On the Construction of Quasimodes Associated with Stable Periodic Orbits[*]

J. V. Ralston

Department of Mathematics, UCLA, Los Angeles, California 90024, USA

Abstract. Let $H(x, D, \varepsilon)$ be a self-adjoint partial differential operator of the form

$$H = \sum_{k=0}^{K} \varepsilon^k H_k(x, \varepsilon D), \quad x \in R^n.$$

Suppose the hamiltonian system

$$\dot{x} = \frac{\partial H_0}{\partial \xi}, \quad \dot{\xi} = -\frac{\partial H_0}{\partial x}$$

has a nondegenerate stable periodic orbit γ on which $\dot{x} \neq 0$. Then it is possible to construct a sequence of real numbers ε_m tending to zero, a sequence of functions u_m concentrated in a tube of radius $\varepsilon_m^{1/2}$ about the projection of γ into x-space, and a polynomial $E(\varepsilon)$ such that

$$\|(H(\varepsilon_m) - E(\varepsilon_m))u_m\| \leq C\varepsilon_m^M \|u_m\|.$$

The power M depends on the order of stability of γ. The constructions are explicit in terms of solutions of linear O.D.E.'s, and are generalizations of "gaussian beams". Actually, instead of just one sequence, one gets a family of sequences parametrized by the multi-indices of order $n-1$, but the constant C is not independent of these multi-indices. The nondegeneracy hypothesis implies γ is part of a one-parameter family of stable periodic orbits, and C is independent of this parameter.

After presenting the constructions, we discuss their application to the quasi-classical limit in quantum mechanics and their relation to work of Keller, Maslov and others.

We wish to study the behavior of the spectrum of a linear partial differential operator, $P(x, D, \varepsilon)$, depending on a parameter ε, as ε tends to zero.. We assume

$$P(x, D, \varepsilon) = \sum a_\alpha(x, \varepsilon)D^\alpha$$

where $D = (i^{-1} \partial/\partial x_1, \ldots, i^{-1} \partial/\partial x_n)$ and each coefficient a_α is a finite sum of smooth functions of x multiplied by integral powers of ε. Replacing $P(x, D, \varepsilon)$ by $\varepsilon^m P(x, D, \varepsilon)$ for some integer m, we may write

$$P(x, D, \varepsilon) = \sum_{k=0}^{S} \varepsilon^k H_k(x, \varepsilon D)$$

where each $H_k(x, \xi)$ is a polynomial in ξ. We will assume that $P(x, D, \varepsilon)$ is "symmetric to second order" in the sense that if $P^*(x, D, \varepsilon)$ denotes the formal adjoint of $P(x, D, \varepsilon)$ with respect to the inner product

$$(u, v) = \int_{R^n} u(x)\bar{v}(x) f(x) dx$$

then $P - P^*$ has an expansion of the form

$$P(x, D, \varepsilon) - P^*(x, D, \varepsilon) = \sum_{k=2}^{T} \varepsilon^k \tilde{H}_k(x, \varepsilon D).$$

This means that $H_0(x, \xi)$, which plays the role of the principal symbol of P in this setting, is real-valued and that the analogue for P of the "subprincipal symbol" is also real-valued.

We are going to construct "quasi-modes" for P associated with a stable periodic orbit γ for the Hamiltonian system

$$\dot{x} = \partial H_0/\partial \xi, \qquad \dot{\xi} = -\partial H_0/\partial x. \tag{1}$$

We assume that $\dot{x} \neq 0$ on γ. Hence the curve Γ formed by projecting γ onto its x-component is a smooth curve in R^n, possibly with self-intersections. In what follows we will make a "nondegeneracy" hypothesis which will imply that γ belongs to a smooth 1-parameter family of periodic solutions of (1),

$$\gamma(E) = \{(x(t, E), \xi(t, E)) : 0 \leq t \leq T(E)\},$$

defined for E near E_0, the value of H_0 on γ, with $H = E$ on $\gamma(E)$. Then, assuming a natural stability condition on γ, we will construct a sequence of functions $u_m(x, E)$ and a sequence $\varepsilon_m(E)$ of real values for ε tending to zero, such that u_m is small outside a tube around $\Gamma(E)$ of radius $\mathcal{O}(\varepsilon_m^{1/2})$ and

$$\|(P(x, D, \varepsilon_m) - E)u_m\| \, \|u_m\|^{-1} \leq C\varepsilon_m^{3/2} \tag{2}$$

where C is independent of E on an open interval containing E_0. Assuming more restrictive but still generic stability conditions on γ, we will refine the construction of the u_m so that

$$\|(P(x, D, \varepsilon_m) - (E + b_2\varepsilon_m^2 + \ldots + b_M\varepsilon_m^M))u_m\| \, \|u_m\|^{-1} \leq C_M \varepsilon_m^{M+1}. \tag{3}$$

