity condition* is fulfilled if

$$\operatorname{meas}\Pi = 0\,. \tag{1.3.9}$$

It follows from Corollary 1.3.6 that (1.3.9) is equivalent to

$$\operatorname{meas}\Pi^a = 0\,. \tag{1.3.10}$$

Therefore, defining the concept of nonperiodicity we can choose between (1.3.9) and (1.3.10) depending on the circumstances. This observation often allows us to simplify matters. For example, when proving spectral results it is more convenient to deal with the seemingly more restrictive condition (1.3.9). On the other hand, applying spectral results to concrete problems it is much easier to check the condition (1.3.10). In particular, from Remark 1.3.3 and Corollary 1.3.6 we obtain

LEMMA 1.3.8. *In the analytic case there are only two possibilities: either the nonperiodicity condition is fulfilled or all the trajectories are periodic with the same period* $T > 0$.

It is clear that failures of the nonperiodicity condition are very rare (at least in the analytic situation). For instance, in the analytic case it is sufficient to find one nonperiodic trajectory or two periodic trajectories with incommensurable periods to make sure that the nonperiodicity condition is fulfilled.

Failure of the nonperiodicity condition is usually due to some strong symmetries. The most obvious example is the geodesic flow on the sphere $\mathbb{S}^n$. However, there exist nontrivial examples.

EXAMPLE 1.3.9. Let M be a closed (i.e., without boundary) connected compact real-analytic two-dimensional surface of revolution in $\mathbb{R}^3$ without self-intersections. We shall call a surface M of this class a *Zoll surface* if there exists a $T > 0$ such that all the geodesics on M are T-periodic (i.e., if the nonperiodicity condition fails). Following [Be] we explicitly describe below all Zoll surfaces.

Let us introduce cylindrical coordinates (φ, r, z) according to the formulae

$$y_1 = r\cos\varphi\,, \qquad y_2 = r\sin\varphi\,, \qquad y_3 = z\,,$$

where (y_1, y_2, y_3) are the Cartesian coordinates in $\mathbb{R}^3$ and the y_3-axis is the axis of revolution. Let us consider the meridian curve of M in the half-plane

$$\varphi = \text{const}\,, \qquad 0 \le r < +\infty\,, \qquad -\infty < z < +\infty\,.$$

It turns out that the surface M is a Zoll surface if and only if its meridian curve can be parametrized as

$$r = R\cos\theta\,, \quad z = R\int_0^\theta \sqrt{\big(1+f(\psi)\big)^2 - \sin^2\psi}\,d\psi\,, \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}\,, \tag{1.3.11}$$

with some constant $R > 0$ and some real-valued function f satisfying the following conditions.

(1) f is odd.
(2) $f(\psi) \ge |\sin\psi| - 1\,,\ \forall\psi \in [-\pi/2, \pi/2]$.
(3) f is real-analytic on $\mathbb{R}$.
(4) f is even with respect to the point $\pi/2$.
(5) The Taylor expansions of the function $\big(1+f(\psi)\big)^2 - \sin^2\psi$ at the points $\pm\pi/2$ start with terms of order $4q_\pm - 2$, where $q_\pm \in \mathbb{N}$.
(6) The function z defined by (1.3.11) satisfies the inequality $z(\theta) > z(-\theta)\,,$ $\forall\theta \in (0, \pi/2]$.

We choose the sign of the the square root in (1.3.11) in such a way that the function $F(\psi) = \sqrt{\big(1+f(\psi)\big)^2 - \sin^2\psi}$ is real-analytic and $F(0) > 0$.

Conditions (1), (2) are the main ones which single out the class of Zoll surfaces. The other ones play only an auxiliary role, ensuring the analyticity of the surface M and excluding self-intersections.

A trivial example of a function f satisfying all the above conditions is $f \equiv 0$. In this case (1.3.11) parametrizes a semicircle which means that M is a sphere. A

nontrivial example is $f(\theta) = (1/3)\sin\theta\cos^2\theta$; the Zoll surface for this f is shown on Fig. 1. It gives a side view of the Zoll surface. The external egg-shaped curve is the meridian of the Zoll surface, and the internal boomerang-shaped curve is the projection of a geodesic (namely, the geodesic intersecting the equator at the angle $\pi/4$).

Figure 1. Example of a Zoll surface. Side view.

The reader can find other examples of manifolds with periodic geodesic flow and interesting discussions of this subject in [Be], [Gui].

Now we proceed to manifolds with boundary. For the sake of simplicity we shall consider only trajectories $(x^*(t;y,\eta),\xi^*(t;y,\eta))$ originating from points $(y,\eta) \in T'\mathring{M}$ (not $T'M$), i.e., strictly from the interior. This will be sufficient for our aims because, as we shall see later, only sets of nonzero measure influence the second asymptotic term of $N(\lambda)$, whereas

$$\operatorname{vol}\left(T'M\big|_{\partial M}\right) \,=\, \operatorname{meas}\left(S^*M\big|_{\partial M}\right) \,=\, 0\,.$$

2. Geodesic billiards. Let M be a Riemannian manifold with boundary, and let $h(x,\xi) = |\xi|_x$ (see Example 1.2.4). In this case the Hamiltonian billiards described below are called *geodesic billiards*.

The trajectory $(x^*(t),\xi^*(t))$ may hit the boundary at some time $t = \tau \neq 0$ (i.e., $x^*(\tau) \in \partial M$) and the question arises how to extend it. Geometric optics gives a natural reflection law: the angle of incidence of a ray equals the angle of reflection. Certainly, the notion of angle is understood in the sense of Riemannian geometry, i.e, at the point x_0 angles are measured in local coordinates x in which $|\xi|_{x_0} = |\xi| = \sqrt{\xi_1^2+\xi_2^2+\cdots+\xi_n^2}$.

We shall now reformulate the *reflection law* in a more explicit manner, in terms of Hamiltonian trajectories. Consider the points $(x^*(\tau-0),\xi^*(\tau-0))$ and $(x^*(\tau+0),\xi^*(\tau+0))$ which we shall call the *point of incidence* and *point of reflection*, respectively. As $\xi^*(t)$ determines the direction of the ray $x^*(t)$, the ξ-components of these two points are, generally speaking, different, and our aim is to determine the unknown covector ξ, which is $\xi^*(\tau+0)$ or $\xi^*(\tau-0)$ depending on whether $\tau > 0$ or $\tau < 0$.

Let us impose the natural condition that the Hamiltonian h is preserved under reflection, i.e.,

$$h(x^*(\tau-0),\xi^*(\tau-0)) \,=\, h(x^*(\tau+0),\xi^*(\tau+0))\,. \tag{1.3.12}$$

From this equality and from the geometric reflection law it follows that the points of incidence and reflection differ only in their ξ_n-components. Therefore we can

denote these points as

$$(x^*(\tau-0),\xi^*(\tau-0)) \,=\, (x^{*\prime},0,\xi^{*\prime},\xi_n^-)$$

and

$$(x^*(\tau+0),\xi^*(\tau+0)) \,=\, (x^{*\prime},0,\xi^{*\prime},\xi_n^+)\,,$$

respectively (recall that near ∂M we use special local coordinates x, see subsection 1.1.2). Now we can rewrite (1.3.12) as

$$h(x^{*\prime},0,\xi^{*\prime},\xi_n^-) \,=\, h(x^{*\prime},0,\xi^{*\prime},\xi_n^+)\,. \tag{1.3.13}$$

This is the equation which describes the reflection law in terms of Hamiltonian trajectories. Note that both the equation (1.3.13) and the condition that the points of incidence and reflection differ only in their conormal components are invariant under changes of coordinates x.

Having fixed $(x^{*\prime},\xi^{*\prime})$, let us consider $h(\xi_n) := h(x^{*\prime},0,\xi^{*\prime},\xi_n)$ as a function of the variable $\xi_n\in\mathbb{R}$. Since h is the square root of a quadratic polynomial it has a unique local (and global) minimum at some point $\xi_n=\xi_n^{\rm st}$. As x_n^* increases from ∂M into the interior of M and $\dot x_n^* = h_{\xi_n}$, we have

$$\xi_n^-\le\xi_n^{\rm st}\,,\qquad \xi_n^+\ge\xi_n^{\rm st}\,, \tag{1.3.14}$$

$$h_{\xi_n}(\xi_n^-)\le 0\,,\qquad h_{\xi_n}(\xi_n^+)\ge 0\,. \tag{1.3.15}$$

The qualities $\xi_n^- = \xi_n^{\rm st}$ $(\tau>0)$ or $\xi_n^+ = \xi_n^{\rm st}$ $(\tau<0)$ describe the situation when the arriving (from $t=0$) ray $x^*(t)$ is tangent to ∂M at the point $x^*(\tau\mp 0)$. Such trajectories trajectories $(x^*(t),\xi^*(t))$ are called *grazing*. We shall exclude this case from consideration because the set of grazing trajectories is sparse (see Lemma 1.3.11 below), and for our aims there is no need to deal with them.

Having excluded grazing trajectories, we obtain instead of (1.3.14), (1.3.15) strict inequalities

$$\xi_n^-<\xi_n^{\rm st}\,,\qquad \xi_n^+>\xi_n^{\rm st}\,, \tag{1.3.16}$$

$$h_{\xi_n}(\xi_n^-)< 0\,,\qquad h_{\xi_n}(\xi_n^+)> 0\,. \tag{1.3.17}$$

Formulae (1.3.13), (1.3.16), (1.3.17) are illustrated by Fig. 2.

Figure 2. Graph of the function $h(\xi_n)$ in the case of a geodesic billiard system.

It is clear from Fig. 2 that, given ξ_n^-, one can uniquely determine ξ_n^+. This means that for $\tau>0$ we uniquely determine $(x^*(\tau+0),\xi^*(\tau+0))$ which serves as

the originating point of the reflected Hamiltonian trajectory. Conversely, if $\tau < 0$ then given ξ_n^+ one can uniquely determine $(x^*(\tau - 0), \xi^*(\tau - 0))$ which is the starting point of the extended Hamiltonian trajectory. In the second case we have deliberately avoided the word "reflected" because according to our terminology the point of reflection is $(x^*(\tau+0), \xi^*(\tau+0))$, whereas we are extending the trajectory back in time from the point $(x^*(\tau - 0), \xi^*(\tau - 0))$.

Exclusion of grazing trajectories means that we consider only *transversal* reflections, i.e., reflections at nonzero angles. Fig. 3 illustrates transversal reflections of a ray.

Figure 3. Transversal reflections.

DEFINITION 1.3.10. A trajectory obtained by consecutive transversal reflections is called a *billiard* trajectory.

We shall retain for billiard trajectories the same notation $(x^*(t; y, \eta), \xi^*(t; y, \eta))$ as for Hamiltonian trajectories.

As the Hamiltonian h is preserved along the billiard trajectories and as these trajectories are homogeneous in η (see (1.3.4)), it is natural to consider only trajectories lying in S^*M.

It is easy to see that only two factors may prevent us from extending a billiard trajectory $(x^*(t; y, \eta), \xi^*(t; y, \eta))$ to all $t \in \mathbb{R}$.

First of all, it may occur that after a finite number of transversal reflections we obtain a grazing Hamiltonian trajectory. We shall denote the set of starting points $(y, \widetilde{\eta}) \in S^*\overset{\circ}{M}$ of such billiard trajectories by $\mathcal{P}^{\mathrm{g}}$.

Secondly, it may occur that a billiard trajectory experiences an infinite number of transversal reflections in a finite (positive or negative) time; see Fig. 4.

Figure 4. A dead-end trajectory.

We shall call such a billiard trajectory a *dead-end* trajectory. By $\mathcal{P}^{\mathrm{d}}$ we shall denote the set of starting points $(y, \widetilde{\eta}) \in S^*\overset{\circ}{M}$ of dead-end billiard trajectories.

LEMMA 1.3.11 ([CorFomSin, Sect. 6.1]). *For a geodesic billiard system*

$$\operatorname{meas} \mathcal{P}^{\mathrm{g}} = 0 \tag{1.3.18}$$

and

$$\operatorname{meas} \mathcal{P}^{\mathrm{d}} = 0\,. \tag{1.3.19}$$

Lemma 1.3.11 means that almost all billiard trajectories can be extended to all $t \in \mathbb{R}$. The group Φ^t of shifts along the billiard trajectories is called the *billiard flow*. The billiard flow preserves the symplectic differential 1-form and, consequently, the symplectic differential 2-form, the symplectic volume and the canonical measure on S^*M.

Adding the phrase "after a finite number of transversal reflections" in Definitions 1.3.1, 1.3.2 we obtain definitions of T-periodic, absolutely T-periodic, periodic and absolutely periodic billiard trajectories and points. As before, we shall denote the sets of such points in $S^*\mathring{M}$ by Π_T, Π_T^a, Π, and Π^a, respectively. Lemmas 1.3.4, 1.3.5 and Corollary 1.3.6 remain valid in this case. We shall retain for geodesic billiards Definition 1.3.7 describing the nonperiodicity condition.

Let us now introduce the important notions of convexity and concavity.

DEFINITION 1.3.12. Denote

$$\mathbf{k}(x',\xi') := \{h_{\xi_n}, h\}|_{x=(x',0),\xi=(\xi',\xi_n^{\mathrm{st}}(x',\xi'))}\,, \tag{1.3.20}$$

$(x',\xi') \in T'\partial M$. Here $\{\cdot,\cdot\}$ are the Poisson brackets (see our list of notation for the proper sign), and $\xi_n^{\mathrm{st}}(x',\xi')$ is the real ξ_n-root of the equation

$$h_{\xi_n}(x',0,\xi',\xi_n) = 0\,, \tag{1.3.21}$$

see also Fig. 2. We shall call $\mathbf{k}$ the *Hamiltonian curvature* of ∂M.

Note that formulae (1.3.20), (1.3.21) have a simple interpretation: $\mathbf{k} = -\ddot{x}_n^*|_{t=0}$, where $x^*(t) \subset \widehat{M}$ is a ray emitted from the point $x^*(0) = (x',0) \in \partial M$ in the direction $\dot{x}^*(0) = h_\xi(x',0,\xi',\xi_n^{\mathrm{st}}(x',\xi'))$ tangent to ∂M. Here we deliberately consider the ray on the extended manifold $\widehat{M}$ (see subsection 1.1.2) because this ray may leave M at arbitrarily small $t \neq 0$. Obviously, under the assumption that the coordinate ξ_n is fixed (see subsection 1.1.2) $\mathbf{k}$ is a well defined function on $T'\partial M$.

DEFINITION 1.3.13. We shall say that the manifold M is *convex* if $\mathbf{k}(x',\xi') \geq 0$, $\forall (x',\xi') \in T'\partial M$, and *strongly convex* if $\mathbf{k}(x',\xi') > 0$, $\forall (x',\xi') \in T'\partial M$.

DEFINITION 1.3.14. We shall say that the manifold M is *concave* if $\mathbf{k}(x',\xi') \leq 0$, $\forall (x',\xi') \in T'\partial M$, and *strongly concave* if $\mathbf{k}(x',\xi') < 0$, $\forall (x',\xi') \in T'\partial M$.

The notion of convexity is illustrated by the following

EXAMPLE 1.3.15 (EUCLIDEAN CASE). Let $M \subset \mathbb{R}^n$ and $h = |\eta|$; in this example we denote by y Cartesian coordinates in $\mathbb{R}^n$, and by η their corresponding duals. Definition 1.3.13 in this case is equivalent to the traditional definition of convexity: the set M is called convex if for any two points $y, z \in M$ the entire segment of the straight line connecting y and z belongs to M. Moreover, the quantity $\mathbf{k}$ in this case has a clear geometric meaning. Let us introduce (curvilinear) local coordinates x such that x_n is the Euclidean distance from the point to ∂M, with positive sign when the point is in $\mathring{M}$ and negative when the point is in $\mathbb{R}^n \setminus M$. Then $\mathbf{k}(x', \xi')$ is the curvature (in the traditional sense) of the normal section of the surface ∂M at the point $y = y(x', 0)$ in the direction $\eta = \eta(x', 0, \xi', \xi_n^{\mathrm{st}}(x', \xi'))$.

The Euclidean case is ill-suited for illustrating the notion of concavity because concave manifolds in this case are noncompact. Therefore we shall consider the following less trivial

EXAMPLE 1.3.16. Let M be a spherical cap, that is, part of a unit n-dimensional sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ cut off by a hyperplane in $\mathbb{R}^{n+1}$:

$$M = \{y \in \mathbb{R}^{n+1} : |y| = 1,\ y_{n+1} \geq c\},$$

where c is some constant, $|c| < 1$, and y are Cartesian coordinates in $\mathbb{R}^{n+1}$. The metric on M is assumed to be the standard one induced by the Euclidean metric in $\mathbb{R}^{n+1}$. In this case the manifold M is convex if it is less than or equal to a hemisphere ($c \geq 0$), strongly convex if it is strictly less than a hemisphere ($c > 0$), concave if it is greater than or equal to a hemisphere ($c \leq 0$), and strongly concave if it is strictly greater than a hemisphere ($c < 0$); see Fig. 5.

Figure 5. Spherical caps: a) convex, b) concave.

LEMMA 1.3.17. *If the manifold M is strongly convex, then there are no grazing billiard trajectories (i.e., $\mathcal{P}^{\mathrm{g}} = \varnothing$). If the manifold M is strongly concave or strongly convex, then there are no dead-end billiard trajectories (i.e., $\mathcal{P}^{\mathrm{d}} = \varnothing$).*

It is easy to see that a somewhat weaker version of Remark 1.3.3 holds for geodesic billiards.

REMARK 1.3.18. In the analytic case, if a trajectory is absolutely T-periodic then all sufficiently close billiard trajectories are T-periodic with the same period T.

Unfortunately, in the general case it is impossible to extend analytically the property of absolute periodicity to all points from $S^*\mathring{M}$, because analyticity may be lost on $\mathcal{P}^{\mathrm{g}}$. A lucky exception is the case of a convex manifold, see Lemma 1.3.17. For a convex manifold, Remark 1.3.3 and Lemma 1.3.8 still hold. Moreover, we have the following important

LEMMA 1.3.19. *Let us consider the analytic case and let the manifold* M *be convex,* $\partial M \neq \varnothing$, *and* $\mathbf{k}(x', \xi') \not\equiv 0$. *Then the nonperiodicity condition is fulfilled.*

Thus, Lemma 1.3.19 gives effective sufficient conditions guaranteeing nonperiodicity.

In particular, applying Lemma 1.3.19 to the Euclidean case (Example 1.3.15) we conclude that if M is convex (in the traditional sense) and ∂M is analytic then the nonperiodicity condition is fulfilled.

Applying Lemma 1.3.19 to Example 1.3.16 we conclude that if M is strictly less than a hemisphere ($c > 0$) then the nonperiodicity condition is fulfilled. On the other hand, it is easy to see directly that if $c \leq 0$ then the nonperiodicity condition is not fulfilled: when M is a hemisphere ($c = 0$) all geodesics are periodic after two reflections, and when M is strictly greater than a hemisphere ($c < 0$) the great circles which do not intersect ∂M form a rich set of periodic geodesics. So in this particular example strict convexity is necessary and sufficient for nonperiodicity.

It is worth noting that Lemma 1.3.19 does not have an analogue for manifolds without boundary. This leads us to the paradoxical conclusion that some manifolds with boundary are simpler objects than manifolds without boundary, at least as far as nonperiodicity is concerned.

Concluding this subsection, let us make the obvious observation that given a second order elliptic self-adjoint differential operator A acting on M one can always introduce a Riemannian metric based on the principal symbol of A and consider the corresponding geodesic billiard system.

3. Branching Hamiltonian billiards. Let us proceed at last to the most difficult case when M is a manifold with boundary and A is an operator of order $2m$, $m > 1$.

As in the preceding subsections, we start by introducing the trajectories

$$(x^*(t; y, \eta), \xi^*(t; y, \eta)), \qquad (y, \eta) \in T'\mathring{M},$$

of the Hamiltonian system (1.3.3). The trajectory may hit the boundary at some time $t = \tau \neq 0$ and our task is to define the reflection law. Unfortunately, in the general case we cannot formulate the reflection law in simple geometric terms because the Hamiltonian $h(x, \xi)$ does not induce a Riemannian metric on M (with the exception of the special case $A_{2m}(x, x) = \big(A_2(x, \xi)\big)^m$, where $A_2(x, \xi)$ is the principal symbol of some second order differential operator). We can, however, formulate the reflection law on the basis of the Hamiltonian formalism developed in subsection 2. Let us recall that this formalism requires $x^*(t)$ and the first $n-1$ components of $\xi^*(t)$ to be continuous at $t = \tau$ and the n-th component of $\xi^*(t)$ to satisfy the equality (1.3.13), $\xi_n^{\pm} = \xi_n^*(\tau \pm 0)$. For $\tau > 0$, ξ_n^- is given and ξ_n^+ must be determined from (1.3.13); for $\tau < 0$, ξ_n^+ is given and ξ_n^- must be determined from (1.3.13). As in subsection 2, we also require $\xi_n^{\pm}$ to satisfy the inequalities (1.3.17). In a weaker form (1.3.15) these inequalities are natural because the ray $x^*(t)$ has to stay on the manifold M. As in subsection 2, the exclusion of cases $h_{\xi_n}(\xi_n^-) = 0$, $h_{\xi_n}(\xi_n^-) = 0$ simplifies matters and (as we shall see later) leads to the loss of a set of initial points (y, η) of zero measure.