The nondegeneracy and stability conditions arise in the following way. Let $(x(t), \xi(t))$, $0 \leq t < T$ be the solution of (1) which traces γ, and consider the system of equations:

$$\dot{y}_i = \frac{\partial^2 H_0}{\partial \xi_i \partial x_j}(x(t), \xi(t))y_j + \frac{\partial^2 H_0}{\partial \xi_i \partial \xi_j}(x(t), \xi(t))\eta_j,$$

$$\dot{\eta}_i = -\frac{\partial^2 H_0}{\partial x_i \partial x_j}(x(t), \xi(t))y_j - \frac{\partial^2 H_0}{\partial x_i \partial \xi_j}(x(t), \xi(t))\eta_j \tag{4}$$

or more compactly in matrix notation

$$\dot{y} = B(t)y + A(t)\eta \, ,$$
$$\dot{\eta} = - C(t)y - B^t(t)\eta \, .$$

These are the equations governing the flow on the tangent space to phase space along γ induced by the flow (1) on phase space. The linear transformation on C^{2n} given by

$$\mathscr{T} : (y(0), \eta(0)) \to (y(T), \eta(T)) \, ,$$

where $(y(t), \eta(t))$ is a solution of (4), preserves the canonical symplectic 2-form, i.e. $\sigma(v, \tilde{v}) = \sigma(\mathscr{T}v, \mathscr{T}\tilde{v})$ where

$$\sigma((y, \eta), (\tilde{y}, \tilde{\eta})) = y \cdot \tilde{\eta} - \tilde{y} \cdot \eta \, .$$

Hence, if λ is an eigenvalue of $\mathscr{T}$ so is λ^{-1} and $\det \mathscr{T} = 1$ (see Weyl [19, Chapter VI] for a proof that $\det \mathscr{T} = 1$). If $\phi_1 = (\dot{x}(0), \dot{\xi}(0))$, we have $\mathscr{T}\phi_1 = \phi_1$. Thus 1 must be an eigenvalue of $\mathscr{T}$ of (algebraic) multiplicity at least 2.

Definition 1. We say γ is nondegenerate if the eigenvalue 1 has multiplicity 2.

If γ is nondegenerate it follows immediately from the Poincaré Continuation Theorem (cf. [16, pp. 145–148]) that γ belongs to a one parameter family of periodic orbits as described earlier. Moreover, since

$$(x(T(E), E), \xi(T(E), E) = (x(0, E), \xi(0, E))$$

and $H_0(x(0, E), \xi(0, E)) = E$, after differentiating these equations with respect to E we see that the 1-eigenspace of $\mathscr{T}$ is spanned by ϕ_1 and $\phi\tilde{\ }_1 = \{\partial x(0, E_0)/\partial E, \partial \xi(0, E_0)/\partial E\}$ and $\sigma(\phi_1, \phi\tilde{\ }_1) = 1$. Thus, letting $V = \{v \in C^{2n} : \sigma(v, \phi_1) = \sigma(v, \phi\tilde{\ }_1) = 0\}$, we see that $\mathscr{T}V = V$ and V has trivial intersection with the span of ϕ_1 and $\phi\tilde{\ }_1$.

Definition 2. We say γ is stable if the eigenvalues of $\mathscr{T}$ restricted to V have modulus 1 and are distinct.

Since the eigenvalues of $\mathscr{T}$ also occur in conjugate pairs, nondegeneracy and stability imply $\mathscr{T}$ has eigenvalues $1, 1, \lambda_2, \bar{\lambda}_2, \ldots, \lambda_n, \bar{\lambda}_n$ where $\lambda_j = e^{i\theta_j}$, $0 < |\theta_j| < \pi$. We may choose eigenvectors ϕ_i with $\mathscr{T}\phi_i = \lambda_i \phi_i$, $i = 2, \ldots, n$. Since $\mathscr{T}$ preserves σ, one checks easily that

$$\sigma(\phi_i, \phi_j) = 0 \quad \text{for all} \quad i, j \, ,$$
$$\sigma(\phi_i, \overline{\phi_j}) = 0 \quad \text{for} \quad i \neq j \text{ and } i = j = 1 \, . \tag{5}$$

By normalizing the ϕ_i and replacing ϕ_i by $\bar{\phi}_i$ when necessary, one can arrange

$$\sigma(\phi_i, \bar{\phi}_i) = -2(-1)^{1/2} \quad \text{for} \quad i > 1 \, . \tag{6}$$

If we now vary E, assuming γ is nondegenerate and stable, $\mathscr{T}$ and $\lambda_2, \ldots, \lambda_n$ will be smooth functions of E for E near E_0, and we can choose eigenfunctions $\phi_2, \ldots, \phi_n$ satisfying (5) and (6) so that they are also smooth functions of E. Moreover, since the eigenvalues of $\mathscr{T}(E)$ must occur in reciprocal and conjugate pairs, the eigenvalues $\lambda_2(E), \ldots, \lambda_n(E)$ will have modulus 1 for E near E_0 and we can choose $\theta_2(E), \ldots, \theta_n(E)$ smoothly varying.

The construction of $u_m(E)$ satisfying (2) is actually possible under a weaker stability hypothesis suggested by the work of Krein [12] and Moser [14].

Definition 2'. γ is stable if the hermitian form $(-1)^{-1/2}\sigma(v, \bar{v})$ is definite on the eigenspaces of $\mathcal{T}$ restricted to V.

We discuss this extension in Appendix A. In the body of the paper stability will mean the property given in Definition 2. To construct $u_m(E)$ satisfying estimates like (2) we will need the following stronger notions of stability.

Definition 3. If $\prod\limits_{2}^{n} \lambda_i^{\alpha_i} \neq 1$ for any nonzero $(\alpha_2, \ldots, \alpha_n) \in Z^{n-1}$ with $\sum |\alpha_i| \leqq N+1$, we say γ is N-stable.

Note that 1-stability is stability, and, if γ is N-stable, so ist $\gamma(E)$ for E near E_0.

The solutions $\phi_i(t)$ to (4) with $\phi_i(0) = \phi_i$ will play a central role in the construction of the u_m. Note that since the flow defined by (4) preserves σ, the Equations (5) and (6) hold for all t when ϕ_i is replaced by $\phi_i(t)$. If we let $\phi_i(t) = (y^i(t), \eta^i(t))$, the set $\{y^i(t): i = 1, \ldots, n\}$ is a basis for C^n for all t. To see this suppose

$$\sum_{i=1}^{n} a_i y^i(t_0) = 0 .$$

Then letting $v = \sum\limits_{i=1}^{n} a_i \phi_i(t_0)$, we have

$$0 = \sigma(v, \bar{v}) = \sum_{i,j} a_i \bar{a}_j \sigma(\phi_i, \bar{\phi}_j) = \sum_{i=2}^{n} -2(-1)^{1/2} |a_i|^2 .$$

Hence $a_2 = \ldots = a_n = 0$. Since we assumed $\dot{x}(t_0) = y^1(t_0) \neq 0$, $a_1 = 0$ also. Thus the matrix $Y(t)$ with columns $y^1(t), \ldots, y^n(t)$ is nonsingular for $t \in [0, T]$.

We are now in a position to specify the sequence ε_m. Actually there is not just one sequence but a family of sequences parametrized by multi-indices $(m_2, \ldots, m_n)$. If γ is stable, there is a sequence $u_m(x, E)$ satisfying (2) when the sequence $\varepsilon_m(E)$ is defined by

$$K(E) = (2\pi m + m_2 \theta_2(E) + \ldots + m_n \theta_n(E) + \Delta(E))\varepsilon_m(E) \tag{7}$$

where $K = \int\limits_{0}^{T(E)} \xi(t, E)\dot{x}(t, E)dt$—the "action" of $\gamma(E)$—and

$$\Delta(E) = \frac{1}{2} \int\limits_{0}^{T(E)} \frac{d \operatorname{Arg} \det Y(t, E)}{dt} dt$$

$$+ \frac{1}{2} \int\limits_{0}^{T(E)} \operatorname{Re}\{H_1(x(t, E), \xi(t, E))\} dt .$$

Note that when $\operatorname{Re} H_1$ vanishes,

$$\Delta = \frac{\theta_1}{2} + \ldots + \frac{\theta_n}{2} + p\pi, \qquad p \in Z . \tag{7'}$$

If γ is N-stable, for $N = \sum\limits_{i=2}^{n} m_i + 2M$, then the u_m can be made to satisfy (3). The formula (7) is the analogue of the quantum law proposed by Bohr, "$K = 2\pi m\hbar$," with Δ playing the role of the correction added by Sommerfeld. The integer p is the natural generalization to this setting of an index introduced by Gelfand and Lidskii in [6], which we discuss in Appendix B.