Thus, in the most general case we define the reflection law by the equality (1.3.13) with the additional conditions (1.3.17), plus the requirement that $x^*(t)$ and $\xi^{*\prime}(t)$ are continuous.

The application of this reflection law leads us to the discovery of a completely new phenomenon compared with the case of a second-order differential operator: for $m \geq 2$ the equation (1.3.13) may have several solutions ξ_n^+ ($\tau > 0$) or ξ_n^- ($\tau < 0$). This becomes clearer if we transform (1.3.13) to an equivalent form by raising both parts to the $2m$-th power:

$$A_{2m}(x', 0, \xi', \xi_n^-) = A_{2m}(x', 0, \xi', \xi_n^+)\,. \tag{1.3.22}$$

Multiplying (1.3.17) by $2mh^{2m-1}$, we obtain

$$A'_{2m}(x', 0, \xi', \xi_n^-) < 0\,, \qquad A'_{2m}(x', 0, \xi', \xi_n^+) > 0\,, \tag{1.3.23}$$

where the prime at A_{2m} denotes differentiation with respect to ξ_n. At this stage we assume, of course, that

$$\operatorname{Im} \xi_n^{\pm} = 0\,; \tag{1.3.24}$$

nonreal roots will be dealt with later, in subsection 2.6.4.

With respect to the unknown quantity ξ_n^+ (or ξ_n^-) (1.3.22) is an algebraic equation of degree $2m$ with real coefficients. Obviously, this equation has at most m ξ_n^+-roots (or ξ_n^--roots) satisfying the conditions (1.3.23), (1.3.24) if ξ_n^- (or ξ_n^+) is given. Moreover, one can easily construct an example when the equation (1.3.22) has exactly m ξ_n^+-roots and m ξ_n^--roots satisfying the conditions (1.3.23), (1.3.24). Thus, the case $m = 1$ considered in subsection 2 was a very special one: there we had a unique ξ_n^+ for each given ξ_n^- and vise versa. Fig. 6 illustrates the nonuniqueness of the solution ξ_n^+ (here ξ_n^- is given).

Figure 6. Graph of the function $h(\xi_n)$ in the case $m \geq 2$.

There is no reason to exclude any of the possible ways of continuation of the trajectory $(x^*(t), \xi^*(t))$ at each reflection. We shall retain the notation $(x^*(t), \xi^*(t))$ for each of the possible continuations (in positive or negative time) of the initial trajectory obtained by consecutive transversal reflections, and in line with Definition 1.3.10 we shall call $(x^*(t), \xi^*(t))$ a billiard trajectory. Here transversality means that both of the inequalities (1.3.17) must be fulfilled; in the case $m \geq 2$ these inequalities may not follow from one another as in the case $m = 1$. Considering billiard trajectories of finite length and with finitely many legs we shall always assume that $x^*(t; y, \eta)$ and $\xi^*(t; y, \eta)$ depend smoothly on $(t; y, \eta)$ at the points $(t; y, \eta)$ such that $x^*(t; y, \eta) \notin \partial M$ (see also end of this subsection where we introduce the notion of type of a billiard trajectory).

If for some incident trajectory there exist several reflected ones, then we shall say that the trajectory *branches*, see Fig. 7. Note that if we have branching the set of trajectories originating from a fixed point (y, η) may be uncountable.

Figure 7. Branching of trajectories.

REMARK 1.3.20. Branching of trajectories occurs also in transmission problems, in which several manifolds are connected by common stretches of boundary with appropriate boundary conditions, see [Sa1], [Sa3].

We shall call a billiard system with a Hamiltonian of general form (that is, not necessarily the square root of a quadratic polynomial) a *branching Hamiltonian billiard system*.

Let us now examine the factors that may prevent the extension of the billiard trajectory $(x^*(t;y,\eta),\xi^*(t;y,\eta))$ to all $t\in\mathbb{R}$.

Let $\mathcal{P}^{\mathrm{g}}$ be the set of points $(y,\widetilde{\eta})\in S^*\mathring{M}$ such that at least one trajectory originating from $(y,\widetilde{\eta})$ after a finite number of transversal reflections (in positive or negative time) hits a point $(x',0,\xi',\xi_n)\in T'M|_{\partial M}$ at which the equation $h(x',0,\xi',\zeta)=1$ has a multiple real ζ-root. (It may well be that $\zeta=\xi_n$ is itself a multiple root.)

Let $\mathcal{P}^{\mathrm{d}}$ be the set of points $(y,\widetilde{\eta})\in S^*\mathring{M}$ such that at least one trajectory originating from $(y,\widetilde{\eta})$ experiences an infinite number of transversal reflections in a finite (positive or negative) time. As before we call such trajectories dead-end.

If $(y,\widetilde{\eta})\notin\mathcal{P}^{\mathrm{g}}\cup\mathcal{P}^{\mathrm{d}}$ then all trajectories originating from $(y,\widetilde{\eta})$ can be extended to all $t\in\mathbb{R}$.

LEMMA 1.3.21. *For a branching Hamiltonian billiard system*

$$\operatorname{meas}\mathcal{P}^{\mathrm{g}}\;=\;0\,. \tag{1.3.25}$$

In the general case we cannot guarantee the fulfilment of equality (1.3.19). Moreover, there exists an effective (even analytic) example [SaVa1] of a branching Hamiltonian billiard system with $\operatorname{meas}\mathcal{P}^{\mathrm{d}}\neq 0$. Therefore we will have to impose (1.3.19) as a condition.

DEFINITION 1.3.22. We shall say that the *nonblocking condition* is fulfilled if $\operatorname{meas}\mathcal{P}^{\mathrm{d}}=0$.

The word *nonblocking* is used here to exhibit the fact that under this condition almost all (in terms of the measure of their starting points) trajectories can be extended to all $t\in\mathbb{R}$.

When the nonblocking condition is fulfilled, the set $\mathcal{P}^{\mathrm{g}}\cup\mathcal{P}^{\mathrm{d}}$ not only has zero measure, but it is also a set of first category in the sense of Baire.

Let us define T-periodic, absolutely T-periodic, periodic and absolutely periodic billiard trajectories in the same manner as in subsections 1 and 2 (Definitions 1.3.1, 1.3.2 plus the phrase "after a finite number of transversal reflections"). As before, we shall denote the sets of starting points of such trajectories in $S^*\mathring{M}$ by Π_T, Π_T^a, Π, and Π^a, respectively. However, we shall refrain from calling the points themselves periodic because some trajectories originating from these points may not be periodic.

The following results are analogues of Lemmas 1.3.4, 1.3.5 and Corollary 1.3.6 for branching Hamiltonian billiards.

LEMMA 1.3.23. *If the nonblocking condition is fulfilled, then the sets* Π *and* $\bigcup_{0<T\le T_+} \Pi_T$, $\forall T_+ > 0$, *are measurable.*

LEMMA 1.3.24. *If the nonblocking condition is fulfilled, then (1.3.7) and (1.3.8) hold.*

In accordance with Definition 1.3.7, we say that the nonperiodicity condition is fulfilled if

$$\operatorname{meas} \Pi = 0 . \tag{1.3.26}$$

It follows from Lemma 1.3.24 that if the nonblocking condition is fulfilled, then (1.3.26) is equivalent to

$$\operatorname{meas} \Pi^a = 0 , \tag{1.3.27}$$

which is very useful in obtaining effective sufficient conditions for nonperiodicity.

Further on it will be convenient for us to avoid considering reflections at negative times. Since our Hamiltonian (1.1.14) is an even function of ξ we have

$$(x^*(-t;y,\eta),\xi^*(-t;y,\eta)) = (x^*(t;y,-\eta),-\xi^*(t;y,-\eta)) , \tag{1.3.28}$$

so all the results for negative t can be deduced from those for positive t. We shall use this observation throughout the book. Note that if A is a pseudodifferential operator formula (1.3.28) does not necessarily hold, so in this case one has to consider negative as well as positive t.

In the process of the effective construction of the wave group (Chapter 3) we will often be forced to restrict ourselves to some finite interval on the time axis and to exclude some awkward reflections. We introduce the following

DEFINITION 1.3.25. Let T_+ be a finite positive number. We shall say that the trajectory

$$(x^*(t;y,\eta),\xi^*(t;y,\eta)) , \qquad 0 \le t \le T_+ , \quad (y,\eta) \in T'\mathring{M} , \tag{1.3.29}$$

is *admissible* if it satisfies the following conditions.

(1) It experiences a finite number of reflections.

(2) At each moment of reflection $t = \tau \in (0, T_+]$ all the ζ-roots of the algebraic equation

$$A_{2m}(x^{*\prime}(\tau;y,\eta),0,\xi^{*\prime}(\tau;y,\eta),\zeta) = A_{2m}(y,\eta) \tag{1.3.30}$$

are simple.

(3) At each moment of reflection $t = \tau \in (0, T_+]$ the number $\nu = A_{2m}(y,\eta)$ is not an eigenvalue of the auxiliary one-dimensional spectral problem (1.1.7), (1.1.8) with $(x',\xi') = (x^{*\prime}(\tau;y,\eta),\xi^{*\prime}(\tau;y,\eta)) \in T^*\partial M$.

The second condition of Definition 1.3.25 implies, in particular, that the reflections are transversal. However this condition has wider implications. When we say "the ζ-roots of the algebraic equation are simple" we mean *all the roots in the complex plane*, not only the real ones. The role of the nonreal roots will become clear in Sections 2.6 and 2.8 when we introduce the concept of a boundary layer oscillatory integral. In a sense, the nonreal ζ-roots correspond to Hamiltonian trajectories which leave the real space after reflection and become complex (in the analytic situation this statement has a precise meaning). Obviously, the equation (1.3.30) may have nonreal ζ-roots only if $m \ge 2$.

Note also that for a second order operator the third condition of Definition 1.3.25 is always fulfilled.

DEFINITION 1.3.26. We shall say that $(y,\eta) \in T'\mathring{M}$ is a T_+-admissible point if all the trajectories (1.3.29) originating from this point are T_+-admissible. We shall say that $(y,\eta) \in T'\mathring{M}$ is an admissible point if it is T_+-admissible for all $T_+ > 0$. A subset of $T'\mathring{M}$ is said to be T_+-admissible (admissible) if all its points are T_+-admissible (admissible).

Denote by O_{T_+} and O_∞ the sets of all T_+-admissible and admissible points in $T'\mathring{M}$, respectively.

LEMMA 1.3.27. *For any finite positive T_+ the set O_{T_+} is open.*

LEMMA 1.3.28. *If the nonblocking condition is fulfilled, then* $\operatorname{vol}(T'\mathring{M} \setminus O_\infty) = \operatorname{meas}(S^*\mathring{M} \setminus O_\infty) = 0$.

Further on we prefer to deal with T_+-admissible billiard trajectories. Most of our results and definitions can be extended to all well defined billiard trajectories, so it is a rather technical assumption. In view of Lemma 1.3.28 this will be sufficient for our purposes. Often we shall not distinguish between Π_{T_+} and $\Pi_{T_+} \cap O_{T_+}$, $\Pi^a_{T_+}$ and $\Pi^a_{T_+} \cap O_{T_+}$, Π and $\Pi \cap O_\infty$, Π^a and $\Pi^a \cap O_\infty$.

Consider a T_+-admissible billiard trajectory (1.3.29) experiencing $\mathbf{r}$ reflections on the time interval $(0, T_+)$. We can associate with this trajectory its *type*, which is the multiindex $\mathfrak{m} = \mathfrak{m}_1|\mathfrak{m}_2|\dots|\mathfrak{m}_i|\dots|\mathfrak{m}_\mathbf{r}$ defined as follows. If $\tau_i \in (0, T_+)$ is the moment of the ith reflection, then $\xi^*_n(\tau_i + 0; y, \eta)$ is the $2\mathfrak{m}_i$th (in order of growth) real ζ-root of the equation (1.3.30). The notion of a billiard trajectory being T_+-admissible ensures that all the real ζ-roots of the equation (1.3.30) are simple; consequently, for any real root with even sequential number we have $h_{\xi_n}(x^{*\prime}(\tau_i; y, \eta), 0, \xi^{*\prime}(\tau_i; y, \eta), \zeta) > 0$, i.e., it really corresponds to a reflected trajectory.

LEMMA 1.3.29. *The number of types of billiard trajectories originating from a T_+-admissible point is finite.*

Let $(x^*(t; y_0, \eta_0), \xi^*(t; y_0, \eta_0))$ be a T_+-admissible billiard trajectory of type $\mathfrak{m}$ such that T_+ is not a moment of reflection. Clearly, if we fix the type $\mathfrak{m}$ then $(x^*(t; y, \eta), \xi^*(t; y, \eta))$ smoothly depends on $(t; y, \eta)$ provided (y, η) is sufficiently close to (y_0, η_0) and t is not a moment of reflection.

The branching of trajectories does not allow us to introduce the group of shifts Φ^t along billiard trajectories (cf. subsection 2). However, shifts along admissible billiard trajectories of fixed type preserve the symplectic differential 1-form and 2-form, as well as the symplectic volume (see Section 2.3).

4. Simple reflection. An important class of Hamiltonian billiard systems is the class of systems satisfying the simple reflection condition.

DEFINITION 1.3.30. We shall say that the *simple reflection condition* is fulfilled if for any $(x', \xi') \in T'\partial M$ the Hamiltonian $h(x', 0, \xi', \xi_n)$ has only one local (and hence global) minimum as a function of $\xi_n \in \mathbb{R}$.

In other words, the simple reflection condition is fulfilled if and only if there is no branching.

LEMMA 1.3.31 (GENERALIZATION OF THE SECOND PART OF LEMMA 1.3.11). *If the Hamiltonian billiard system satisfies the simple reflection condition, then the nonblocking condition is fulfilled, i.e.,* $\operatorname{meas} \mathcal{P}^{\mathrm{d}} = 0$.

For a Hamiltonian billiard system satisfying the simple reflection condition we can introduce the group of shifts Φ^t along billiard trajectories. This billiard flow has properties similar to those of the geodesic billiard flow.

DEFINITION 1.3.32. We shall say that the *strong simple reflection condition* is fulfilled if for any $(x', \xi') \in T'\partial M$ the equation $h_{\xi_n}(x', 0, \xi', \xi_n) = 0$ has only one real ξ_n-root and this root is simple ($h_{\xi_n \xi_n} \neq 0$).

REMARK 1.3.33. Geodesic billiards obviously satisfy the strong simple reflection condition.

Hamiltonian billiards satisfying the strong simple reflection condition are a subclass of Hamiltonian billiards satisfying the simple reflection condition. Billiards from this subclass are of special interest because **all** the facts described in subsection 2 (devoted to geodesic billiards) remain true for them as well. In particular, Lemma 1.3.19 remains true; due to its importance we shall state it again.

LEMMA 1.3.34 (GENERALIZATION OF LEMMA 1.3.19). *Let us consider the analytic case and let the strong simple reflection condition be satisfied, the manifold* M *be convex (see Definitions 1.3.12, 1.3.13),* $\partial M \neq \varnothing$, *and* $\mathbf{k}(x', \xi') \not\equiv 0$. *Then the nonperiodicity condition is fulfilled.*

Note that under the assumptions of Lemma 1.3.34 the nonblocking condition is fulfilled as well, by virtue of Lemma 1.3.31.

In the case of a manifold without boundary there are no reflections, but in order to simplify our subsequent statements we shall assume throughout the book that a billiard system without reflections satisfies the strong simple reflection condition.

Concluding this subsection, let us make a remark on the relationship between the concepts of nonperiodicity and ergodicity. As pointed out earlier, a Hamiltonian billiard system satisfying the simple reflection condition generates the group of shifts Φ^t which preserves the symplectic volume; such groups are called *flows*. A flow is called *ergodic* if any set invariant with respect to this flow has zero or full measure [CorFomSin]. If the billiard flow is ergodic, then the nonperiodicity condition is fulfilled. The converse is, in general, not true. An example is the case of Euclidean billiards in a closed domain $M \subset \mathbb{R}^2$ bounded by an ellipse. The fact that such a billiard system is not ergodic is well known [CorFomSin, Sect. 6.3], whereas its nonperiodicity follows from Lemma 1.3.19 or Lemma 1.3.34. Thus, nonperiodicity is a weaker (in fact, considerably weaker) requirement than ergodicity. This explains why constructing effective sufficient conditions for nonperiodicity is easier and more rewarding than constructing effective sufficient conditions for ergodicity.

5. Euclidean billiards. Let us finally look at the most basic billiards, that is, Euclidean billiards. Euclidean billiards are a special case of geodesic billiards. In this case the manifold M is a region in $\mathbb{R}^n$, the metric is Euclidean, the rays are straight lines, and reflections from the boundary satisfy the usual "angle of incidence equals angle of reflection" law.

For Euclidean billiards the nonblocking condition is always fulfilled due to simple reflection. But nonperiodicity has to be checked.

We state the following

CONJECTURE 1.3.35. *In the case of a Euclidean billiard system there are no absolutely periodic billiard trajectories.*

Note that this conjecture, if proven, would imply the nonperiodicity of Euclidean billiard systems (see Lemma 1.3.6).

The arguments in favour of this conjecture are the following.

If one assumes the existence of an absolutely periodic billiard trajectory this leads to certain algebraic equations involving the Taylor coefficients which describe the shape of the boundary near the points of reflection. There are infinitely many algebraic equations and they involve infinitely many Taylor coefficients, which makes analysis difficult. However, one can truncate this infinite system of equations and look at the resulting finite subsystem of l equations. One can check that if l is sufficiently large, then the number of unknowns (Taylor coefficients) is less than l. Therefore it is highly unlikely that this system has a solution.

There is also a physical argument in favour of conjecture 1.3.35. If the conjecture were not true one could construct an ideal optical system with a finite number of mirrors. Here "ideal" means focusing up to infinite order, i.e., without any aberrations.

Conjecture 1.3.35 is purely geometrical, and is remarkably simple and natural. It contains no direct reference to partial differential equations, spectral theory, symplectic geometry, measure theory, etc. The authors, however, are unaware of a mathematical proof of this conjecture.

We know Conjecture 1.3.35 to be true only for special classes of shapes of M. In particular, we know it to be true when M is convex (in the usual sense) and ∂M analytic; see proof of Lemma 1.3.34 given in Appendix D. If one allows piecewise smooth boundaries [Va6], we can state two other sets of conditions under which the conjecture is true.

(1) Each smooth component of ∂M has nonpositive normal curvature (say, a polyhedron satisfies this condition). In this case a divergent beam of rays becomes more and more divergent after each reflection and can not focus. This is a standard argument from ergodic theory.
(2) A very special set of conditions [Wo].

1.4. Hamiltonian billiards II: reflection matrix

As we shall discover later (Sections 3.3, 3.4), the physical meaning of a branching Hamiltonian billiard system is that it describes the propagation of waves governed by the equation $D_t^{2m}u = Au$ (analogue of the wave equation). Billiard trajectories trace the movement of these waves, and our billiard trajectories play the role of a "skeleton" upon which we will consequently build the wave group $\exp(-it\mathcal{A}^{1/(2m)})$.

Tracing the path of a wave is only the first (and simplest) step. The next step is to describe the partition of energy between different branches of reflected waves. The notion of the reflection matrix introduced in this section serves this purpose. The reflection matrix is not used directly in the formulation of the main result of Section 1.6. However, it is used later on in Section 1.6 when we perform the effective calculation of the second asymptotic coefficient.

Under the simple reflection condition energy is not redistributed at reflections (there is no branching), and the reflection matrix is just a complex number of the form $e^{i\mu}$, $\mu \in \mathbb{R}$. In this case it is natural to study the *phase shift* which the wave

gains in the process of its propagation. The phase shift comes from three sources:

(1) phase shift induced by reflections from the boundary (the sum of arguments μ of the corresponding reflection matrices);
(2) phase shift induced by the passage of the trajectory through caustics;
(3) phase shift induced by the subprincipal symbol of the differential operator A.

This section deals with the first source and Section 1.5 deals with the second one. Dealing with the third source will not require a special section because the corresponding phase shift will be given by a simple integral over the billiard trajectory (see subsection 1.7.2). The notion of the phase shift is necessary for the formulation of the results of Sections 1.7 and 1.8.

Let us fix an arbitrary point $(x', \xi') \in T^*\partial M$ and consider the one-dimensional spectral problem (1.1.7), (1.1.8) on the half-line $x_n \in \mathbb{R}_+$. For the time being let us omit the parameters (x', ξ') and rewrite (1.1.7), (1.1.8) as

$$A(D_{x_n})v = \nu v\,, \tag{1.4.1}$$

$$\Big(B^{(j)}(D_{x_n})v\Big)\Big|_{x_n=0} = 0\,, \qquad j = 1, 2, \dots, m\,, \tag{1.4.2}$$

where $A(D_{x_n}) = A_{2m}(x', 0, \xi', D_{x_n})$, $B^{(j)}(D_{x_n}) = B^{(j)}_{m_j}(x', \xi', D_{x_n})$. Obviously, the spectral problem (1.4.1), (1.4.2) is formally self-adjoint with respect to the inner product

$$(v, w)_+ = \int_0^{+\infty} v(x_n)\,\overline{w(x_n)}\,dx_n\,, \tag{1.4.3}$$

and one can associate with (1.4.1), (1.4.2) a self-adjoint operator $\mathbf{A}^+$ in $L_2(\mathbb{R}_+)$.