The construction of "quasi-modes" of the sort described above can be traced back to the classical WKB method, but it was Keller [11] who suggested using such constructions to study the spectrum of nonseparable equations. In present day terminology Keller discussed quasimodes associated with compact Lagrangian manifolds. This approach was developed, greatly generalized and made rigorous through the work of Maslov [13], Duistermaat [5], Weinstein [18] and Colin de Verdière [4]. For this approach to succeed the Hamiltonian $H_0(x, \xi)$ must be at least "close" to completely integrable, see [4] and [5].

In [3] Babich and Lazutkin discussed the possibility of constructing quasimodes associated with a single periodic orbit. Using the "method of parabolic equations," they constructed a sequence satisfying (2) in the case $n = 2$. More recently Voros [17] deduced the formula (7) for general n by replacing H_0 by its completely integrable linearization near γ and using the theory described in the preceding paragraph. Still more recently Guillemin and Weinstein [8] proved the existence of sequence satisfying (2) in the case that P is the Laplace-Beltrami operator on an orientable manifold M, by constructing an isometry from $L^2(S^1)$ into $L^2(M)$ which approximately intertwines $d^2/d\theta^2$ and P. The isometry from $L^2(S^1)$ into $L^2(M)$ was a Fourier integral operator of a new type introduced by Guillemin in [9].

The approach that we are going to present here is closest to that of Babich and Lazutkin. We explicitly construct quasimodes of the form

$$u = e^{(i/\varepsilon)\psi(x)}(a_0(x) + \varepsilon a_1(x) + \ldots + \varepsilon^N a_N(x)) \tag{8}$$

where ψ is real-valued on Γ with $(\mathrm{Im}\, \partial^2 \psi(x(t))/\partial x_i \partial x_j)$ positive definite on vectors orthogonal to $\dot{x}(t)$. The ansatz (8) is the familiar ansatz of geometric optics but with a complex phase. In physics literature (8) is known as the "Gaussian beam" ansatz or the beam optics approximation, and it has been extensively studied. In particular it has been used to find analogues of the quantum law (7). The article [1] by J. A. Arnaud provides an excellent survey of this work. The ansatz (8) was also used in [15], but our aim in this paper has been to make the computations more explicit and general.

The plan of this paper is as follows. We construct the phase function ψ and the amplitudes $a_1, \ldots, a_N$ in Sections 1 and 2. The refinements necessary to construct quasimodes satisfying (3) are given in Section 3. In this connection we note that the trick of replacing E by $E + b_2 \varepsilon^2 + \ldots + b_M \varepsilon^M$ in (3) as suggested to us by the results of Colin de Verdière [4]. In Section 4 we discuss the conclusions on the spectrum of P which one can draw from (2) and (3). Finally in Section 5 we give applications and discuss the relation of our results to those obtained by the method of Keller and Maslov. For simplicity in Sections 1, 2, 3, and 4 we ignore the parameter E and construct only $u_m(x, E_0)$. However, since the ϕ_i, $i = 1, \ldots, n$, and λ_i, $i = 2, \ldots, n$ can be chosen as smooth functions of E for E in a neighborhood

of E_0, it will be apparent that all estimates are uniform in E. In the case that Γ has self-intersections the approximate eigenfunction u must be a sum of functions of the form (8). Since the modifications necessary for this case will be obvious once the construction is completed, we will assume in Sections 1, 2, 3, and 4 that Γ has no self-intersections.

I am very much indebted to Jürgen Moser for many helpful suggestions and for pointing out the relevance of [6], [12], and [14] for this work. I am also grateful to George Papanicolaou for calling [1] to my attention, and to Mark Levi for explaining [6].