The equation (1.4.1) is an ordinary differential equation of order $2m$ with constant coefficients and it can be explicitly solved in exponential functions. This simple observation opens the way to the effective study of the spectral problem (1.4.1), (1.4.2). In fact, all the constructions described below (as well as in subsection 1.6.3 and Appendix A) are of purely algebraic nature.

In this section we give a list of basic properties of the spectral problem (1.4.1), (1.4.2), referring the reader to Appendix A for details.

The problem (1.4.1), (1.4.2) has a finite number of eigenvalues. Moreover, this number is uniformly bounded over all problems of this type with fixed order $2m$, see Proposition A.1.13.

The problem (1.4.1), (1.4.2) has no singular continuous spectrum (see Remark A.2.8). As the number of eigenvalues is finite, the continuous spectrum coincides with the essential. Thus, the spectrum of (1.4.1), (1.4.2) is the union of the (absolutely) continuous spectrum and of the set of eigenvalues. Note that some eigenvalues may be embedded in the continuous spectrum.

The continuous spectrum of the problem (1.4.1), (1.4.2) is the interval $[\nu_1^{\mathrm{st}}, +\infty)$, where $\nu_1^{\mathrm{st}} = \min\limits_{\xi_n \in \mathbb{R}} A(\xi_n)$ (see subsections A.1.4 and A.1.5).

We shall call the number ν^{st} a *threshold* (or a *stationary value* of the symbol) if the equation $A(\xi_n) = \nu^{\mathrm{st}}$ has a multiple real ξ_n-root. Let us enumerate the

thresholds: $\nu_1^{\rm st} < \nu_2^{\rm st} < \cdots < \nu_s^{\rm st}$. Clearly, $1 \le s \le 2m-1$. After the removal of thresholds the continuous spectrum separates into zones

$$(\nu_1^{\rm st}, \nu_2^{\rm st}),\ (\nu_2^{\rm st}, \nu_3^{\rm st}), \dots,\ (\nu_{s-1}^{\rm st}, \nu_s^{\rm st}),\ (\nu_s^{\rm st}, +\infty). \tag{1.4.4}$$

Further on we use the (scalar) variable ζ instead of ξ_n in order to simplify notation.

Assume that ν belongs to one of the zones (1.4.4), and consider the algebraic equation

$$A(\zeta) = \nu. \tag{1.4.5}$$

By $2q$ we shall denote the number of real ζ-roots of this equation. The number q is called the *multiplicity of the continuous spectrum* at the point ν. Obviously, this number is the same for all ν from a given zone (1.4.4), and $1 \le q \le m$.

It will be convenient for us to impose temporarily the following technical conditions (which are related to the admissibility condition).

Condition 1.4.1. Our ν is such that the complex ζ-roots of (1.4.5) are simple.

Condition 1.4.2. Our ν is not an eigenvalue of (1.4.1), (1.4.2).

Let us denote the real ζ-roots of (1.4.5) by $\zeta^{\mp}(\nu)$, where the superscript $\mp$ indicates the sign of the derivative $A'(\zeta^{\mp}(\nu))$, $A' \equiv dA/d\zeta$; note that this derivative is non-zero because our ν is not a threshold. Let us enumerate the real roots $\zeta^{\mp}(\nu)$ in order of growth:

$$\zeta_1^-(\nu) < \zeta_1^+(\nu) < \zeta_2^-(\nu) < \zeta_2^+(\nu) < \dots \zeta_q^-(\nu) < \zeta_q^+(\nu). \tag{1.4.6}$$

Formula (1.4.6) is illustrated by Fig. 8 (here $q = 2$). Note that apart from different notation Fig. 8 is essentially the same as Fig. 6 from subsection 1.3.3.

Figure 8. Real roots of the equation (1.4.5).

Let us denote by $\zeta_l^{\mp}(\nu)$, $l = q+1, q+2, \dots, m$, the complex roots of (1.4.5) with negative and positive imaginary part, respectively. Note that under the Condition 1.4.1 we have $A'(\zeta^{\mp}(\nu)) \ne 0$, $l = q+1, q+2, \dots, m$.

Let us call nontrivial bounded solutions of (1.4.1), (1.4.2) *generalized eigenfunctions* (or *eigenfunctions of the continuous spectrum*) corresponding to our ν. As we have excluded eigenvalues (Condition 1.4.2) generalized eigenfunctions defined in this way are not in $L_2(\mathbb{R}_+)$. With account of Condition 1.4.1 we shall search for generalized eigenfunctions in the form

$$v(x_n) \;=\; \sum_{l=1}^{q} \frac{a_l^-\, e^{ix_n\zeta_l^-(\nu)}}{\sqrt{-2\pi A'(\zeta_l^-(\nu))}} \;+\; \sum_{l=1}^{m} \frac{a_l^+\, e^{ix_n\zeta_l^+(\nu)}}{\sqrt{2\pi A'(\zeta_l^+(\nu))}}\,, \tag{1.4.7}$$

where the a_l^-, $l = 1, 2, \dots, q$, and a_l^+, $l = 1, 2, \dots, m$, are some complex constants. In the right-hand side of (1.4.7) all the square roots in the first sum and the first q square roots in the second sum are chosen to be positive, whereas the last $m - q$ square roots in the second sum can be chosen arbitrarily. The normalizing factors

$$\frac{1}{\sqrt{\mp 2\pi A'(\zeta_l^{\mp}(\nu))}}$$

in (1.4.7) are introduced for the sake of convenience, in order to make the reflection matrix defined below a unitary one. Another reason for introducing these normalizing factors is that they will naturally appear when we will start dealing with our global invariant oscillatory integrals, see Lemmas 2.9.10, 2.9.11.

Proposition A.2.1 establishes that the linear space of generalized eigenfunctions is q-dimensional (this justifies our notion of multiplicity defined above) and that each generalized eigenfunction (1.4.7) is uniquely defined by the set of coefficients a_l^-, $l = 1, 2, \dots, q$, or by the set of coefficients a_l^+, $l = 1, 2, \dots, q$. Denote by a^- and a^+ the columns of coefficients a_l^-, $l = 1, 2, \dots, q$, and a_l^+, $l = 1, 2, \dots, q$, respectively. As a generalized eigenfunction is uniquely determined by either of the two columns a^- or a^+, we have a linear relation

$$a^+ = R(\nu)\, a^- , \tag{1.4.8}$$

where $R(\nu)$ is an invertible $q \times q$ matrix. The matrix $R(\nu)$ can be effectively constructed by substituting (1.4.7) into the boundary conditions (1.4.2) and solving the resulting system of linear algebraic equations with respect to the a_l^+.

The matrix $R(\nu)$ is called the *reflection matrix*. By Proposition A.2.1 the reflection matrix is unitary.

Obviously, the matrix $R(\nu)$ is real-analytic as a function of ν in each of the zones (1.4.4) apart, maybe, from a finite number of points at which either of the Conditions 1.4.1 or 1.4.2 fails. By Proposition A.2.1 $R(\nu)$ admits an analytic extension to such points. Thus $R(\nu)$ is well defined and real-analytic in each of the zones (1.4.4).

The thresholds are, generally speaking, branching points for the elements of $R(\nu)$. In addition, passage through a threshold normally results in a change of the size of $R(\nu)$; this corresponds to the change of the multiplicity of the continuous spectrum.

Note, however, that the elements of the reflection matrix always have one-sided limits at thresholds. This follows from the fact that these elements are bounded (recall that the matrix is unitary!) and that the branching is of root type.

Definition 1.4.3. We shall say that the one-dimensional problem (1.4.1), (1.4.2) satisfies the *simple reflection condition* if the continuous spectrum has multiplicity one in all the zones (1.4.4).

Of course, in the case of simple reflection the reflection matrix $R(\nu)$ is of size 1×1 (complex number) in all the zones (1.4.4), and in view of (1.4.8) $R(\nu) = a_1^+/a_1^-$. Naturally, $|R(\nu)| = 1$. The quantity $\arg R(\nu)$ is called the *phase shift generated by the reflection*. This phase shift is defined modulo 2π.

DEFINITION 1.4.4. We call the threshold $\nu^{\rm st}$ *normal* if the algebraic equation $A(\zeta) = \nu^{\rm st}$ has only one multiple real root $\zeta = \zeta^{\rm st}$ and $A''(\zeta^{\rm st}) \neq 0$.

Note that the situation of a normal threshold can be viewed as a generic one, in the sense that all the thresholds can be made normal by a small self-adjoint perturbation of the coefficients of our spectral problem (1.4.1), (1.4.2).

DEFINITION 1.4.5. We shall say that the one-dimensional problem (1.4.1), (1.4.2) satisfies the *strong simple reflection condition* if the problem has only one threshold and this threshold is normal.

Definitions 1.4.3, 1.4.5 given above match well with Definitions 1.3.30, 1.3.32. Namely, the one-dimensional problem (1.4.1), (1.4.2) satisfies the simple reflection condition (strong simple reflection condition) for all $(x', \xi') \in T'\partial M$ if and only if the Hamiltonian billiard system satisfies the simple reflection condition (strong simple reflection condition).

From now on it will be convenient for us to indicate the dependence of all our quantities on the parameters $(x', \xi') \in T^*\partial M$.

Let us examine how the reflection matrix $R(\nu; x', \xi')$ behaves under changes of local coordinates. Let $x = x(\widetilde{x})$ be a change of local coordinates on M such that $x_n \equiv \widetilde{x}_n$ (see subsection 1.1.2), and let

$$f'(\widetilde{x}') \,:=\, \frac{\partial x'}{\partial \widetilde{x}_n}\bigg|_{\widetilde{x}_n = 0} .$$

Our original change of local coordinates on M generates a change of local coordinates $x' = x'(\widetilde{x}') := x'(\widetilde{x}', 0)$ on ∂M, which leads to a change of dual coordinates

$$\xi' \,=\, \xi'(\widetilde{x}', \widetilde{\xi}') \,:=\, \sum_{k=1}^{n-1} \widetilde{\xi}_k \, \frac{\partial \widetilde{x}_k}{\partial x'}\bigg|_{x' = x'(\widetilde{x}')}$$

on the fibres of $T'\partial M$.

Suppose now that $v(x_n; x', \xi')$ is a solution of the problem (1.1.7), (1.1.8). Set

$$\widetilde{v}(\widetilde{x}_n; \widetilde{x}', \widetilde{\xi}') \,:=\, e^{i\widetilde{x}_n \langle f'(\widetilde{x}'), \xi'(\widetilde{x}', \widetilde{\xi}')\rangle} \, v(\widetilde{x}_n; x'(\widetilde{x}'), \xi'(\widetilde{x}', \widetilde{\xi}')) \,. \tag{1.4.9}$$

Direct substitution demonstrates that this function is a solution of the problem

$$A_{2m}(\widetilde{x}', 0, \widetilde{\xi}', D_{\widetilde{x}_n})\widetilde{v} \,=\, \nu\widetilde{v} \,,$$

$$\Big(B^{(j)}_{m_j}(\widetilde{x}', \widetilde{\xi}', D_{\widetilde{x}_n})\widetilde{v}\Big)\Big|_{\widetilde{x}_n = 0} \,=\, 0 \qquad j = 1, 2, \ldots, m \,.$$

Formula (1.4.9) implies that under changes of local coordinates x the ordering of the real roots (1.4.6) is preserved and, moreover, the reflection matrix $R(\nu; x', \xi')$ behaves as a function on $T'\partial M$. In the case of simple reflection the same applies to the phase shift.

It remains only to relate the quantities defined above to billiard trajectories. Let $(x^*(t; y, \eta), \xi^*(t; y, \eta))$ be a trajectory which experiences a reflection at $t = t^*(y, \eta)$. Then

$$\nu = A_{2m}(y, \eta) \,, \qquad (x', \xi') = (x^{*\prime}(t^*(y, \eta); y, \eta), \xi^{*\prime}(t^*(y, \eta); y, \eta)) \in T^*\partial M$$

are the values of parameters corresponding to this particular reflection, and these are the values to be substituted into all the functions defined in this section. In particular, we get

$$\xi_n^*(t^*(y,\eta)-0;y,\eta)=\zeta_k^-(A_{2m}(y,\eta);x^{*\prime}(t^*(y,\eta);y,\eta),\xi^{*\prime}(t^*(y,\eta);y,\eta))\,,$$

$$\xi_n^*(t^*(y,\eta)+0;y,\eta)=\zeta_l^+(A_{2m}(y,\eta);x^{*\prime}(t^*(y,\eta);y,\eta),\xi^{*\prime}(t^*(y,\eta);y,\eta))$$

for some k and l. Consequently

$$R_{lk}(A_{2m}(y,\eta);x^{*\prime}(t^*(y,\eta);y,\eta),\xi^{*\prime}(t^*(y,\eta);y,\eta))$$

is the element of the reflection matrix corresponding to our particular reflection; here l is the number of the row, and k is the number of the column.

In the case of simple reflection

$$\arg R(A_{2m}(y,\eta);x^{*\prime}(t^*(y,\eta);y,\eta),\xi^{*\prime}(t^*(y,\eta);y,\eta))$$

is the phase shift generated by our particular reflection. Recall that in the case of simple reflection

$$|R(A_{2m}(y,\eta);x^{*\prime}(t^*(y,\eta);y,\eta),\xi^{*\prime}(t^*(y,\eta);y,\eta))|=1\,,$$

and consequently the 1×1 reflection matrix is uniquely determined by the phase shift. The phase shift itself is, of course, defined modulo 2π; we shall always choose it to be locally continuous in (y,η).

1.5. Hamiltonian billiards III: Maslov index

In this section we introduce an important geometric characteristic of a billiard trajectory which is called the *Maslov index*. We also state some simple results which justify the definitions and can be used for the calculation of the Maslov index. These results will be proved in Appendix D.

In geometrical optics the Maslov index, multiplied by $-\pi/2$, is interpreted as the phase shift generated by the passage of the trajectory through caustics. It should be mentioned that the words "Maslov index" are often used for absolutely different objects and, probably, in our case it would be more appropriate to call it the Morse index rather than the Maslov index. However, we will follow the tradition.

The notion of the Maslov index introduced in this section is necessary for the formulation of the results of Sections 1.7, 1.8, but it is not needed in Section 1.6.

Consider a T-admissible billiard trajectory

$$(1.5.1)_0\qquad \Gamma\ =\ \big(x^*(t;y_0,\eta_0)\,,\,\xi^*(t;y_0,\eta_0)\big)\qquad 0\le t\le T\,,$$

such that $x^*(T;y_0,\eta_0)\notin\partial M$. Let $O\in T'M$ be a sufficiently small conic neighbourhood of the point (y_0,η_0) and

$$(1.5.1)\qquad \big(x^*(t;y,\eta)\,,\,\xi^*(t;y,\eta)\big)\,,\qquad 0\le t\le T\,,\quad (y,\eta)\in O\,,$$

be the family of T-admissible billiard trajectories with the same type (see end of subsection 1.3.3). We denote by $0<t_1^*(y,\eta)<\dots<t_{\mathbf{r}}^*(y,\eta)<T$ the moments of reflection.

Let us introduce in local coordinates the matrices x^*_η and ξ^*_η with elements $(x^*_j)_{\eta_i}$ and $(\xi^*_j)_{\eta_i}$ respectively (j being the number of the row and i that of the column). Certainly, at the moments of reflection these matrices are not smooth in t, and at these points we have to consider the matrices $x^*_\eta(t^*_k(y,\eta)-0;y,\eta)$, $\xi^*_\eta(t^*_k(y,\eta)-0;y,\eta)$ and $x^*_\eta(t^*_k(y,\eta)+0;y,\eta)$, $\xi^*_\eta(t^*_k(y,\eta)+0;y,\eta)$.

Throughout the section we assume that

$$x^*_\eta(T;y_0,\eta_0) \;=\; 0\,. \tag{1.5.2}$$

Under change of coordinates x^*_η behaves as a vector in x and as a vector in y, i.e., in new coordinates $\widetilde{x}^*_{\widetilde{\eta}} = (\partial\widetilde{x}/\partial x)\cdot x^*_\eta \cdot (\partial\widetilde{y}/\partial y)^T$ where $\partial\widetilde{x}/\partial x$ and $\partial\widetilde{y}/\partial y$ are the Jacobi matrices. Therefore the condition (1.5.2) is invariant under changes of coordinates.

The matrix ξ^*_η does not behave as a tensor. We will discuss its properties in Section 2.3. In particular, we will prove (Lemma 2.3.2) that for any fixed point $(t_0;y_0,\eta_0)$ and for any coordinates y one can choose coordinates x in a neighbourhood of $x^*(t_0;y_0,\eta_0)$ such that

$$\det \xi^*_\eta(t;y,\eta) \;\neq\; 0 \tag{1.5.3}$$

for $(t;y,\eta)$ close to $(t_0;y_0,\eta_0)$.

1. First definition. We start with a definition which is closer to the standard definition of the Morse index.

Let us choose functions $t_j(y,\eta)\in C^\infty(O)$, $j=0,\dots,N$, and coordinate patches $\Omega_j\subset M_x$, $j=0,\dots,N-1$, in such a way that

(1) $0\equiv t_0(y,\eta)<t_1(y,\eta)<\ldots<t_{N-1}(y,\eta)<t_N(y,\eta)\equiv T$;
(2) for each $k=1,2,\dots,\mathbf{r}$ there exists a j_k such that $t_{j_k}(y,\eta)\equiv t^*_k(y,\eta)$;
(3) $x^*(t;y,\eta)\in\Omega_j$ for all $(t;y,\eta)\in\check{\mathfrak{O}}_j$, $j=0,1,\dots,N-1$;
(4) the local coordinates defined on Ω_j satisfy (1.5.3) for all $(t;y,\eta)\in\check{\mathfrak{O}}_j$, $j=0,\dots,N-1$ (here we may need to use coordinates in which $\partial M\neq\{x_n=0\}$, see Remark 2.3.3).

We use the notation

$$\check{\mathfrak{O}}_j \;:=\; \{(t;y,\eta)\,:\,(y,\eta)\in O\,,\;t\in[t_j(y,\eta),t_{j+1}(y,\eta)]\}\,, \tag{1.5.4}$$

$j=0,1,\dots,N-1$. Clearly, conditions (1), (2) imply that $N\geq\mathbf{r}+1$.

Let us introduce the matrix-functions $C_j=C_j(t;y,\eta):=(\xi^*_\eta)^T\cdot x^*_\eta$ defined on the sets (1.5.4). If $t=t_{j_k}(y,\eta)=t^*_k(y,\eta)$ is a moment of reflection then we set $C_{j_k-1}(t;y,\eta):=C_{j_k-1}(t-0;y,\eta)$ and $C_{j_k}(t;y,\eta):=C_{j_k}(t+0;y,\eta)$. Since the shift along billiard trajectories preserves the canonical 2-form $dx\wedge d\xi$ the matrices C_j are symmetric, i.e., $C_j=(x^*_\eta)^T\cdot\xi^*_\eta$ (see Section 2.3). Since the functions x^*_j are homogeneous in η of degree 0, by the Euler identity we have $x^*_\eta\,\eta\equiv 0$. Therefore $\operatorname{rank}x^*_\eta(t;y,\eta)\leq n-1$ and, consequently, $\operatorname{rank}C_j(t;y,\eta)\leq n-1$ for all j and $(t;y,\eta)\in\check{\mathfrak{O}}_j$.

Let $r^-_j(t;y,\eta)$ be the number of strictly negative eigenvalues of the matrix $C_j(t;y,\eta)$. By $r^-_j(t\pm0;y,\eta)$ we shall denote the one-sided limits of the functions r^-_j (whenever these limits exist).

DEFINITION 1.5.1. The integer number

$$\alpha_\Gamma \;=\; -\sum_{j=0}^{N-1} \big(r_j^-(t_{j+1}(y_0,\eta_0);y_0,\eta_0) \;-\; r_j^-(t_j(y_0,\eta_0);y_0,\eta_0) \big) \tag{1.5.5}$$

is called the Maslov index of the trajectory Γ.

PROPOSITION 1.5.2. *Let $\widetilde{x}$ be another coordinate system on Ω_j satisfying (1.5.3) and $\widetilde{r}_j^-(t;y,\eta)$ be the number of negative eigenvalues of the corresponding matrix $\widetilde{C}_j(t;y,\eta)$. Then*

$$r_j^-(t;y,\eta) \;-\; \widetilde{r}_j^-(t;y,\eta) \;=\; \text{const}\,,$$

i.e. this difference is independent of $(t;y,\eta)\in\check{\mathfrak{O}}_j$.

By Proposition 1.5.2 the difference

$$r_j^-(t_{j+1}(y_0,\eta_0);y_0,\eta_0) \;-\; r_j^-(t_j(y_0,\eta_0);y_0,\eta_0)$$

does not depend on the choice of the coordinates x. This implies that the definition of the Maslov index is independent of the choice of the t_j, the covering $\{\Omega_j\}$ and the coordinate systems on Ω_j. Besides, Proposition 1.5.2 shows that the jumps $r_j^-(t+0;y,\eta)-r_j^-(t;y,\eta)$ and $r_j^-(t;y,\eta)-r_j^-(t-0;y,\eta)$ are invariant objects. Up to the factor -1, the Maslov index α_Γ is equal to the sum of all these jumps along the billiard trajectory Γ provided the number of jumps is finite.

LEMMA 1.5.3. *For all $k=1,\dots,\mathbf{r}$*

$$\operatorname{rank} x_\eta^*(t_k^*(y,\eta)-0;y,\eta) \;=\; \operatorname{rank} x_\eta^*(t_k^*(y,\eta)+0;y,\eta)$$

(here we first take the limit in t and then evaluate the rank.)

Let $\mathcal{R}=\mathcal{R}(t;y,\eta):=\operatorname{rank} x_\eta^*(t;y,\eta)$. By Lemma 1.5.3 $\mathcal{R}(t;y,\eta)$ is well defined at the points of reflection, and we can consider it as a function on $[0,T]\times O$.

If $\mathcal{R}(t;y,\eta)$ is constant on some subset of (1.5.4) then $r_j^-(t;y,\eta)$ is also constant on this subset. This means that the parts of the trajectory where $\mathcal{R}$ does not change do not contribute to the Maslov index. Thus, the Maslov index α_Γ depends only on the behaviour of the trajectory Γ in the neighbourhoods of the points at which $\mathcal{R}$ has jumps.

If $\mathcal{R}(t;y,\eta)<n-1$ then the point $x^*(t;y,\eta)$ is said to be the *conjugate point* of the ray $x^*(t;y,\eta)$ and the number $n-1-\mathcal{R}(t;y,\eta)$ is called its multiplicity. In geometrical optics the set of conjugate points of the rays starting at a fixed point y is often called the *caustic set* (or just *caustic*). It is known that when a ray $x^*(t;y,\eta)$ passes through the caustic set the wave amplitude is multiplied by

$$\exp\big(i\,\pi\,(r_j^-(t+0;y,\eta)-r_j^-(t-0;y,\eta))/2\big)\,,$$

or, in other words, the wave acquires the phase shift

$$\pi\big(r_j^-(t+0;y,\eta) \;-\; r_j^-(t-0;y,\eta)\big)/2\,.$$

Thus, the quantity $\mathfrak{f}_c(t;y_0,\eta_0)=\alpha_\Gamma\pi/2$ is the total phase shift generated by the passage of the ray $x^*(t;y_0,\eta_0)$ through caustics.

REMARK 1.5.4. Since $C_0(0;y,\eta)\equiv 0$ and $C_{N-1}(T;y_0,\eta_0)=0$,

$$\alpha_\Gamma \;=\; -\sum_{j=1}^{N-1}\big(r^-_{j-1}(t_j(y_0,\eta_0);y_0,\eta_0)\;-\;r^-_j(t_j(y_0,\eta_0);y_0,\eta_0)\big)\,. \tag{1.5.6}$$

Let $\partial M=\varnothing$. Then by Proposition 1.5.2 for any $j=1,\dots,N-1$ the expression

$$\sigma_{j-1,j}\;:=\;r^-_{j-1}(t;y,\eta)\;-\;r^-_j(t;y,\eta)\,,\qquad \forall (t;y,\eta)\in\check{\mathfrak{O}}_{j-1}\cap\check{\mathfrak{O}}_j\,,$$

is an integer constant (note that our $\sigma_{j-1,j}$ differ by sign from Hörmander's [Hö2, p. 92] [DuiGuiHö, p. 36(92)]). Let us assume that all the trajectories (1.5.1) are T-periodic and $\Omega_N=\Omega_1$. We identify the points $(0;y,\eta)$ and $(T;y,\eta)$, and then the set of intersections $\check{\mathfrak{O}}_{j-1}\cap\check{\mathfrak{O}}_j$ and the constant functions $\sigma_{j-1,j}$ defined on them is an integer cocycle on $\mathbb{S}^1\times O$. From the representation (1.5.6) it is clear that $-\alpha_\Gamma$ coincides with the value of this cocycle on the closed curve $(t;y_0,\eta_0)$, $0\le t\le T$, so α_Γ is determined by the corresponding Čech cohomology class.

2. Some explicit formulae. Let $r_+(x,\xi)$ and $r_-(x,\xi)$ be the numbers of the positive and negative eigenvalues of the matrix $\partial_{\xi\xi}h(x,\xi)$. Changing coordinates $x\to\widetilde{x}$ we obtain $h_{\widetilde{\xi}\,\widetilde{\xi}}=(\partial\widetilde{x}/\partial x)\cdot h_{\xi\xi}\cdot(\partial\widetilde{x}/\partial x)^T$. Therefore the definition of the functions $r_+(x,\xi)$ and $r_-(x,\xi)$ is independent of the choice of coordinates. Since $h_\xi(x,\xi)$ is positively homogeneous in ξ of degree zero Euler's identity implies $\operatorname{rank}\partial_{\xi\xi}h(x,\xi)\le n-1$ for all $(x,\xi)\in T'M$.

We have seen that in "good" cases the Maslov index α_Γ is the sum of the jumps of $r^-_j(t;y_0,\eta_0)$. It turns out that under some assumptions these jumps can be expressed in terms of $r_+(x^*,\xi^*)$, $r_-(x^*,\xi^*)$ and $\mathcal{R}(t;y,\eta)$. Here and below $(x^*,\xi^*)=(x^*(t;y,\eta),\xi^*(t;y,\eta))$.

LEMMA 1.5.5. *Let $x^*_\eta(t;y,\eta)=0$ and $\operatorname{rank}\partial_{\xi\xi}h(x^*,\xi^*)=n-1$ for some $(t;y,\eta)\in\check{\mathfrak{O}}_j$. If $s\ne 0$ is sufficiently small and $t+s\in[t_j(y,\eta),t_{j+1}(y,\eta)]$ then*

$$\operatorname{rank}C_j(t+s;y,\eta)\;=\;n-1 \tag{1.5.7}$$

and

$$r^-_j(t+s;y,\eta)\;=\;\begin{cases} r_+(x^*,\xi^*)\,, & s<0\,,\\ r_-(x^*,\xi^*)\,, & s>0\,.\end{cases}$$

LEMMA 1.5.6. *Let $r_+(x^*,\xi^*)=n-1$ for some $(t;y,\eta)\in\check{\mathfrak{O}}_j$. If s is sufficiently small and $t+s\in[t_j(y,\eta),t_{j+1}(y,\eta)]$ then*

$$r^-_j(t+s;y,\eta)\;=\;\begin{cases} r^-_j(t;y,\eta)-\mathcal{R}(t;y,\eta)+n-1\,, & s<0\,,\\ r^-_j(t;y,\eta)\,, & s\ge 0\,,\end{cases}$$

and (1.5.7) hold.

REMARK 1.5.7. In the general case, when the conditions of Lemma 1.5.5 or Lemma 1.5.6 are not fulfilled, the jumps of r^-_j depend not only on the matrix $\partial_{\xi\xi}h$ but also on the higher order derivatives of h.

EXAMPLE 1.5.8. Let M be a Riemannian manifold and $h(x,\xi) = |\xi|_x$ (see Example 1.2.4). Then $r_+(x,\xi) = n-1$ for all $(x,\xi) \in T'M$. Therefore by Lemma 1.5.6 and Definition 1.5.1 the Maslov index is the number of conjugate points counted with their multiplicities. Note that in this case the number of conjugate points on the trajectory Γ is finite; this fact follows from (1.5.7) and the obvious formula $\mathcal{R}(t;y,\eta) = \operatorname{rank} C_j(t;y,\eta)$.

3. Another definition. In this subsection we give another definition of the Maslov index which is more traditional. In fact, it is a modification of the classical definition suggested by V. Arnol'd [Ar1].

In Section 2.4 we will associate with the k-th leg of our family of billiard trajectories (1.5.1) a class $\mathfrak{F}_k$ of smooth complex-valued non-degenerate phase functions φ_k which are defined in a neighbourhood of the set

$$\{\,(t,x;y,\eta)\,:\, t \in [t_k^*(y,\eta), t_{k+1}^*(y,\eta)]\,,\ x = x^*(t;y,\eta)\,\}\,, \qquad k = 0,\dots,\mathbf{r},$$

where $t_0^*(y,\eta) := 0$ and $t_{\mathbf{r}+1}^*(y,\eta) := T$. The fact that $\varphi_k(t,x;y,\eta)$ is a phase function means that $\operatorname{Im}\varphi_k \ge 0$ and φ_k is positively homogeneous in η of degree 1. A phase function φ_k belongs to $\mathfrak{F}_k$ if

$$(\varphi_k)_\eta(t,x;y,\eta)\big|_{x=x^*} = 0\,, \quad (\varphi_k)_x(t,x;y,\eta)\big|_{x=x^*} = \xi^*(t;y,\eta)\,, \tag{1.5.8}$$

and

$$\det(\varphi_k)_{x\eta}(t,x;y,\eta)\big|_{x=x^*} \ne 0$$

for all $t \in [t_k^*(y,\eta), t_{k+1}^*(y,\eta)]$ (see Section 2.4 for more precise definitions).

Let us choose a matching sequence of phase functions $\varphi_k \in \mathfrak{F}_k$, $k = 0,\dots,\mathbf{r}$, for which

$$\varphi_k(t,x;y,\eta)\big|_{x\in\partial M} = \varphi_{k+1}(t,x;y,\eta)\big|_{x\in\partial M} \tag{1.5.9}$$

(by Lemma 2.6.3 such a sequence always exists). The matrices $(\varphi_k)_{x\eta}$ behave as tensors under changes of coordinates, i.e., they are multiplied by the Jacoby matrices (see Section 2.2). So under changes of coordinates the expressions $(\det(\varphi_k)_{x\eta})^2$ are multiplied by positive numbers, which does not change their arguments. Moreover, the condition (1.5.9) implies that

$$\begin{aligned}\big(\det{}^2(\varphi_k)_{x\eta} \,/\, |\det{}^2(\varphi_k)_{x\eta}|\big)\big|_{x=x^*(t_{k+1}^*;y,\eta)} \\ = \big(\det{}^2(\varphi_{k+1})_{x\eta} \,/\, |\det{}^2(\varphi_{k+1})_{x\eta}|\big)\big|_{x=x^*(t_{k+1}^*;y,\eta)}\end{aligned}$$

(see formula (2.6.14)). Therefore we can introduce on $[0,T]\times O$ the continuous function f defined by the equalities

$$f(t;y,\eta) := \big(\det{}^2(\varphi_k)_{x\eta} \,/\, |\det{}^2(\varphi_k)_{x\eta}|\big)\big|_{x=x^*(t;y,\eta)} \tag{1.5.10}$$

where $t \in [t_k^*(y,\eta), t_{k+1}^*(y,\eta)]$, $k = 0,\dots,\mathbf{r}$.

Obviously, $|f| \equiv 1$. Since $x^*|_{t=0} \equiv y$, the conditions (1.5.8) imply that $(\varphi_0)_{x\eta} \equiv I$ for $t = 0$, $x = y$ (see subsection 2.4.1) and, consequently, $f(0;y,\eta) \equiv 1$. Therefore there exists a unique continuous branch $\arg_0 f(t;y,\eta)$ of the argument $\arg f(t;y,\eta)$ such that $\arg_0 f(0;y,\eta) \equiv 0$.

Under the condition (1.5.2) the matrix $(\varphi_{\mathbf{r}})_{x\eta}$ is real for $t = T$, $x = x^*(T;y_0,\eta_0)$. Consequently, $\arg_0 f(T;y_0,\eta_0)$ is a multiple of 2π.

DEFINITION 1.5.9. The integer number $-(2\pi)^{-1}\arg_0 f(T;y_0,\eta_0)$ is called the Maslov index of the trajectory Γ.

In other words, the Maslov index is equal to the sum over $k=0,\dots,\mathbf{r}$ of the variations of $-(2\pi)^{-1}\arg(\det(\varphi_k)_{x\eta})^2\big|_{x=x^*(t;y_0,\eta_0)}$ as t goes from $t_k^*(y_0,\eta_0)$ to $t_{k+1}^*(y_0,\eta_0)$.

The equivalence of Definitions 1.5.1 and 1.5.9 will be proved in Appendix D.

REMARK 1.5.10. When $\partial M=\varnothing$ we have only one phase function $\varphi=\varphi_0$ (instead of the sequence $\{\varphi_k\}$). Let us introduce the differential 1-form

$$-(2\pi)^{-1}\,d\left(\arg(\det\varphi_{x\eta})^2\big|_{x=x^*}\right).$$

Obviously, the Maslov index α_Γ coincides with the integral of this 1-form over the curve $(t;y_0,\eta_0)$, $0\le t\le T$. Under the assumptions of Remark 1.5.4 this integral is determined by the corresponding de Rham cohomology class, which is the image of the Čech cohomology class from Remark 1.5.4 provided by the standard isomorphism of the Čech and de Rham cohomology groups (see [LapSaVa] for details).

1.6. Classical two-term asymptotic formula for $N(\lambda)$

1. Statement of the result. Having introduced the geometrical concepts of nonperiodicity (see Definition 1.3.7 and its subsequent generalizations in subsections 1.3.2, 1.3.3) and nonblocking (Definition 1.3.22), we can now formulate the theorem which gives the classical two-term asymptotic formula for $N(\lambda)$. For the case $\partial M=\varnothing$ it was established in [DuiGui], and for the case $m=1$ in [Iv1], [Me]. The general result appeared in [Va3], [Va4], [Va7].

THEOREM 1.6.1. *If the nonperiodicity and nonblocking conditions are fulfilled, then*

$$N(\lambda)=c_0\lambda^n+c_1\lambda^{n-1}+o(\lambda^{n-1})\qquad\lambda\to+\infty. \tag{1.6.1}$$

Here the coefficient c_0 is the same as in Theorem 1.2.1 and

$$c_1=\int\limits_{T^*\partial M}\operatorname{shift}^+(1;x',\xi')\,dx'\,đ\xi', \tag{1.6.2}$$

where $\operatorname{shift}^+(\nu;x',\xi')$ for fixed $(x',\xi')\in T^*\partial M$ is the spectral shift (see subsection 3) of the auxiliary one-dimensional problem (1.1.7), (1.1.8) on the half-line. For manifolds without boundary $c_1=0$.

REMARK 1.6.2. Formula (1.6.1) can be written down in an equivalent form

$$\lambda_k=c_0^{-1/n}k^{1/n}-c_1(nc_0)^{-1}+o(1)\qquad k\to+\infty. \tag{1.6.1$'$}$$

The remainder of this section is split into five subsections. In subsection 2 we briefly discuss the conditions appearing in Theorem 1.6.1. In subsection 3 we define the spectral shift and produce formulae for its effective evaluation. In subsection 4 we discuss the formula for the second asymptotic coefficient and give two elementary examples of its calculation. In subsection 5 we formulate a modified version of Theorem 1.6.1 suited for the case of a pseudodifferential operator acting on a manifold without boundary.

2. Discussion of the result. Let us turn once again to Example 1.2.5 (Laplacian on a sphere). We have already noted that in this example the second asymptotic term of $N(\lambda)$ is of the order $\lambda^{(n-1)/(2m)}$. However, it is not of the form $c_1\lambda^{(n-1)/(2m)}$, i.e., the classical two-term Weyl formula (1.6.1) does not hold. Indeed, it follows from (1.2.4), (1.2.5) that there is a sequence $\Lambda_j \to +\infty$ such that $N(\Lambda_j + 0) - N(\Lambda_j) \geq \operatorname{const} \Lambda_j^{n-1}$, $\operatorname{const} > 0$, whereas (1.6.1) would imply $N(\lambda+0)-N(\lambda) = o(\lambda^{n-1})$ as $\lambda \to +\infty$. In other words, looking at the fine structure of the spectrum we see that in Example 1.2.5 the eigenvalues are unevenly distributed along the spectrum, they cluster into groups of growing multiplicity.

The reason for the uneven distribution of eigenvalues in Example 1.2.5 is clear — in this case the spectral problem (1.1.1), (1.1.2) has a very rich group of symmetries and, consequently, the eigenvalues have very high multiplicities. Such "pathologically" (in the terminology of [DuiGui]) symmetric cases have to be excluded if we want to obtain a classical (polynomial) two-term asymptotic formula for $N(\lambda)$, and this is why we need the nonperiodicity condition. It was first introduced in [DuiGui] for manifolds without boundary and it appears in all the subsequent works on classical two-terms spectral asymptotics.

The fact that the (asymptotic) symmetries of the spectral problem (1.1.1), (1.1.2) can be described in terms of billiard trajectories is a nontrivial one. The precise role of billiard trajectories will be revealed in further chapters in the course of our proof. However, the underlying idea is worth mentioning here. Let us introduce the time variable t by a change $\lambda \to D_t$ $(= -i\partial/\partial t)$, i.e., let us consider the nonstationary equation

$$Av = D_t^{2m} v \tag{1.6.3}$$

with boundary conditions (1.1.2). One can single out (see Section 3.1) solutions of (1.6.3), (1.1.2) of the form $v(x,t) = \exp(-itA^{1/(2m)})v_0(x)$. Clearly, such solutions contain full information about the spectral problem (1.1.1), (1.1.2). Analysis of the nonstationary problem (1.6.3), (1.1.2) shows that the singularities of such solutions propagate along the billiard trajectories defined in Section 1.3, so it is natural to use them for describing global geometrical characteristics.

Another way of explaining the appearance of billiard trajectories is to try to solve the spectral problem (1.1.1), (1.1.2) asymptotically. Let us search for the eigenfunction v in the form

$$v = a(x,\lambda)\, \exp(i\lambda\varphi(x))\,, \tag{1.6.4}$$

where $a(x,\lambda) = a_0(x) + \lambda^{-1}a_{-1}(x) + \lambda^{-2}a_{-2}(x) + \dots$, $\lambda \to +\infty$. Substituting (1.6.4) into (1.1.1) and leaving only the terms with the leading power of λ, we obtain the eikonal equation $A_{2m}(x, \varphi_x) = 1$ the solution of which can be expressed in terms of Hamiltonian trajectories (1.3.1); see, e.g., [Sh]. A similar construction with m compensating exponents added to (1.6.4) allows one to satisfy asymptotically the boundary conditions (1.1.2), with the phase functions $\varphi^{(l)}$ of the compensating exponents being expressed in terms of reflected billiard trajectories. Unfortunately, the exponential representation technique works only locally, in the neighbourhood of a fixed point $(x_0, \varphi_x|_{x=x_0}) \in T'M$. In the general case one can not solve the equations (1.1.1) and (1.1.2) on the whole of M and ∂M using this technique, but it gives an idea why billiard trajectories appear in spectral asymptotics.

The nonblocking condition is more technical and its necessity in Theorem 1.6.1 is less obvious. Spectral problems which do not satisfy the nonblocking condition have not yet been studied; we can only say that for such problems the geometric picture may be very complicated and modern microlocal techniques fail to describe it adequately.

The nonblocking condition was introduced by the authors in two independent works [Va4] and [Sa1].

3. Spectral shift of the auxiliary one-dimensional problem. Let us consider the one-dimensional spectral problem (1.1.7), (1.1.8). In this subsection we define for (1.1.7), (1.1.8) the notion of the spectral shift, and give (without proof) a convenient trace formula for its evaluation. A more detailed analysis with the proofs is carried out in Appendix A.

As in Section 1.4, let us omit for the time being the parameter (x', ξ'), and rewrite (1.1.7), (1.1.8) as (1.4.1), (1.4.2). Recall that by $\mathbf{A}^+$ we denote the self-adjoint operator in $L_2(\mathbb{R}_+)$ associated with (1.4.1), (1.4.2).

Let us also consider the spectral problem (1.4.1) on the whole line $\mathbb{R}$ without boundary conditions, and let us denote by $\mathbf{A}$ the corresponding self-adjoint operator in $L_2(\mathbb{R})$.

Denote by $\mathbf{E}^+_\nu$, $\mathbf{E}_\nu$ the spectral projections of the operators $\mathbf{A}^+$, $\mathbf{A}$, respectively. For definiteness we choose $\mathbf{E}^+_\nu$, $\mathbf{E}_\nu$ to be left-continuous in ν. For each fixed ν our spectral projections are integral operators

$$\mathbf{E}^+_\nu \;=\; \int_0^{+\infty} \mathbf{e}^+(\nu, x_n, y_n)\,(\,\cdot\,)\,dy_n\,, \qquad \mathbf{E}_\nu \;=\; \int_{-\infty}^{+\infty} \mathbf{e}(\nu, x_n, y_n)\,(\,\cdot\,)\,dy_n$$

with continuous kernels $\mathbf{e}^+$, $\mathbf{e}$.

Denote by $\theta : L_2(\mathbb{R}) \to L_2(\mathbb{R}_+)$ the restriction operator; its adjoint $\theta^* : L_2(\mathbb{R}_+) \to L_2(\mathbb{R})$ is the extension operator defined by

$$(\theta^* v)(x_n) \;=\; \begin{cases} 0\,, & x_n < 0\,, \\ v(x_n)\,, & x_n \ge 0\,. \end{cases}$$

By θ_X we shall denote the characteristic function of the interval $[0, X]$, and by Tr we shall denote the trace of an operator.

Definition 1.6.3. The *spectral shift* of the problem (1.4.1), (1.4.2) is the function on $\mathbb{R}$ defined as follows:

(1) for ν which are not thresholds (see Section 1.4)

$$\text{shift}^+(\nu) \;=\; \int_0^{+\infty} (\mathbf{e}^+(\nu, x_n, x_n) - \mathbf{e}(\nu, x_n, x_n))\,dx_n \tag{1.6.5}$$
$$\equiv \lim_{X \to +\infty} \operatorname{Tr}(\theta_X(\mathbf{E}^+_\nu - \theta \mathbf{E}_\nu \theta^*)\theta_X)\,;$$

(2) at thresholds $\text{shift}^+(\nu)$ is defined by left-continuity.

Convergence of the integral (1.6.9) follows from the asymptotic formula

$$\mathbf{e}^+(\nu, x_n, x_n) \,-\, \mathbf{e}(\nu, x_n, x_n) \;= \tag{1.6.6}$$
$$\sum_{k,l=1}^{q} \frac{\sqrt{-\,A'(\zeta_l^+(\nu))\,A'(\zeta_k^-(\nu))}}{\pi\, x_n \big(A'(\zeta_l^+(\nu)) - A'(\zeta_k^-(\nu))\big)}\, \mathrm{Im}\big(R_{lk}(\nu)\, e^{i x_n (\zeta_l^+(\nu) - \zeta_k^-(\nu))}\big) \;+\; O(x_n^{-2})$$

as $x_n \to +\infty$; here the $R_{lk}(\nu)$ are the elements of the reflection matrix and the $\zeta_l^{\pm}(\nu)$ are the real ζ-roots of the algebraic equation $A(\zeta) = \nu$, see Section 1.4 for details. Formula (1.6.6) implies that normally the integral (1.6.5) is not absolutely convergent. The absence of absolute convergence in (1.6.5) implies, in turn, that the operator $\mathbf{E}_\nu^+ - \theta \mathbf{E}_\nu \theta^*$ is not of trace class.

For a positive self-adjoint operator with a purely discrete spectrum the trace of the spectral projection is the counting function (i.e., the number of eigenvalues below a given ν). Hence the notion of the spectral shift is very similar to that of the counting function. The difference is that the spectral shift may be defined for self-adjoint operators the spectra of which are not purely discrete; in this situation the spectral projection is not of trace class and one has to perform a regularization procedure. In our particular case (1.6.5) regularization is performed in two stages: firstly, by subtracting the spectral projection of a basic (reference) operator, and secondly, by introducing the regularizing spatial cut-off θ_X.

Denote by $\sigma(\mathbf{A}^+)$, $\sigma(\mathbf{A})$ the spectra of the operators $\mathbf{A}^+$, $\mathbf{A}$. We have

$$[\nu_1^{\mathrm{st}}, +\infty) = \sigma(\mathbf{A}) \subset \sigma(\mathbf{A}^+) \subset [0, +\infty)\,,$$

where ν_1^{st} is as defined in Section 1.4, and the last inclusion is a consequence of Conditions 1.1.1′, 1.1.4′. Thus, for any $\mu \in \mathbb{C} \setminus \sigma(\mathbf{A}^+)$ we can define the resolvents $\mathbf{R}_\mu^+ = (\mathbf{A}^+ - \mu I)^{-1}$ and $\mathbf{R}_\mu = (\mathbf{A} - \mu I)^{-1}$. For each fixed μ these resolvents are integral operators

$$\mathbf{R}_\mu^+ \;=\; \int_0^{+\infty} \mathbf{r}^+(\mu, x_n, y_n)\,(\,\cdot\,)\,dy_n\,, \qquad \mathbf{R}_\mu \;=\; \int_{-\infty}^{+\infty} \mathbf{r}(\mu, x_n, y_n)\,(\,\cdot\,)\,dy_n$$

with continuous kernels $\mathbf{r}^+$, $\mathbf{r}$.

For any $\mu \in \mathbb{C} \setminus \sigma(\mathbf{A}^+)$ the operator $\mathbf{R}_\mu^+ - \theta \mathbf{R}_\mu \theta^*$ is of trace class (moreover, it is of finite rank), and $\mathbf{r}^+(\mu, x_n, x_n) - \mathbf{r}(\mu, x_n, x_n)$ tends to zero exponentially as x_n tends to $+\infty$. Thus, the expression

$$\mathbf{f}^+(\mu) := \mathrm{Tr}(\mathbf{R}_\mu^+ - \theta \mathbf{R}_\mu \theta^*) \equiv \int_0^{+\infty} (\mathbf{r}^+(\mu, x_n, x_n) - \mathbf{r}(\mu, x_n, x_n))\,dx_n \tag{1.6.7}$$

is well defined. The function $\mathbf{f}^+$ is analytic in $\mathbb{C} \setminus \sigma(\mathbf{A}^+)$, and for any given $\nu \in \sigma(\mathbf{A}^+)$ which is not a threshold and not an eigenvalue of the problem (1.4.1), (1.4.2) $\mathbf{f}^+(\mu)$ is bounded as $\mathbb{C} \setminus \sigma(\mathbf{A}^+) \ni \mu \to \nu$.

Definition 1.6.3 is equivalent to

DEFINITION 1.6.4. The *spectral shift* of the problem (1.4.1), (1.4.2) is the function on $\mathbb{R}$ defined as follows:

(1) $\operatorname{shift}^+(\nu) = 0$ for $\nu \le 0$;
(2) for positive ν which are not thresholds and are not eigenvalues

$$\operatorname{shift}^+(\nu) \;=\; (-2\pi i)^{-1} \int\limits_{L(\nu)} \mathbf{f}^+(\mu)\, d\mu \;, \tag{1.6.8}$$

where $L(\nu)$ is the oriented arc of the circumference $|\mu| = \nu$ in the complex μ-plane going counterclockwise from $\nu + i0$ to $\nu - i0$;
(3) at thresholds and eigenvalues $\operatorname{shift}^+(\nu)$ is defined by left-continuity.

Let us now list the basic properties of our spectral shift.

LEMMA 1.6.5. *The function* shift^+ *is bounded:*

$$|\operatorname{shift}^+(\nu)| \le m, \qquad \forall \nu \in \mathbb{R}.$$

Recall that $2m$ is the order of the differential equation (1.4.1).

Let us denote by $\mathbf{N}^+(\nu)$ the counting function of the spectral problem (1.4.1), (1.4.2), that is the number of eigenvalues below a given ν. Here we count both eigenvalues outside the continuous spectrum and those embedded in the continuous spectrum.

LEMMA 1.6.6. *The function* $\operatorname{shift}^+(\nu) - \mathbf{N}^+(\nu)$ *is continuous in each of the zones (1.4.4) of the continuous spectrum.*

Lemma 1.6.6 implies that the function shift^+ may be discontinuous only at thresholds and eigenvalues. With account of Proposition A.1.13 this means that the number of discontinuities of the function shift^+ is finite, and, moreover, uniformly bounded over all problems of this type with fixed order $2m$.

LEMMA 1.6.7. *For any* $\mu \in \mathbb{C} \setminus [\min \sigma(\mathbf{A}^+), +\infty)$

$$\mathbf{f}^+(\mu) \;=\; \int\limits_0^{+\infty} \frac{\operatorname{shift}^+(\nu)}{(\nu - \mu)^2}\, d\nu \,.$$

A consequence of Lemma 1.6.7 is the following result which can be used for the evaluation of the jumps of the function shift^+.

COROLLARY 1.6.8. *For any* $\nu \in \mathbb{R}$

$$\operatorname{shift}^+(\nu + 0) \;-\; \operatorname{shift}^+(\nu) \;=\; \lim_{\substack{\mu \to \nu \\ \mu \in \mathbb{C} \setminus \sigma(\mathbf{A}^+)}} (\nu - \mu)\, \mathbf{f}^+(\mu) \;.$$

If ν is an eigenvalue which is not a threshold, then it it is not necessary to use Corollary 1.6.8 because by Lemma 1.6.6 $\operatorname{shift}^+(\nu + 0) - \operatorname{shift}^+(\nu)$ is just the multiplicity of the eigenvalue ν. Corollary 1.6.8 is really needed for computing the jumps of the function shift^+ at thresholds. Corollary 1.6.8 is, however, not very convenient. Below we consider the important special case of a normal threshold (see Definition 1.4.4). Then the formula for the jump of the function shift^+ becomes completely explicit.

DEFINITION 1.6.9. We call the normal threshold ν^{st} *soft* if for $\nu = \nu^{\mathrm{st}}$ the problem (1.4.1), (1.4.2) has a solution of the form

$$v(x_n) = e^{ix_n\zeta^{\mathrm{st}}} + w(x_n)\,, \quad \text{where} \quad w(x_n) \to 0 \quad \text{as} \quad x_n \to +\infty\,,$$

and *rigid* if it has no solution of this form. Here ζ^{st} is as in Definition 1.4.4.

LEMMA 1.6.10. *If the threshold ν^{st} is normal then*

$$\operatorname{shift}^+(\nu^{\mathrm{st}}+0) - \operatorname{shift}^+(\nu^{\mathrm{st}}) = \mathbf{N}^+(\nu^{\mathrm{st}}+0) - \mathbf{N}^+(\nu^{\mathrm{st}}) \pm \frac{1}{4}\,,$$

where the plus or minus sign is chosen according to whether the threshold ν^{st} is soft or rigid, respectively.

It follows from Lemma 1.6.10 that a normal threshold produces a jump in the spectral shift which in absolute value is one quarter of the jump produced by an eigenvalue.

Lemmas 1.6.6, 1.6.10 and Corollary 1.6.8 provide us with an effective description of the jumps of the spectral shift, but we are still left with the problem of determining its continuous part. Using Definitions 1.6.3 or 1.6.4 for this purpose is inconvenient. Indeed, Definition 1.6.3 involves the spectral function $\mathbf{e}^+(\nu, x_n, y_n)$ of the problem (1.4.1), (1.4.2), the effective construction of which is not completely trivial, and, moreover, it involves integration in x_n. Definition 1.6.4 is better in the sense that it does not use the spectral function, but on the other hand it contains a double integral (see (1.6.7), (1.6.8)). Fortunately, this inconvenience can be overcome. The following lemma gives a formula for the continuous part of the spectral shift which does not contain the spectral function or any integrations. Everything is expressed through the reflection matrix $R(\nu)$ introduced in Section 1.4.

LEMMA 1.6.11. *For each zone (1.4.4) of the continuous spectrum there exists a continuous branch* $\arg_0 \det\bigl(i\,R(\nu)\bigr)$ *of the argument of* $\det\bigl(i\,R(\nu)\bigr)$ *such that*

$$\operatorname{shift}^+(\nu) = \mathbf{N}^+(\nu) + \frac{\arg_0 \det\bigl(i\,R(\nu)\bigr)}{2\pi} \tag{1.6.9}$$

in this zone. For ν lying below the continuous spectrum $\operatorname{shift}^+(\nu) = \mathbf{N}^+(\nu)$.

REMARK 1.6.12. Formula (1.6.9) is a special case of the standard trace formula from scattering theory

$$\operatorname{shift}(\nu) = \mathbf{N}(\nu) + \frac{\arg_0 \det S(\nu)}{2\pi}\,, \tag{1.6.10}$$

where $\mathbf{N}(\nu)$ is the counting function of the perturbed problem and $S(\nu)$ is the scattering matrix; see, e.g., [Ya]. There is, however, a technical difficulty in that we compare two spectral problems in different Hilbert spaces, namely, $L_2(\mathbb{R}_+)$ and $L_2(\mathbb{R})$. The natural way of overcoming this difficulty is to consider instead of the problem on the half-line $\mathbb{R}_+$ a problem on the perforated line $\mathbb{R}_+ \setminus \{0\}$ with boundary conditions

$$\Bigl(B^{(j)}(D_{x_n})v\Bigr)\Big|_{x_n=+0} = 0\,, \qquad j = 1, 2, \dots, m, \tag{1.6.11$^+$}$$

$$\Bigl(\overline{B^{(j)}}(D_{x_n})v\Bigr)\Big|_{x_n=-0} \,=\, 0\,, \qquad j=1,2,\dots,m \tag{1.6.11$^-$}$$

(cf. (1.4.2)), where $\overline{B^{(j)}}(D_{x_n}) \equiv \bigl(B^{(j)}(D_{x_n})\bigr)^*$. Then the problem (1.4.1), (1.6.11$^+$), (1.6.11$^-$) on the perforated line $\mathbb{R}_+ \setminus \{0\}$ is a spectral problem in the Hilbert space $L_2(\mathbb{R})$, and one can view it as a perturbation of the reference problem (1.4.1) on the full line $\mathbb{R}$ in the same Hilbert space $L_2(\mathbb{R})$. In this case one can develop a consistent scattering theory, see Appendix A, and show that (1.6.10) holds. Moreover, the scattering matrix in this case is given by the formula

$$S(\nu) \,=\, \begin{pmatrix} 0 & R(\nu) \\ R^T(\nu) & 0 \end{pmatrix}, \tag{1.6.12}$$

where $R(\nu)$ is the reflection matrix defined in Section 1.4. Formulae (1.6.10), (1.6.12) imply

$$\operatorname{shift}(\nu) \,=\, \mathbf{N}(\nu) \,+\, \frac{\arg_0 \det{}^2\bigl(i\,R(\nu)\bigr)}{2\pi}\,. \tag{1.6.13}$$

But the problem (1.4.1), (1.6.11$^+$), (1.6.11$^-$) on the perforated line $\mathbb{R}_+ \setminus \{0\}$ is invariant under the change $v(x_n) \to \overline{v(-x_n)}$, consequently

$$\operatorname{shift}(\nu) \,=\, 2\,\operatorname{shift}^+(\nu)\,, \qquad \mathbf{N}(\nu) \,=\, 2\,\mathbf{N}^+(\nu)\,. \tag{1.6.14}$$

From (1.6.13), (1.6.14) we obtain

$$\operatorname{shift}^+(\nu) \,=\, \mathbf{N}^+(\nu) \,+\, \frac{\arg_0\bigl(\pm\det\bigl(i\,R(\nu)\bigr)\bigr)}{2\pi}$$

with either a plus or a minus. A more detailed analysis shows that the sign in the latter formula is a plus, and we arrive at (1.6.9).

Further on we use the notation $\arg_0\det\bigl(i\,R(\nu)\bigr)$ for the particular branch of the argument $\arg\det\bigl(i\,R(\nu)\bigr)$ specified by Lemma 1.6.11. It will be convenient for us to set $\arg_0\det\bigl(i\,R(\nu)\bigr) \equiv 0$ for ν lying below the continuous spectrum, and to define $\arg_0\det\bigl(i\,R(\nu)\bigr)$ at thresholds by left-continuity. Under such a convention the trace formula (1.6.9) holds for all $\nu \in \mathbb{R}$.

Lemmas 1.6.10, 1.6.11 immediately imply

Corollary 1.6.13. *If the threshold ν^{st} is normal then*

$$\bigl|\arg_0\det\bigl(i\,R(\nu^{\mathrm{st}}+0)\bigr) - \arg_0\det\bigl(i\,R(\nu^{\mathrm{st}})\bigr)\bigr| \,=\, \frac{\pi}{2}\,.$$

If ν^{st} is a normal threshold and if we know the choice of the branch $\arg_0\det\bigl(i\,R(\nu)\bigr)$ of the argument $\arg\det\bigl(i\,R(\nu)\bigr)$ for $\nu \le \nu^{\mathrm{st}}$, then Corollary 1.6.13 allows us to choose uniquely the branch $\arg_0\det\bigl(i\,R(\nu)\bigr)$ of the argument $\arg\det\bigl(i\,R(\nu)\bigr)$ for $\nu > \nu^{\mathrm{st}}$. This is possible because according to Corollary 1.6.13 $\arg_0\det\bigl(i\,R(\nu^{\mathrm{st}}+0)\bigr)$ belongs to the interval

$$\bigl[\arg_0\det\bigl(i\,R(\nu^{\mathrm{st}})\bigr) - \pi/2\,,\, \arg_0\det\bigl(i\,R(\nu^{\mathrm{st}})\bigr) + \pi/2\bigr]$$

the length of which is less than 2π. One does not even have to check whether the threshold ν^{st} is soft or rigid. Thus, if all the thresholds are normal we can determine consecutively the branches $\arg_0 \det\bigl(i\,R(\nu)\bigr)$ on all the intervals (1.4.4) of the continuous spectrum.

Matters are facilitated even further if our one-dimensional problem (1.4.1), (1.4.2) satisfies the strong simple reflection condition (see Definition 1.4.5). In this case

$$\arg_0 \det\bigl(i\,R(\nu)\bigr) \;=\; \begin{cases} 0\,, & \text{if} \quad \nu \le \nu_1^{\mathrm{st}}\,, \\ \arg_0 \left(\frac{i\,a_1^+(\nu)}{a_1^-(\nu)}\right)\,, & \text{if} \quad \nu > \nu_1^{\mathrm{st}}\,, \end{cases} \tag{1.6.15}$$

where the $a_1^{\pm}(\nu)$ are the coefficients from (1.4.7) and the branch of the argument is uniquely specified by the condition

$$\left|\arg_0 \left(\frac{i\,a_1^+(\nu_1^{\mathrm{st}}+0)}{a_1^-(\nu_1^{\mathrm{st}}+0)}\right)\right| \;=\; \frac{\pi}{2}\,. \tag{1.6.16}$$

Further on we indicate the dependence of all our quantities on the parameter $(x',\xi') \in T'\partial M$. The reasoning from the end of Section 1.4 (see formula (1.4.9)) shows that under changes of local coordinates $\mathrm{shift}^+(\nu;x',\xi')$, $\mathbf{N}^+(\nu;x',\xi')$, and $\arg_0 \det\bigl(i\,R(\nu;x',\xi')\bigr)$ behave as functions on $T'\partial M$.

We shall often be using the following two properties of the spectral shift:

$$\mathrm{shift}^+(\lambda^{2m}\nu;x',\lambda\xi') \;=\; \mathrm{shift}^+(\nu;x',\xi')\,, \qquad \forall \lambda>0\,; \tag{1.6.17}$$

for any given local coordinate system x there exists a positive constant such that

$$\mathrm{shift}^+(\nu;x',\xi') \;=\; 0 \qquad \text{for} \qquad |\xi'|^{2m} > \mathrm{const}\;\nu\,. \tag{1.6.18}$$

The rescaling property (1.6.17) follows from the homogeneity of $A(x,\xi)$, $B_{m_j}^{(j)}(x',\xi)$ in ξ. The property (1.6.18) follows from (1.6.17) and the fact that for any given $\xi' \ne 0$ the spectra of the operators $\mathbf{A}^+$, $\mathbf{A}$ are strictly positive.

4. The second asymptotic coefficient: discussion and examples. First, let us show that the integral (1.6.2) exists in the usual Riemann sense. In view of (1.6.18) the domain of integration in (1.6.2) is, in fact, bounded. Denote by Σ' the set of points $(x',\xi') \in T^*\partial M$ such that 1 is an eigenvalue or a threshold of the one-dimensional problem (1.1.7), (1.1.8); obviously, the set Σ' is closed and bounded. By Lemma A.4.1 the function $\mathrm{shift}(1;x',\xi')$ is continuous on $T^*\partial M \setminus \Sigma'$, and by Lemma 1.6.5 it is uniformly bounded on $T^*\partial M$. So in order to prove Riemann integrability it is sufficient to show that the $(2n-2)$-dimensional Jordan measure of Σ' is zero. As the set Σ' is closed and bounded, this is equivalent to the $(2n-2)$-dimensional Lebesgue measure of Σ' being zero. In view of Tonelli's theorem, in order to prove the latter it is sufficient to show that for any $(x'_0,\xi'_0) \in T'\partial M$ the ray $\{(x'_0,\lambda\xi'_0),\ \lambda>0\}$ intersects Σ' at a finite number of points. By rescaling, a point $(x'_0,\lambda\xi'_0)$ belongs to Σ' if and only if $\nu=\lambda^{-2m}$ is is an eigenvalue or a threshold of the one-dimensional problem (1.1.7), (1.1.8) with $(x',\xi') \equiv (x'_0,\xi'_0)$, and we already know that the number of such ν is finite. This completes the proof of Riemann integrability.

Substituting (1.6.14) into (1.6.2) we obtain

$$c_1 \,=\, \int\limits_{T^*\partial M} \left(\mathbf{N}^+(1; x', \xi') \,+\, \frac{\arg_0 \det\bigl(i\,R(1; x', \xi')\bigr)}{2\pi} \right) dx'\,đ\xi' \,. \tag{1.6.19}$$

Formula (1.6.19) proves to be very convenient for the practical evaluation of the coefficient c_1.

Let us illustrate the algorithm of computing the coefficient c_1 by a few elementary examples.

EXAMPLE 1.6.14. Let us consider the spectral problem

$$-\Delta v \,=\, \lambda^2 v \,, \tag{1.6.20}$$

$$v|_{\partial M} \,=\, 0 \tag{1.6.21}$$

in a region $M \subset \mathbb{R}^2$. Here $\Delta = \partial^2/\partial y_1^2 + \partial^2/\partial y_2^2$ is the Laplacian and $y = (y_1, y_2)$ are Cartesian coordinates in $\mathbb{R}^2$. We already know (see Example 1.2.3) that for this problem $c_0 \,=\, S/(4\pi)$, where S is the surface area of M. For computing the coefficient c_1 we introduce special local coordinates $x = (x_1, x_2)$ in the small neighbourhood of ∂M (see Fig. 9): we associate with a point $P \in M$ the nearest point $\widetilde{P} \in \partial M$ and then we take x_2 to be the length of the (straight) line segment $P\widetilde{P}$ and x_1 to be the length of the (curvilinear) arc $P_0\widetilde{P} \subset \partial M$, where $P_0 \in \partial M$ is some fixed point.

Figure 9. Local coordinates x.

The auxiliary one-dimensional spectral problem associated with (1.6.20), (1.6.21) is

$$-d^2v/dx_2^2 + \xi_1^2 v \,=\, \nu v \,, \tag{1.6.22}$$

$$v|_{x_2=0} \,=\, 0 \,, \tag{1.6.23}$$

where $v \equiv v(x_2)$, $x_2 \in \mathbb{R}_+$. In the subsequent analysis we assume that $\xi_1 \neq 0$; this can be done because the integral (1.6.19) does not depend on the value of the integrand at a particular ξ_1.

The problem (1.6.22), (1.6.23) has no eigenvalues, so

$$\mathbf{N}^+(\nu; x_1, \xi_1) \equiv 0 \,. \tag{1.6.24}$$

The problem (1.6.22), (1.6.23) has only one threshold $\nu_1^{\mathrm{st}} = \xi_1^2$, and the continuous spectrum is the semi-infinite interval $[\xi_1^2, +\infty)$. The points $\nu > \xi_1^2$ of the continuous spectrum have multiplicity one, and the corresponding generalized eigenfunctions have the form

$$v(x_2) \,=\, \sin\Bigl(x_2\sqrt{\nu-\xi_1^2}\Bigr) \,=\, \frac{a_1^-\, e^{ix_2\zeta_1^-(\nu)}}{\sqrt{-2\pi A'(\zeta_1^-(\nu))}} \,+\, \frac{a_1^+\, e^{ix_2\zeta_1^+(\nu)}}{\sqrt{2\pi A'(\zeta_1^+(\nu))}}$$

(cf. (1.4.7)), where $\zeta_1^{\pm}(\nu) = \pm\sqrt{\nu-\xi_1^2}$ and

$$a_1^{\pm} \,=\, \mp i\,\sqrt{\pi\sqrt{\nu-\xi_1^2}}\,. \tag{1.6.25}$$

As the strong simple reflection condition is satisfied, we can use formulae (1.6.15), (1.6.16). Substituting (1.6.25) into (1.6.15) we obtain for $\nu > \xi_1^2$

$$\arg_0 \det\bigl(i\,R(\nu; x_1, \xi_1)\bigr) \,=\, -\pi/2 + 2\pi k \tag{1.6.26}$$

with an unknown integer k. Substituting (1.6.26) into (1.6.16) we establish that $k = 0$. Thus,

$$\arg_0 \det\bigl(i\,R(\nu; x_1, \xi_1)\bigr) \,=\, \begin{cases} 0\,, & \text{if} \quad \nu \le \xi_1^2\,, \\ -\pi/2\,, & \text{if} \quad \nu > \xi_1^2\,. \end{cases} \tag{1.6.27}$$

Finally, substituting (1.6.24) and (1.6.27) into (1.6.19) we obtain

$$c_1 \,=\, \int_0^L \int_{-\infty}^{+\infty} \frac{\arg_0 \det\bigl(i\,R(1; x_1, \xi_1)\bigr)}{2\pi}\, đ\xi_1\, dx_1 \,=\, -\frac{1}{4}\int_0^L \int_{-1}^{1} đ\xi_1\, dx_1 \,=\, -\frac{L}{4\pi}\,, \tag{1.6.28}$$

where L is the length of ∂M.

It is interesting to determine whether in our case the normal threshold ν_1^{st} is soft or rigid, and check whether Lemma 1.6.10 gives the proper sign for the jump. The function $v(x_n)$ from Definition 1.6.9 in this case has the form

$$v(x_2) \,\equiv\, 1\,. \tag{1.6.29}$$

Obviously, the function (1.6.29) does not satisfy the boundary condition (1.6.23), so the threshold is rigid. Accordingly, the sign of $\arg_0 \det\bigl(i\,R(\nu_1^{\mathrm{st}} + 0; x_1, \xi_1)\bigr)$ is negative.

Example 1.6.15. Consider the same spectral problem as in Example 1.6.14, but with the Neumann boundary condition

$$\partial v/\partial x_2|_{\partial M} \,=\, 0 \tag{1.6.30}$$

instead of (1.6.21). Of course, the coefficient c_0 is the same as in Example 1.6.14. In order to compute the coefficient c_1 we have to consider the auxiliary one-dimensional spectral problem associated with (1.6.20), (1.6.30). This one-dimensional problem is described by the equation (1.6.22) with boundary condition

$$dv/dx_2|_{x_2=0} \,=\, 0\,. \tag{1.6.31}$$

The problem (1.6.22), (1.6.31) has no eigenvalues, so (1.6.24) remains true.

The threshold and the continuous spectrum are the same as in Example 1.6.14, but the generalized eigenfunctions now have the form

$$v(x_2) \,=\, \cos\left(x_2\sqrt{\nu-\xi_1^2}\right) \,=\, \frac{a_1^-\, e^{ix_2\zeta_1^-(\nu)}}{\sqrt{-2\pi A'(\zeta_1^-(\nu))}} \,+\, \frac{a_1^+\, e^{ix_2\zeta_1^+(\nu)}}{\sqrt{2\pi A'(\zeta_1^+(\nu))}}\,,$$

where

$$a_1^+ \,=\, a_1^- \,=\, \sqrt{\pi\sqrt{\nu-\xi_1^2}} \tag{1.6.32}$$

(cf. (1.6.25)). Substituting (1.6.32) into (1.6.15) we obtain for $\nu > \xi_1^2$

$$\arg_0 \det\bigl(i\,R(\nu;x_1,\xi_1)\bigr) \,=\, \pi/2 + 2\pi k \tag{1.6.33}$$

(cf. (1.6.26)) with an unknown integer k. Substituting (1.6.33) into (1.6.16) we establish that $k = 0$. Thus,

$$\arg_0 \det\bigl(i\,R(\nu;x_1,\xi_1)\bigr) \,=\, \begin{cases} 0\,, & \text{if} \quad \nu \le \xi_1^2\,, \\ \pi/2\,, & \text{if} \quad \nu > \xi_1^2\,, \end{cases} \tag{1.6.34}$$

which differs from (1.6.27) only in sign.

Finally, substituting (1.6.24) and (1.6.34) into (1.6.19) we obtain

$$c_1 \,=\, \frac{L}{4\pi}\,, \tag{1.6.35}$$

which also differs from (1.6.28) only in sign.

It is easy to see that the function (1.6.29) satisfies the boundary condition (1.6.31), so the threshold in Example 1.6.15 is soft. Accordingly, the sign of $\arg_0 \det\bigl(i\,R(\nu_1^{\mathrm{st}}+0;x_1,\xi_1)\bigr)$ is positive.

We are now prepared to explain the origin of the names "rigid" and "soft" with respect to normal thresholds. The spectral problem considered in Examples 1.6.14, 1.6.15 can be interpreted as the problem of free harmonic vibrations of a membrane. The Dirichlet boundary condition (1.6.21) (Example 1.6.14) describes the situation when the edge of the membrane is fixed ("rigid" boundary condition), and, accordingly, the threshold is called "rigid". The Neumann boundary condition (1.6.30) (Example 1.6.15) describes the situation when the edge of the membrane is free in the direction normal to the surface ("soft" boundary condition), and, accordingly, the stationary point is called "soft".

Example 1.6.16. Let M be an n-dimensional Riemannian manifold and $A = -\Delta$, where Δ is the Laplacian (see Example 1.2.4). We shall consider the Dirichlet $v|_{x_n} = 0$ or Neumann $\partial v/\partial x_n|_{x_n} = 0$ boundary conditions, where x_n is the geodesic distance to ∂M. Direct calculations along the lines of Examples 1.6.14, 1.6.15 show that

$$c_1 \,=\, \mp\,\frac{1}{4}\,(2\pi)^{1-n}\,\omega_{n-1}\,\operatorname{Meas}\partial M\,,$$

where the minus corresponds to the Dirichlet condition, the plus to the Neumann one, $\operatorname{Meas}\partial M$ is Riemannian $(n-1)$-dimensional volume of ∂M and ω_{n-1} is the volume of the unit ball in $\mathbb{R}^{n-1}$. Note that in deriving the above formula it is convenient to fix a point $x_0 = (x_0', 0) \in \partial M$, choose normal geodesic coordinates with origin x_0, and first integrate in (1.6.19) with respect to $đ\xi'$ only. The result will be a density on ∂M, so the integral with respect to dx' will be well defined and lead to the appearance of the factor $\operatorname{Meas}\partial M$.

Examples with nontrivial functions $\mathbf{N}^+(\nu; x', \xi')$ and $\arg_0 \det\bigl(i\,R(\nu; x', \xi')\bigr)$ will be considered in Section 6.2.

Let us now compare the structure of the terms $c_0\,\lambda^n$ and $c_1\,\lambda^{n-1}$ from the asymptotic expansion (1.6.1).

By (1.2.2) we have

$$c_0\,\lambda^n \;=\; \int\limits_{T'M} \mathbf{N}_A(\nu; x, \xi)\,dx\,đ\xi\,, \tag{1.6.36}$$

where

$$\mathbf{N}_A(\nu; x, \xi) \;=\; \begin{cases} 0\,, & \text{if} \quad \nu \le A_{2m}(x,\xi)\,, \\ 1\,, & \text{if} \quad \nu > A_{2m}(x,\xi) \end{cases} \tag{1.6.37}$$

(the subscript $_A$ in $\mathbf{N}_A$ indicates that this quantity depends only on the differential operator A, but not on the boundary conditions). For fixed $(x,\xi) \in T'M$ the function (1.6.37) is the counting function of the operator of multiplication by $A_{2m}(x,\xi)$ in $\mathbb{R}$. In other words, (1.6.37) is the counting function of the 1×1 matrix $A_{2m}(x,\xi)$. This trivial spectral problem

$$A_{2m}(x,\xi)v \;=\; \nu v \tag{1.6.38}$$

(where v is a real number, not a function) has only one eigenvalue $\nu = A_{2m}(x,\xi)$, and (1.6.37) is indeed the corresponding counting function.

With account of (1.6.19) and the rescaling properties of the functions $\mathbf{N}^+(\nu; x', \xi')$ and $\arg_0 \det\bigl(i\,R(\nu; x', \xi')\bigr)$ (cf. (1.6.17)) we have

$$c_1\,\lambda^{n-1} \;=\; \int\limits_{T^*\partial M} \left(\mathbf{N}^+(\nu; x', \xi') \;+\; \frac{\arg_0 \det\bigl(i\,R(\nu; x', \xi')\bigr)}{2\pi} \right) dx'\,đ\xi'\,. \tag{1.6.39}$$

Comparing formulae (1.6.36) and (1.6.39) we see that the terms $c_0\,\lambda^n$ and $c_1\,\lambda^{n-1}$ have roughly the same structure. In both cases integration is carried out over a cotangent bundle and the integrand contains the counting function ($\mathbf{N}_A$ or $\mathbf{N}^+$) of some elementary spectral problem. The analogy seems to be spoiled by the presence of the term $\arg_0 \det\bigl(i\,R(\nu; x', \xi')\bigr)$ in (1.6.39), but this is in fact quite natural: it is known that such a quantity plays the role of the counting function for the continuous spectrum.

It is interesting to note that formula (1.6.19) for the coefficient c_1 has a mechanical interpretation.

If for some $(x', \xi') \in T'\partial M$ the auxiliary one-dimensional spectral problem (1.1.7), (1.1.8) has an eigenvalue, then the original spectral problem (1.1.1), (1.1.2)

usually has a subsequence of eigenvalues corresponding to eigenfunctions which are localized in a small neighbourhood of ∂M ($x_n \lesssim \lambda^{-1}$, with exponential decay for $x_n \gg \lambda^{-1}$). Such eigenfunctions are well known in the theory of elasticity and are associated with the so-called Rayleigh surface waves. In scalar problems this boundary localization effect may also occur, see Section 6.2 in which we consider the two-dimensional biharmonic operator with free boundary conditions. The quantity

$$\int\limits_{T^*\partial M} \mathbf{N}^+(\nu; x', \xi')\, dx'\, đ\xi'$$

is the contribution of the subsequence of eigenvalues described above to the total counting function $N(\lambda)$.

The mechanical interpretation of the quantity

$$\int\limits_{T^*\partial M} \frac{\arg_0 \det\bigl(i\, R(\nu; x', \xi')\bigr)}{2\pi}\, dx'\, đ\xi' \tag{1.6.40}$$

is somewhat more complicated. It describes the influence of the boundary conditions on the "generic" eigenfunctions of the problem (1.1.1), (1.1.2), i.e., on those eigenfunctions which are not localized near the boundary. If the strong simple reflection condition is fulfilled then the integrand in (1.6.40) is expressed (see (1.6.15), (1.6.16)) through the phase shift induced by the reflection. The acquisition of a phase shift is equivalent to a slight displacement of the boundary, and (1.6.40) can be viewed as the resulting correction to the first term of the asymptotic expansion of $N(\lambda)$.

5. Pseudodifferential case. In subsection 1.1.8 we stated that apart from the main spectral problem (1.1.1), (1.1.2) in which A is the differential operator, we are also interested in the case when A is a pseudodifferential operator acting on a manifold without boundary. In order to honour this commitment we state below a modified version of Theorem 1.6.1, which is due to J.J. Duistermaat and V.W. Guillemin [DuiGui], [DuiGuiHö].

THEOREM 1.6.1′. *Let $\partial M = \varnothing$ and A be a pseudodifferential operator. If the nonperiodicity condition is fulfilled, then (1.6.1) holds. Here c_0 is the same as in Theorem 1.2.1, and*

$$\begin{aligned} c_1 \;&=\; -\frac{1}{2m} \int\limits_{S^*M} A_{\mathrm{sub}}\,(x, \widetilde{\xi})\, dx\, đ\widetilde{\xi} \\ &=\; -\left(1 + \frac{n-1}{2m}\right) \int\limits_{A_{2m} \le 1} A_{\mathrm{sub}}\,(x, \xi)\, dx\, đ\xi\,. \end{aligned} \tag{1.6.41}$$

Note that for differential operators of even order the subprincipal symbol is an odd function with respect to ξ , $A_{\mathrm{sub}}\,(x, -\xi) = -A_{\mathrm{sub}}\,(x, \xi)$, so for such operators the integral (1.6.41) is zero. This explains why formula (1.6.2) does not contain the integral (1.6.41) as an additional term.

1.7. Nonclassical two-term asymptotic formulae for $N(\lambda)$

1. Preliminary discussion. In this section we turn to the case when the measure of the set Π is nonzero. One can expect that in this situation the two-term asymptotic formula contains an extra term of order $O(\lambda^{n-1})$ which reflects the influence of the periodic trajectories. The most natural conjecture is that

$$N(\lambda) \;=\; c_0\,\lambda^n \;+\; c_1\,\lambda^{n-1} \;+\; \widetilde{\mathbf{Q}}(\lambda)\,\lambda^{n-1} \;+\; o(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \tag{1.7.1}$$

where c_0 and c_1 are the same constants as in (1.4.1), and $\widetilde{\mathbf{Q}}(\lambda)$ is a bounded function.

Of course, a function $\widetilde{\mathbf{Q}}$ satisfying (1.7.1) always exists; for example, we can take any function $\widetilde{\mathbf{Q}}$ of the form

$$\widetilde{\mathbf{Q}}(\lambda) \;=\; \lambda^{1-n}\,N(\lambda) \;-\; c_0\,\lambda \;-\; c_1 \;+\; o(1)\,, \qquad \lambda \to +\infty\,.$$

But this observation is of no use unless we can say something more definite about $\widetilde{\mathbf{Q}}$. In particular, we would like this function to depend only on the higher order terms of the differential operator A and the boundary operators $B^{(k)}$.

However, Example 1.7.1 (see below) shows that, generally speaking, (1.7.1) can not hold with a $\widetilde{\mathbf{Q}}$ independent of the lower order terms. Since in the general case the lower order terms are not invariantly defined, it seems to be impossible to obtain any general results concerning $\widetilde{\mathbf{Q}}$. Therefore we adopt another approach. We shall prove that there exists a bounded function $\mathbf{Q}$ depending only on the principal and subprincipal symbols of A and on the principal symbols of $B^{(k)}$, such that

$$\mathbf{Q}(\lambda+\varepsilon) \;-\; \mathbf{Q}(\lambda) \;\geq\; -\varepsilon\,n\,c_0\,, \qquad \forall \lambda \in \mathbb{R}\,, \quad \forall \varepsilon \geq 0\,, \tag{1.7.2}$$

and

$$\begin{aligned} &c_0\,\lambda^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}\,(\lambda - o(1))\;\lambda^{n-1} \;-\; o(\lambda^{n-1}) \;\leq\; N(\lambda) \\ &\qquad \leq\; c_0\,\lambda^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}\,(\lambda + o(1))\;\lambda^{n-1} \;+\; o(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \end{aligned} \tag{1.7.3}$$

where $o(1)$ is some positive function tending to zero as $\lambda \to +\infty$ and $o(\lambda^{n-1}) = \lambda^{n-1}o(1)$. In view of (1.7.2)

$$\begin{aligned} &c_0\,\lambda^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}(\lambda)\,\lambda^{n-1} \\ &\qquad \leq\; c_0\,(\lambda+\varepsilon)^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}(\lambda+\varepsilon)\,\lambda^{n-1} \;+\; o(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \end{aligned}$$

for any fixed $\varepsilon \geq 0$, so (1.7.3) makes sense.

The asymptotic estimate (1.7.3) implies that the "distance" between the graphs of the functions

$$(N(\lambda))^{1/n} \quad \text{and} \quad \bigl(c_0\lambda^n + c_1\lambda^{n-1} + \mathbf{Q}(\lambda)\lambda^{n-1}\bigr)^{1/n}$$

tends to zero as $\lambda \to +\infty$. The asymptotic formula (1.7.1) would mean that the distance between these graphs in the vertical direction tends to zero, whereas (1.7.3) allows also an error $o(1)$ in the horizontal direction.

In other words, we shall prove that for all $\varepsilon > 0$

$$(1.7.4)\quad \begin{aligned} & c_0\,(\lambda-\varepsilon)^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}(\lambda-\varepsilon)\,\lambda^{n-1} \;-\; o(\lambda^{n-1}) \;\le N(\lambda) \\ & \le\; c_0\,(\lambda+\varepsilon)^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}(\lambda+\varepsilon)\,\lambda^{n-1} \;+\; o(\lambda^{n-1})\,, \qquad \lambda\to+\infty\,, \end{aligned}$$

where, generally speaking, $o(\lambda^{n-1})$ depends on ε. Obviously, (1.7.4) is equivalent to (1.7.3).

If the function $\mathbf{Q}$ is uniformly continuous then we obtain from (1.7.4)

$$(1.7.5)\qquad N(\lambda) \;=\; c_0\,\lambda^n \;+\; c_1\,\lambda^{n-1} \;+\; \mathbf{Q}(\lambda)\,\lambda^{n-1} \;+\; o(\lambda^{n-1})\,, \qquad \lambda\to+\infty\,.$$

We shall call (1.7.5) with a uniformly continuous $\mathbf{Q}$ the *quasi-Weyl* asymptotic formula.

If

$$(1.7.6)\qquad \mathbf{Q}(\mu_j+\varepsilon_j) \;-\; \mathbf{Q}(\mu_j-\varepsilon_j) \;\ge\; c\,, \qquad j=1,2,\dots,$$

where $\mu_j \to +\infty$, $\varepsilon_j \to +0$ are some positive sequences and c is a positive constant, then (1.7.4) implies

$$(1.7.7)\qquad N(\mu_j+\varepsilon_j) \;-\; N(\mu_j-\varepsilon_j) \;\ge\; c\,\mu_j^{n-1} \;+\; o(\mu_j^{n-1})\,.$$

This means that the spectrum of the operator $\mathcal{A}^{1/(2m)}$ contains contracting groups of eigenvalues lying in the ε_j-neighbourhoods of the points μ_j, whose total multiplicities are estimated from below by const μ_j^{n-1}. Such groups of eigenvalues are called *clusters*. The precise location of eigenvalues inside the clusters normally depends on the lower order terms of the operators A and $B^{(k)}$. Therefore in the case of cluster asymptotics one can not expect to have (1.7.5) with a function $\mathbf{Q}$ depending only on the higher order terms.

If the quasi-Weyl formula (1.7.5) holds then there are no clusters in the spectrum. Indeed, since $\mathbf{Q}$ is uniformly continuous,

$$N(\mu_j+\varepsilon_j) \;-\; N(\mu_j-\varepsilon_j) \;=\; o(\mu_j^{n-1})$$

for all sequences $\mu_j \to +\infty$ and $\varepsilon_j \to 0$.

Example 1.7.1. Let

$$(1.7.8)\qquad A \;=\; (-\Delta + (n-1)^2/4)^m \;+\; V\,,$$

where Δ is the Laplacian on the unit sphere and V is the operator of multiplication by a smooth nonnegative function. Recall that the spectrum of the operator $(-\Delta+(n-1)^2/4)^{1/2}$ consists of the eigenvalues

$$\Lambda_j \;=\; j+(n-1)/2\,, \qquad j=0,1,2,\dots,$$

with multiplicities

$$\frac{(n+j-2)!\,(n+2j-1)}{(n-1)!\,j!} \;\ge\; \text{const}\;\Lambda_j^{n-1}$$

(see Example 1.2.5).

We shall see later that for the operator (1.7.8) formula (1.7.4) holds with a periodic function $\mathbf{Q}$ which has jumps at the points $j+(n-1)/2$ (Example 1.7.11). Such a function satisfies (1.7.6) with $\mu_j = \Lambda_j$, so we have clusters around the points Λ_j. This result can be also obtained by means of perturbation theory: if we add the lower order operator V to the operator $(-\Delta+(n-1)^2/4)^m$ then the multiple eigenvalues Λ_j of $(-\Delta+(n-1)^2/4)^m$ turn into contracting groups of eigenvalues with the same total multiplicities.

For a general potential V the quasi-Weyl formula (1.7.5) is not true. However, the difference

$$N(\lambda) \;-\; c_0\,\lambda^n \;-\; c_1\,\lambda^{n-1} \;-\; \mathbf{Q}(\lambda)\,\lambda^{n-1}$$

is zero outside ε_j-neighbourhoods of the points $j+(n-1)/2$, for some $\varepsilon_j \to 0$. Note that there is no function $\widetilde{\mathbf{Q}}$ depending only on the principal and subprincipal symbols which satisfies (1.7.1) for all potentials V. Indeed, if $V \equiv V_0 = \text{const} > 0$ then the spectrum of $\mathcal{A}^{1/(2m)}$ consists of the eigenvalues

$$\Big((j+(n-1)/2)^{2m} + V_0\Big)^{1/(2m)}, \qquad j = 0,1,2,\dots,$$

with the same multiplicities as Λ_j. Since these eigenvalues depend on the constant V_0, the function $\widetilde{\mathbf{Q}}$ in (1.7.1) must also depend on V_0.

2. The case of simple reflection. Assume that the Hamiltonian billiard system satisfies the simple reflection condition (see Definition 1.3.30).

Definition 1.7.2. For a T-admissible trajectory

$$\Gamma \;=\; (x^*(t;y,\eta),\xi^*(t;y,\eta))\,, \qquad 0 \le t \le T\,,$$

let

$\mathfrak{f}_r(T;y,\eta)$ be the total phase shift generated by the reflections of the trajectory Γ, that is, the sum of the phase shifts generated by the reflections of Γ (see end of Section 1.4);

$\mathfrak{f}_s(T;y,\eta) = -\,(2m)^{-1}\,h^{1-2m}(y,\eta)\,\int_0^T A_{\text{sub}}\big(x^*(t;y,\eta),\xi^*(t;y,\eta)\big)\,dt$ (this quantity is interpreted as the phase shift generated by the subprincipal symbol).

When $x_\eta(T;y,\eta)=0$ we also define

$\mathfrak{f}_c(T;y,\eta)$ to be the phase shift generated by the passage of the trajectory Γ through caustics, that is, $\mathfrak{f}_c(T;y,\eta) = -\alpha_\Gamma\pi/2$ where α_Γ is the Maslov index of Γ (see Section 1.5).

The quantity

$$\mathfrak{f}(T;y,\eta) \;=\; \mathfrak{f}_r(T;y,\eta) \;+\; \mathfrak{f}_s(T;y,\eta) \;+\; \mathfrak{f}_c(T;y,\eta)$$

is said to be the *total phase shift* along the trajectory Γ.

The total phase shift is an additive function in the sense that

$$\mathfrak{f}(T_1+T_2;y,\eta) \;=\; \mathfrak{f}(T_1;y,\eta) \;+\; \mathfrak{f}\,(T_2;x^*(T_1;y,\eta),\xi^*(T_1;y,\eta))\,,$$

and the same is valid for $\mathfrak{f}_r$, $\mathfrak{f}_s$ and $\mathfrak{f}_c$.

DEFINITION 1.7.3. For a periodic point (y,η) we denote by $\mathbf{T}(y,\eta)$ the corresponding minimal positive period. If, in addition, (y,η) is absolutely $\mathbf{T}(y,\eta)$-periodic and admissible, then we set $\mathbf{q}(y,\eta) = \mathfrak{f}(\mathbf{T}(y,\eta);y,\eta)$.

Obviously, $\mathbf{q}(y,\eta)$ is the total phase shift along the primitive closed trajectory originating from (y,η). In view of Lemmas 1.3.5 and 1.3.28 the function $\mathbf{q}$ is defined almost everywhere (a.e.) on Π^a, as well as on Π.

LEMMA 1.7.4. *There exists a constant $C>0$ such that $\mathbf{T}(y,\widetilde{\eta}) \ge C$ for all $(y,\widetilde{\eta}) \in \Pi$.*

LEMMA 1.7.5. *The functions $\mathbf{T}$ and $\mathbf{q}$ are measurable.*

Let us denote by $\{\tau\}_{2\pi}$ the residue of the real number τ modulo 2π, that is,

$$\{\tau\}_{2\pi} \;:=\; \tau \,+\, 2\pi k \;\in\; [-\pi,\pi)\,, \qquad k\in\mathbb{Z}\,.$$

Clearly, $\{\tau\}_{2\pi} = 2\pi\{\tau/(2\pi)+1/2\} - \pi$, where $\{\,\cdot\,\}$ is the fractional part.

THEOREM 1.7.6 ([Sa4], [Sa5], [Sa7]). *If the Hamiltonian billiard system satisfies the simple reflection condition then (1.7.4) (and (1.7.3)) hold with the function*

$$\mathbf{Q}(\lambda) \;=\; \int_{\Pi^a} \frac{\{\pi - \mathbf{q} - \lambda\mathbf{T}\}_{2\pi}}{\mathbf{T}}\; dy\, đ\widetilde{\eta}\,. \tag{1.7.9}$$

Note that for each fixed $(y,\eta)\in T^*M$ we have

$$\{\pi - \mathbf{q} - \lambda\mathbf{T}\}_{2\pi} \;=\; 2\sum_{k=1}^{\infty} k^{-1}\,\sin k(\lambda\mathbf{T}+\mathbf{q}) \tag{1.7.10}$$

where the series converges in the sense of distributions. From (1.7.9) and (1.7.10) one can easily deduce the following elementary properties of the function $\mathbf{Q}$:

(1) $\mathbf{Q}$ is uniformly bounded, $|\,\mathbf{Q}(\lambda)\,| \le \pi \int_{\Pi^a} \mathbf{T}^{-1}\, dy\, đ\widetilde{\eta}$;
(2) $\mathbf{Q}$ is an oscillating function, i.e., the integral $\int_0^\lambda \mathbf{Q}(\mu)\,d\mu$ is bounded uniformly with respect to $\lambda\in\mathbb{R}$;
(3) $\mathbf{Q}$ is left-continuous;
(4) for all $\varepsilon>0$ we have $\mathbf{Q}(\lambda+\varepsilon) - \mathbf{Q}(\lambda) \ge -\varepsilon(2\pi)^{-n}\,\mathrm{meas}\,\Pi^a$;
(5) $\mathbf{Q}(\lambda+0) - \mathbf{Q}(\lambda) = 2\pi\int_{\Pi^a(\lambda)} \mathbf{T}^{-1}\, dy\, đ\widetilde{\eta}$, where

$$\Pi^a(\lambda) \;:=\; \{\,(y,\eta)\in\Pi^a \,:\, \mathbf{q}(y,\eta) + \lambda\mathbf{T}(y,\eta) = 0 \pmod{2\pi}\,\};$$

(6) if $\mathbf{T} = T = \mathrm{const}$ a.e. on Π^a then $\mathbf{Q}$ is $2\pi T^{-1}$-periodic.

In view of (1.2.2′), (4) implies (1.7.2). Moreover, from Theorem 1.7.6 and (4) it follows that

$$N(\lambda+\varepsilon) \;-\; N(\lambda) \;\ge\; \varepsilon\,(2\pi)^{-n}\,\mathrm{meas}\,(S^*M\setminus\Pi^a)\,\lambda^{n-1} \;+\; o(\lambda^{n-1})\,.$$

Therefore we obtain the following

COROLLARY 1.7.7. *Let the conditions of Theorem 1.7.6 be fulfilled. Assume that there exists a sequence of points* $\mu_j \to +\infty$ *and an* $\varepsilon > 0$ *such that the number of eigenvalues lying in the* ε*-neighbourhood of* μ_j *is* $o(\mu_j^{n-1})$. *Then* $\operatorname{meas} \Pi^a = \operatorname{meas} S^*M$.

The condition $\mathbf{T} = T = \text{const}$ in (6) seems to be very restrictive. However, the following lemma shows that it is always satisfied when M and h are analytic and the set Π^a is connected.

LEMMA 1.7.8. *In the analytic case* $\mathbf{T} \equiv \text{const}$ *almost everywhere on any open connected subset of* Π^a.

Assume that (6) is fulfilled. Then, since a periodic function is uniformly continuous if and only if it is continuous, Theorem 1.7.6 and (5) immediately imply

COROLLARY 1.7.9. *Let the conditions of Theorem 1.7.6 be fulfilled and* $\mathbf{T} = T = \text{const}$ *a.e. on* Π^a. *Then the quasi-Weyl formula (1.7.5) holds with the function* $\mathbf{Q}$ *given by (1.7.9) if and only if*

$$\operatorname{meas}\{\,(y,\eta) \in \Pi^a \,:\, \mathbf{q}(y,\eta) = \mu\,\} = 0\ , \qquad \forall \mu \in \mathbb{R}\,.$$

If

$$\operatorname{meas}\{\,(y,\eta) \in \Pi^a \,:\, \mathbf{q}(y,\eta) = q\,\} = C_q > 0 \tag{1.7.11}$$

for some constants q and C_q then, in view of (5) and (6),

$$\mathbf{Q}\left(T^{-1}(q+2\pi j)+0\right) - \mathbf{Q}\left(T^{-1}(q+2\pi j)\right) = (2\pi)^{1-n}\,T^{-1}\,C_q\,, \qquad j = 1,2,\dots$$

This implies

COROLLARY 1.7.10. *Let the simple reflection condition be fulfilled,* $\mathbf{T} = T = \text{const}$ *a.e. on* Π^a, *and (1.7.11) hold for some* q. *Then we have (1.7.7) with* $c = (2\pi)^{1-n}\,T^{-1}\,C_q$, $\mu_j = T^{-1}(q+2\pi j)$ *and some* $\varepsilon_j \to 0$, *i.e., the spectrum contains clusters.*

Thus, if $\mathbf{T} = T = \text{const}$ a.e. on Π^a then either the quasi-Weyl formula holds or the spectrum contains clusters. If, modulo a set of measure zero, Π^a consists of several open connected components then $\mathbf{Q}$ is the sum of functions corresponding to these components. Each of these functions either generates clusters or gives a contribution to the quasi-Weyl part of the asymptotics.

EXAMPLE 1.7.11. For the operator (1.7.8) we have $\Pi^a = S^*M$, $\mathbf{T} \equiv 2\pi$, $\mathbf{q} = \mathbf{q}_c$, and $\mathbf{q}_c \equiv \pi(n-1) \pmod{2\pi}$. Therefore

$$\mathbf{Q}(\lambda) = (2\pi)^{-n} \operatorname{meas} S^*M\ \frac{\{\pi - \pi(n-1) - 2\pi\,\lambda\}_{2\pi}}{2\pi} = n\,c_0\ (\{n/2-\lambda\} - 1/2)\ ,$$

where $\{n/2-\lambda\}$ is the fractional part of $n/2-\lambda$. The constant c_0 is equal to $c_0 = 2/n!$ (see Example 1.2.5), and $c_1 = 0$ since $\partial M = \varnothing$.

EXAMPLE 1.7.12. Let M be an n-dimensional unit sphere and $A = -\Delta + B$ where B is a self-adjoint first order differential operator. Again, we have $\Pi^a = S^*M$, $\mathbf{T} \equiv 2\pi$, and $\mathbf{q}_c \equiv \pi(n-1) \pmod{2\pi}$. However, now $\mathbf{q} = \mathbf{q}_c + \mathbf{q}_s$, where

$$\mathbf{q}_s(y,\eta) \;=\; -\frac{1}{2h}\int B_1\, ds\,,$$

B_1 is the principal symbol of B, and the integral is taken over the primitive periodic trajectory originating from (y,η). In this case the type of asymptotics is determined by the principal symbol B_1. It can even happen that $\mathbf{Q} \equiv 0$, and then we have the classical Weyl formula for $N(\lambda)$ (with the same constants as in Example 1.7.11) though all the trajectories are periodic. For instance, $\mathbf{Q}$ is identically zero if M is a unit three-dimensional sphere embedded in $\mathbb{R}^4$ and $B = \mathcal{B} + \mathcal{B}^*$, where $\mathcal{B}$ is the first order differential operator generated by the vector field with components $(-x_2, x_1, -x_4, x_3)$ (this vector field is tangent to the sphere); see [Sa2], [Sa8].

EXAMPLE 1.7.13. Let M be a two-dimensional hemisphere of radius R, and A be the biharmonic operator Δ^2 on M with Dirichlet boundary condition. Then

$$c_0 \;=\; R^2/2\,, \qquad c_1 \;=\; -\frac{R}{2}\left(1 \,+\, \pi^{-1/2}\,\frac{\Gamma(3/4)}{\Gamma(5/4)}\right),$$

$$\mathbf{Q}(\lambda) \;=\; -\,R\left(\{R\lambda\} \,+\, (\cos\pi\{R\lambda\})_+^{1/2} \,-\, \frac{\mathbf{K}(\sqrt{2}/2)+\sqrt{2}}{2\,\mathbf{K}(\sqrt{2}/2)}\right),$$

where $\Gamma(\cdot)$ is the Gamma function, $\{R\lambda\}$ is the fractional part of $R\lambda$, and $\mathbf{K}(\cdot)$ is the complete elliptic integral of the first type. Note that the coefficient c_1 coincides with that for a plate with the same length of ∂M, see Section 6.2. The function $\mathbf{Q}$ is R^{-1}-periodic and continuous, so the quasi-Weyl formula holds. The graph of $\mathbf{Q}$ is given below.

Figure 10. The function $\mathbf{Q}$ for the Dirichlet bi-Laplacian on a hemisphere.

Elementary analysis of the formula for $\mathbf{Q}$ shows that the number of eigenvalues λ_k lying in the intervals $[(j-1/2+\varepsilon)/R\,,(j-\varepsilon)/R]$, $0<\varepsilon<1/4$, $j=1,2,\dots,$ is $o(j)$ as $j\to\infty$. This means that asymptotically the spectrum contains gaps. Another elementary observation is that the left derivative of $\mathbf{Q}$ is infinite at the points $(j-1/2)/R$, so one expects the density of eigenvalues λ_k around these points to be abnormally high.

3. The general case. A similar result is true in the general case of a branching Hamiltonian billiards.

THEOREM 1.7.14. *Let the nonblocking condition (Definition 1.3.22) be fulfilled. Assume, in addition, that there exist positive numbers* $T_1, T_2, \dots$ *tending to* $+\infty$ *such that* $\operatorname{meas}\left(\bigcup_{T\neq 0,\, T\neq T_j} \Pi_T^a\right) = 0$. *Then (1.7.4) holds with a bounded function* $\mathbf{Q}$ *depending only on the principal and subprincipal symbols of* A *and on the principal symbols of* $B^{(k)}$, *and satisfying (1.7.2). The function* $\mathbf{Q}$ *is left-continuous and almost periodic with periods* $T_1, T_2, \dots$

In the general case there is no simple formula for the function $\mathbf{Q}$. One can only prove that for positive λ this function is given by a trigonometric series

$$\mathbf{Q}(\lambda) \;=\; \sum_{j=1}^{\infty} C_j \sin \lambda T_j \,, \tag{1.7.12}$$

where C_j are constants depending on some characteristics of the corresponding periodic trajectories.

Theorem 1.7.14 is obtained in the same manner as Theorem 1.7.6, and therefore we shall not prove it in this book. A rigorous proof of (1.7.12) and formulae for C_j are given in [Sa5] where even a more general situation (with A being a system of partial differential operators) is examined.

4. Pseudodifferential case. Theorem 1.7.6 remains true for a pseudodifferential operator acting on a manifold without boundary [Sa4]. The only difference is that in this case the constant c_1 is given by formula (1.6.41).

1.8. Two-term asymptotic formulae for the spectral function

1. The spectral function. The *spectral function* of the operator $\mathcal{A}$ is defined by

$$e(\lambda, x, y) = \sum_{\lambda_k < \lambda} v_k(x)\, \overline{v_k(y)} \,,$$

where $\lambda_k, v_k(x)$, $k = 1, 2, \dots$, are the eigenvalues and the orthonormalized eigenfunctions of the problem (1.1.1), (1.1.2). Clearly, $e(\lambda, x, y)$ is a smooth half-density on $M \times M$ depending on the spectral parameter λ, and $e(\lambda, x, y) \equiv 0$ if $\lambda \le 0$.

The operator

$$E_\lambda v \;=\; \sum_{\lambda_k < \lambda} (v, v_k)\, v_k \,,$$

where $(\cdot\,,\cdot)$ stands for the inner product (1.1.10), is called the *spectral projection* of the problem (1.1.1), (1.1.2) (or of the corresponding self-adjoint operator $\mathcal{A}^{1/(2m)}$). The spectral function is the integral kernel of the spectral projection:

$$E_\lambda v \;=\; \int\limits_M e(\lambda, x, y)\, v(y)\, dy \,.$$

Clearly, $e(\lambda, y, y)$ is a density on M depending on λ. Integrating over M we obtain

$$N(\lambda) \;=\; \operatorname{Tr} E_\lambda \;=\; \int\limits_M e(\lambda, y, y)\, dy \,. \tag{1.8.1}$$

Thus, $e(\lambda, y, y)$ contains full information about the spectrum.

In this section we shall discuss the asymptotic behaviour of $e(\lambda, y, y)$ as $\lambda \to +\infty$ at a fixed interior point y. More generally, we shall consider the functions

$$e_{P,Q}(\lambda, x, y) \;=\; \sum_{\lambda_k < \lambda} Q^* v_k(x)\, \overline{P^* v_k(y)}\,,$$

where P, Q are pseudodifferential operators of the class $\Psi_0^{\mathbf{l}}$, $2\mathbf{l} > 1-n$. The principal and subprincipal symbols of P and Q are denoted by $P_{\mathbf{l}}, Q_{\mathbf{l}}$ and $P_{\mathrm{sub}}, Q_{\mathrm{sub}}$ respectively. (See Section 2.1 for definitions of pseudodifferential operators and their symbols.)

Obviously, $e_{P,Q}(\lambda, x, y)$ is the integral kernel of the operator $Q^* E_\lambda P$. We have

$$\begin{aligned} e_{P,Q}(\lambda, x, y) \;=\; & \frac{1}{4}\, e_{P+Q,P+Q}(\lambda, x, y) \;-\; \frac{1}{4}\, e_{P-Q,P-Q}(\lambda, x, y) \\ & +\; \frac{i}{4}\, e_{P+iQ,P+iQ}(\lambda, x, y) \;-\; \frac{i}{4}\, e_{P-iQ,P-iQ}(\lambda, x, y) \end{aligned}$$

(the *polarization* formula). Applying polarization, one can easily deduce the results for $e_{P,Q}(\lambda, x, y)$ from those for $e_{P,P}(\lambda, x, y)$.

2. Notation and definitions. Throughout this section we shall always assume that the Hamiltonian billiard system satisfies the simple reflection condition (see Definition 1.3.30).

Let us denote

$$\Pi_{y,T} \;=\; \{\, \widetilde{\eta} \in S_y^* M \,:\, x^*(T; y, \widetilde{\eta}) = y \,\}\,, \qquad \Pi_y \;=\; \bigcup_{T>0} \Pi_{y,T}\,.$$

Let $\Pi_{y,T}^a \subset \Pi_{y,T}$ be the set of $\widetilde{\eta}$ such that

$$\partial_\eta^\alpha \left(|y - x^*(T; y, \eta)|^2 \right)\big|_{\eta=\widetilde{\eta}} = \; 0\,, \qquad \forall \alpha\,,$$

and $\Pi_y^a = \cup_{T>0} \Pi_{y,T}^a$.

Definition 1.8.1. A point $y \in \overset{\circ}{M}$ is said to be *regular* if for almost all $\widetilde{\eta} \in S_y^* M$ the points $(y, \widetilde{\eta}) \in S^* M$ are admissible.

Lemma 1.8.2. *If y is a regular point then the sets Π_y and $\bigcup\limits_{0<T\le T_+} \Pi_{y,T}$, $\forall T_+ > 0$, are measurable.*

Lemma 1.8.3. *If y is a regular point then* $\mathrm{meas}_y(\Pi_y \setminus \Pi_y^a) = 0$. *Moreover,* $\mathrm{meas} \cup_{T>0} \left(\Pi_{y,T} \setminus \Pi_{y,T}^a\right) = 0$.

Definition 1.8.4. A regular point $y \in \overset{\circ}{M}$ is said to be *focal* if $\mathrm{meas}_y \Pi_y^a > 0$.

3. Classical asymptotics. The following well known result is an analogue of Theorem 1.2.1.

THEOREM 1.8.5. *If* $y \in \mathring{M}$ *then*

$$e_{P,Q}(\lambda, y, y) \;=\; c_{0;P,Q}(y)\, \lambda^{n+2\mathbf{l}} \;+\; O(\lambda^{n+2\mathbf{l}-1})\,, \qquad \lambda \to +\infty\,, \tag{1.8.2}$$

where

$$c_{0;P,Q}(y) \;=\; (n+2\mathbf{l})^{-1} \int_{S^*_y M} P_{\mathbf{l}}(y,\widetilde{\eta})\, \overline{Q_{\mathbf{l}}(y,\widetilde{\eta})}\, đ\widetilde{\eta}\,. \tag{1.8.3}$$

The asymptotic formula (1.8.2) is uniform on compact subsets of $\mathring{M}$*.*

Note that $e_{P,Q}(\lambda, y, y)$ is a density depending on λ, and the coefficient $c_{0;P,Q}(y)$ also behaves as a density on M. If P and Q are operators of multiplication by some functions $\widetilde{P}, \widetilde{Q} \in C_0^\infty(\mathring{M})$ and $\widetilde{P}(y) = \widetilde{Q}(y) = 1$, then $e_{P,Q}(\lambda, y, y) = e(\lambda, y, y)$ and (1.8.2) takes the form

$$e(\lambda, y, y) \;=\; c_0(y)\, \lambda^n \;+\; O(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \tag{1.8.4}$$

where

$$c_0(y) \;=\; (2\pi)^{-n}\, \mathrm{vol}_y\, \{\eta : A_{2m}(y,\eta) \le 1\} \;=\; \int\limits_{A_{2m}(y,\eta)\le 1} đ\eta\,.$$

If M is a manifold without boundary then Theorem 1.2.1 is obtained by integrating (1.8.4) over M.

REMARK 1.8.6. By (1.8.2), if $\operatorname{ord} \widetilde{P} > \operatorname{ord} P$ and the principal symbols of $\widetilde{P}$ and P do not vanish at y, then

$$\frac{\sum_{\lambda_k<\lambda} |\widetilde{P} v_k(y)|^2}{\sum_{\lambda_k<\lambda} |P v_k(y)|^2} \;\to\; +\infty\,, \qquad \lambda \to +\infty\,.$$

This means that, in a sense, the eigenfunctions v_k become more and more oscillating as $k \to \infty$.

THEOREM 1.8.7 ([Sa6]). *Let* $y \in \mathring{M}$ *be a regular non-focal point. Then*

$$e_{P,P}(\lambda, y, y) \;=\; c_{0;P}(y)\, \lambda^{n+2\mathbf{l}} \;+\; c_{1;P}(y)\, \lambda^{n+2\mathbf{l}-1} \;+\; o(\lambda^{n+2\mathbf{l}-1}) \tag{1.8.5}$$

as $\lambda \to +\infty$*, where* $c_{0;P}(y) = c_{0;P,P}(y)$ *and*

$$\begin{aligned} c_{1;P}(y) \;=\; (n+2\mathbf{l}-1)^{-1} &\int_{S^*_y M} \left(2\,\mathrm{Re}(P_{\mathbf{l}}\, \overline{P_{\mathrm{sub}}}) + \frac{\mathrm{Im}\{\overline{P_{\mathbf{l}}}, P_{\mathbf{l}}\}}{2} \right) đ\widetilde{\eta} \\ &+ \int_{S^*_y M} \left(\mathrm{Im}(P_{\mathbf{l}}\{h, \overline{P_{\mathbf{l}}}\}) - \frac{A_{\mathrm{sub}}\, |P_{\mathbf{l}}|^2}{2m} \right) đ\widetilde{\eta}\,. \end{aligned} \tag{1.8.6}$$

The asymptotic formula (1.8.5) is uniform on compact subsets of $\mathring{M}$ *which do not contain any focal or non-regular points.*

Assuming that P is the operator of multiplication by a C_0^∞-function which is equal to one in a neighbourhood of y, we obtain

$$e(\lambda, y, y) \;=\; c_0(y)\, \lambda^n \;+\; o(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,.$$

Indeed, in this case $P_{\mathrm{sub}} = 0$ at the point y and the second term in the right-hand side of (1.8.6) disappears because A_{sub} is an odd function of $\widetilde{\eta}$. If M is a manifold without boundary and all the points $y \in M$ are non-focal then the last formula implies (1.6.1) (with $c_1 = 0$).

4. Non-classical asymptotics. Given $\widetilde{\eta} \in \Pi_y$ we denote

$$\mathbf{T}_y(\widetilde{\eta}) \;=\; \min\{\, t>0 \,:\, x^*(t;y,\widetilde{\eta}) = y \,\}\,.$$

If, in addition, $\widetilde{\eta} \in \Pi^a_{y,\mathbf{T}_y(\widetilde{\eta})}$ and $(y,\widetilde{\eta})$ is admissible, then we set

$$\mathbf{q}_y(\widetilde{\eta}) \;=\; \mathfrak{f}\,(\mathbf{T}_y(\widetilde{\eta});y,\widetilde{\eta})\;,$$

where $\mathfrak{f}$ is the total phase shift from Definition 1.7.2. In the case of a geodesic flow $\mathbf{T}_y(\widetilde{\eta})$ is the length of the minimal geodesic loop originating from the point y and going in the direction $\widetilde{\eta}$, and $\mathbf{q}_y(\widetilde{\eta})$ is the total phase shift along this loop.

Further on in this section we assume that the point y is regular. Then Definition 1.8.1 and Lemma 1.8.3 imply that the function $\mathbf{q}_y$ is defined a.e. on Π^a_y, as well as on Π_y.

Let us define a map $\Phi_y : \Pi^a_y \to S^*_yM$ by

$$\Phi_y\widetilde{\eta} \;=\; \xi^*\,(\mathbf{T}_y(\widetilde{\eta});y,\widetilde{\eta})\;,$$

and let

$$J_y(\widetilde{\eta}) \;=\; |\,\det \xi^*_\eta\,(\mathbf{T}_y(\widetilde{\eta});y,\widetilde{\eta})\,|\,.$$

If $x^*_\eta(t;y,\eta)=0$ then $|\det \xi^*_\eta(t;y,\eta)| \neq 0$ behaves under change of coordinates as a density in x and a density to the power -1 in y. Therefore the restriction of J_y to Π^a_y is independent of the choice of coordinates y, assuming that we take the same coordinates for x and for y. Moreover, since the shift along billiard trajectories is a non-degenerate map and $x^*_\eta = 0$ on Π^a_y, we have $J_y \geq \text{const} > 0$ uniformly on Π^a_y.

We shall need the following two lemmas.

LEMMA 1.8.8. *The functions $\mathbf{T}_y$ and $\mathbf{q}_y$ are measurable.*

LEMMA 1.8.9. *For any measurable set $\Omega \subset \Pi^a_y$ the set $\Phi_y\Omega \subset S^*_yM$ is also measurable and*

$$\text{meas}_y(\Phi_y\Omega) \;=\; \int_\Omega J_y(\widetilde{\eta})\,d\widetilde{\eta}\,. \tag{1.8.7}$$

In other words, Lemma 1.8.9 states that J_y is the Radon–Nikodym derivative of the measure $\text{meas}_y(\Phi_y\,\cdot\,)$ with respect to meas_y.

Let $U_y : L_2(S^*_yM) \to L_2(S^*_yM)$ and $U_{y,\lambda} : L_2(S^*_yM) \to L_2(S^*_yM)$ be the linear operators defined as follows:

$$(U_yf)\,(\widetilde{\eta}) \;=\; \begin{cases} 0\,, & \text{if } \widetilde{\eta} \notin \Pi^a_y\,,\\ e^{-i\mathbf{q}_y(\widetilde{\eta})}\,\sqrt{J_y(\widetilde{\eta})}\,f\,(\Phi_y\widetilde{\eta})\;, & \text{if } \widetilde{\eta} \in \Pi^a_y\,, \end{cases}$$

$$(U_{y,\lambda}f)\,(\widetilde{\eta}) \;=\; e^{-i\lambda\mathbf{T}_y(\widetilde{\eta})}\,(U_yf)\,(\widetilde{\eta})$$

(here λ is considered as a parameter). By Lemma 1.8.8 the operators U_y and $U_{y,\lambda}$ are partially isometric with kernel

$$\{f \in L_2(S^*_yM) \,:\, \text{supp}\, f \cap \Phi_y\Pi^a_y = \varnothing\}$$

and range

$$U_yL_2(S^*_yM) \;=\; U_{y,\lambda}L_2(S^*_yM) \;=\; \{f \in L_2(S^*_yM) \,:\, \text{supp}\, f \subset \Pi^a_y\}\,.$$

LEMMA 1.8.10. *For each compact set* $K \subset \mathring{M}$ *there exists a constant* $C_K > 0$ *such that* $\mathbf{T}_y(\widetilde{\eta}) \geq C_K$ *for all* $y \in K$, $\widetilde{\eta} \in \Pi_y$.

LEMMA 1.8.11. *For all* $f \in L_2(S^*_y M)$ *the series*

$$\|f\|^2_{L_2(S^*_y M)} \; + \; \sum_{k=1}^{\infty} 2\,\mathrm{Re}\,\big(f, U^k_{y,\lambda}\, f\big)_{L_2(S^*_y M)}$$

converges in the sense of distributions in λ. *Its sum is a positive Borel measure on* $\mathbb{R}$.

By Lemma 1.8.11 the series

$$\sum_{k=1}^{\infty} \big(U^k_{y,\lambda} + (U^k_{y,\lambda})^*\big) \tag{1.8.8}$$

converges in the weak operator topology and defines a Borel measure with values in the space of bounded operators in $L_2(S^*_y M)$.

LEMMA 1.8.12. *The Fourier transform* $\mathcal{F}_{\lambda\to t}\left[\sum_{k=1}^{\infty}\left(U^k_{y,\lambda} + (U^k_{y,\lambda})^*\right)\right]$ *vanishes in a neighbourhood of zero.*

Let $\mathbf{Q}(y,\lambda)$ be the operator-valued distribution function of the measure (1.8.8) such that $\mathcal{F}_{\lambda\to t}[\mathbf{Q}(y,\lambda)] = 0$ for sufficiently small t. Clearly, $\mathbf{Q}(y,\lambda) = \mathbf{Q}^*(y,\lambda)$, $\forall \lambda \in \mathbb{R}$.

LEMMA 1.8.13. *The function* $\mathbf{Q}(y,\lambda)$ *is uniformly bounded and oscillating in the sense that* $\int_0^\lambda \mathbf{Q}(y,\mu)\,d\mu$ *is uniformly bounded.*

Now we can state the following general result (it is an analogue of Theorem 1.7.6).

THEOREM 1.8.14 ([Sa6]). *Let* $y \in \mathring{M}$ *be a regular point. Then*

$$\begin{aligned} c_{0;P}(y)\,\lambda^{n+2\mathbf{l}} + c_{1;P}(y)\,\lambda^{n+2\mathbf{l}-1} + \mathbf{Q}_P(y,\lambda - o(\lambda))\,\lambda^{n+2\mathbf{l}-1} \; - \; o(\lambda^{n+2\mathbf{l}-1}) \\ \leq \; e_{P,P}(\lambda,y,y) \\ \leq \; c_{0;P}(y)\,\lambda^{n+2\mathbf{l}} \; + \; c_{1;P}(y)\,\lambda^{n+2\mathbf{l}-1} + \mathbf{Q}_P(y,\lambda + o(\lambda))\,\lambda^{n+2\mathbf{l}-1} + o(\lambda^{n+2\mathbf{l}-1}) \end{aligned} \tag{1.8.10}$$

as $\lambda \to +\infty$, *where* $c_{0;P}(y)$, $c_{1;P}(y)$ *are the same as in Theorem 1.8.7 and* $\mathbf{Q}_P(y,\lambda) = \Big(P_{\mathbf{l}}|_{S^*_y M}\,, \mathbf{Q}(y,\lambda)\,P_{\mathbf{l}}|_{S^*_y M}\Big)_{L_2(S^*_y M)}$.

Note that, in view of Lemma 1.8.11,

$$\mathbf{Q}_P(y,\lambda+\varepsilon) \; - \; \mathbf{Q}_P(y,\lambda) \; \geq \; -\,\varepsilon\,(n+2\mathbf{l})\,c_{0;P}(y)\,, \qquad \forall \lambda \in \mathbb{R}\,, \quad \forall \varepsilon \geq 0\,,$$

so (1.8.10) makes sense.

5. The case $\mathbf{T}_y \equiv \text{const}$. If the function $\mathbf{Q}_P(y,\lambda)$ is uniformly continuous in λ then (1.8.10) implies the quasi-Weyl formula

$$e_{P,P}(\lambda,y,y) \;=\; c_{0;P}(y)\,\lambda^{n+2\mathbf{l}} \;+\; c_{1;P}(y)\,\lambda^{n+2\mathbf{l}-1} \;+\; \mathbf{Q}_P(y,\lambda)\,\lambda^{n+2\mathbf{l}-1} \;+\; o(\lambda^{n+2\mathbf{l}-1})\,. \tag{1.8.11}$$

In the general case the definition of $U_{y,\lambda}$ involves two non-commuting operators (U_y and multiplications by $\mathbf{T}_y$) which makes it very difficult to give any simple sufficient conditions for the function $\mathbf{Q}_P(y,\lambda)$ to be uniformly continuous.

In this subsection we consider a special case assuming that there exists a constant T such that

$$\mathbf{T}_y(\widetilde{\eta}) \;=\; T \quad \text{almost everywhere on } \Pi_y^a\,. \tag{1.8.12}$$

This assumption is motivated by the following

LEMMA 1.8.15. *In the analytic case $\mathbf{T}_y \equiv \text{const}$ almost everywhere on any open connected subset of Π_y^a.*

Under the condition (1.8.12) $U_{y,\lambda} = e^{-i\lambda T}U_y$, so the operator-valued function $\mathbf{Q}(y,\lambda)$ is defined by the series

$$\sum_{k=1}^{\infty} \big((ikT)^{-1}\,e^{ik\lambda T}(U_y^k)^* \;-\; (ikT)^{-1}\,e^{-ik\lambda T}U_y^k\big)\,.$$

In this case $\mathbf{Q}(y,\lambda)$ is $2\pi T^{-1}$-periodic in λ, so it is uniformly continuous if and only if it is continuous.

PROPOSITION 1.8.16. *Under the condition (1.8.12)*

$$\mathbf{Q}(y,\lambda) \;=\; \{\pi + \arg U_y - \lambda T\}_{2\pi}$$

where $\{\cdot\}_{2\pi}$ denotes the residue modulo 2π, and the function $\{\pi+\arg U_y-\lambda T\}_{2\pi}$ of the contraction operator U_y is understood in the sense of the Sz.-Nagy–Foias calculus [Sz.-NaFoi].

According to [Sz.-NaFoi], any contraction operator U in the Hilbert space H can be represented as $\mathbf{P}\widetilde{U}\Big|_H$, where $\widetilde{U}$ is a unitary operator acting in a wider Hilbert space $\widetilde{H} \supset H$ and $\mathbf{P}$ is an orthogonal projection in $\widetilde{H}$ such that $\mathbf{P}\widetilde{H} = H$ and $U^j = \mathbf{P}\widetilde{U}^j\Big|_H$, $j = 0,1,2,\dots$. The operator $\widetilde{U}$ is called a *unitary dilation* of U. For an arbitrary measurable function F on $\mathbb{S}$ one defines $F(U) = \mathbf{P}F(\widetilde{U})\Big|_H$.

Proposition 1.8.16 implies

THEOREM 1.8.17 ([Sa6]). *Under the condition (1.8.12) the function $\mathbf{Q}_P(y,\lambda)$ is uniformly continuous with respect to λ if and only the function $P_1|_{S_y^*M}$ is orthogonal in $L_2(S_y^*M)$ to all the eigenfunctions of U_y corresponding to eigenvalues lying on $\mathbb{S}$ (i.e., with modulus 1).*

On the other hand, if $P_1|_{S_y^*M}$ is an eigenfunction of U_y corresponding to an eigenvalue e^{is_0}, $s_0 \in (-\pi, \pi]$, then

$$\mathbf{Q}_P(y,\lambda) \;=\; \{\pi + s_0 - \lambda T\}_{2\pi} \int_{S_y^*M} |P_1|^2 \, đ\widetilde{\eta}$$

and $\mathbf{Q}_P(y,\lambda)$ has jumps at the points

$$\Lambda_j \;=\; T^{-1}\,(2\pi j + s_0)\,, \qquad j = 1, 2, \dots\,,$$

In this case, by Theorem 1.8.14,

$$e_{P,P}(y,y,\Lambda_{j+1}{-}\varepsilon) - e_{P,P}(y,y,\Lambda_j{+}\varepsilon) \;=\; o(\Lambda_j^{n+2\mathbf{l}-1})\,, \quad \forall \varepsilon \in (0, \pi T^{-1})\,, \quad \Lambda_j \to +\infty\,,$$

and there exist positive numbers $\varepsilon_j \to 0$ such that

$$e_{P,P}(y,y,\Lambda_j{+}\varepsilon_j) \;-\; e_{P,P}(y,y,\Lambda_j{-}\varepsilon_j) \;=\; 2\pi\,\Lambda_j^{n+2\mathbf{l}-1} \int_{S_y^*M} |P_1|^2 \, đ\widetilde{\eta} \;+\; o(\Lambda_j^{n+2\mathbf{l}-1})\,.$$

In other words, $e_{P,P}(y,y,\lambda)$ has a purely cluster asymptotics with clusters around the points Λ_j. This implies, in particular, that there exist eigenvalues of $\mathcal{A}^{1/(2m)}$ lying in small neighbourhoods of the points Λ_j, i.e., the eigenvalue of U_y generates a series of eigenvalues of $\mathcal{A}^{1/(2m)}$.

EXAMPLE 1.8.18. Let M be a two-dimensional surface of revolution and A be the Laplace–Beltrami operator. Then $e(y,y,\lambda)$ has purely cluster asymptotics at the poles. The corresponding series of eigenvalues is $\lambda_j = \pi(j+1/2)/l$, where l is the length of the meridian.

EXAMPLE 1.8.19. Let M be a ball in $\mathbb{R}^n$ and A be the Laplace operator subject to Dirichlet or Neumann boundary condition. Then (1.8.5) holds for all points y excluding the centre of the ball. At the centre $e(y,y,\lambda)$ has a purely cluster asymptotics.

EXAMPLE 1.8.20. Let M be the flat domain bounded by an ellipse and A be the Laplace operator subject to Dirichlet or Neumann boundary condition. Then (1.8.5) (with $c_{0,P}^{(0)} = (4\pi)^{-1}$, $c_{1,P}^{(0)} = 0$) holds at all interior points excluding the foci of the ellipse. The foci are focal points. The corresponding operators U_y have purely continuous spectrum, so the quasi-Weyl formula (1.8.11) holds. In particular, if y is a focus and P is the operator of multiplication by a $C_0^\infty(\mathring{M})$-function equal to 1 at y, then

$$\mathbf{Q}_P(y,\lambda) \;=\; \frac{1}{8\pi^3 a} \int_{-\infty}^{+\infty} \left\{ \mu \,\ln \frac{1+\varepsilon}{1-\varepsilon} - 4a\,\lambda \right\}_{2\pi} f^2(\mu)\, d\mu \;\equiv\; \sum_{k=1}^{\infty} b_k \sin(4ak\lambda)\,,$$

where

$$f(\mu) \;=\; \int_{-\infty}^{+\infty} \frac{\cos \mu\tau}{\sqrt{\cosh 2\tau}}\, d\tau\,, \qquad b_k \;=\; \frac{(-1)^k}{4\pi^3 ak} \int_{-\infty}^{+\infty} \cos\left(k\mu \ln \frac{1+\varepsilon}{1-\varepsilon} \right) f^2(\mu)\, d\mu\,,$$

$2a$ is the length of the major axis and $0 < \varepsilon < 1$ is the eccentricity. The graph of this function for $\varepsilon = 1/2$ is given below.

Figure 11. The function $\mathbf{Q}_P(y, \lambda)$ for $a = 1\,,\ \varepsilon = 1/2\,.$

6. Off-diagonal asymptotics. We do not consider in this book the asymptotic behaviour of $e(\lambda, x, y)$ when $x \neq y\,.$ This is a much more complicated problem, mostly because the function $e(\lambda, x, y)$ is not monotone in $\lambda\,.$ It is known [Hö1], [DuiGuiHö] that at interior points

$$e(\lambda, x, y) \;=\; O(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \qquad x \neq y\,. \tag{1.8.13}$$

More advanced results can be found in [Sa6]. In particular, in [Sa6] it is shown that under some additional conditions

$$e(\lambda, x, y) \;=\; o(\lambda^{n-1})\,, \qquad \lambda \to +\infty\,, \qquad x \neq y\,,$$

and that the estimate (1.8.13) can not be improved in the general case.

7. Pseudodifferential case. All the main results of this section remain true for a pseudodifferential operator acting on a manifold without boundary [Sa6].