1. Construction of the Phase Function

Letting $E(\varepsilon) = E_0 + b_2\varepsilon^2 + \ldots + b_M\varepsilon^M$, one sees immediately

$$P(x, D, \varepsilon)u - E(\varepsilon)u$$

$$= e^{i\psi/\varepsilon}\left(\left(H_0\left(x, \frac{\partial\psi}{\partial x}\right) - E_0\right)a_0(x) + \sum_{i=1}^{p} c_i(x)\varepsilon^i\right), \tag{9}$$

where $c_i(x)$ is independent of ε. As was indicated earlier we are not going to attempt to solve $P(x, D, \varepsilon)u - E(\varepsilon)u = 0$ exactly. In view of the rapid decrease of $P(x, D, \varepsilon)u - E(\varepsilon)u$ away from Γ, it will suffice for the estimates (2) and (3) to have $H(x, \partial\psi/\partial x) - E_0$ and $c_i(x)$, $i = 1, \ldots, P$ vanish to sufficiently high orders on Γ. In particular for (2) we only need to have $H(x, \partial\psi/\partial x) - E_0$ vanish to third order on Γ, and this section is devoted to the construction of a phase function ψ such that $H(x, \partial\psi/\partial x) - E_0$ vanishes to third order on Γ.

In order that $H_0(x, \partial\psi/\partial x) - E_0 = 0$ for $x = x(t)$, we require $\partial\psi(x(t))/\partial x = \xi(t)$, $0 \leq t < T$. Unless $K = \int_0^T \xi(t)\dot{x}(t)dt = 0$, this already implies that ψ cannot be single-valued. Instead we will require that all derivatives of ψ are single-valued and ψ itself continues to $\psi + K$ when one goes once around Γ.

We proceed by differentiating $H_0(x, \partial\psi/\partial x) - E_0$ and evaluating the derivatives for $x = x(t)$. All of the first order derivatives will vanish if for $i = 1, \ldots, n$

$$\frac{\partial H_0}{\partial x_i}(x(t), \xi(t)) + \frac{\partial H}{\partial \xi_j}(x(t), \xi(t))\frac{\partial^2\psi}{\partial x_j\partial x_i}(x(t))$$

$$= -\dot{\xi}_i(t) + \frac{\partial^2\psi}{\partial x_j\partial x_i}(x(t))\dot{x}_j(t) = 0.$$

However, this is implied by $\dfrac{\partial\psi}{\partial x}(x(t)) = \xi(t)$.

Letting $M(t) = \left(\dfrac{\partial^2\psi}{\partial x_i\partial x_j}(x(t))\right)$ and noting that for any function $f(x)$

$$\sum_{1}^{n} \frac{\partial H_0}{\partial \xi_i}(x(t), \xi(t))\frac{\partial f}{\partial x_i}(x(t)) = \frac{df(x(t))}{dt},$$

one computes easily that all the second order derivatives of $H_0(x, \partial\psi/\partial x) - E_0$ will vanish on Γ provided

$$0 = dM/dt + C + B^t M + MB + MAM$$

for $0 \leq t \leq T$. Here we have used the matrix notation from (4).

Our choice for $M(t)$ will be the matrix defined by $M(t)y^i(t) = \eta^i(t)$, $i = 1, \ldots, n$. Since $\sigma(\phi_i(t), \phi_j(t)) \equiv 0$, $M(t) = M^t(t)$. If we let $N(t)$ denote the matrix with columns $\eta^1(t), \ldots, \eta^n(t)$, then $M(t) = N(t)Y^{-1}(t)$.

$$\frac{dM}{dt} = \frac{dN}{dt} Y^{-1} - NY^{-1} \frac{dY}{dt} Y^{-1}$$

$$= (-CY - B^t N)Y^{-1} - NY^{-1}(BY + AN)Y^{-1}$$

$$= -C - B^t M - MB - MAM.$$

Thus, if we require $\partial\psi(x(t))/\partial x = \xi(t)$ and $\partial^2\psi(x(t))/\partial x_i \partial x_j = M(t)$, $H_0(x, \partial\psi/\partial x) - E_0$ will vanish on Γ together with its first and second derivatives.

The computations in the preceding paragraph are not as surprising as they may seem at first sight. The equation

$$dM/dt + C + B^t M + MB + MAM = 0 \qquad (10)$$

is a matrix Riccati equation and it follows from the general theory of such equations that $M(t) = N(t)Y^{-1}(t)$ will be a solution to it. A more satisfying explanation of why the computations worked comes from an observation of Hörmander [9]. If ψ were real-valued and $H_0(x, \partial\psi/\partial x) = 0 \ \forall x$, then the manifold $(x, d\psi|_x)$ would be invariant under the flow defined in (1). Hence the tangent spaces to this manifold along γ, i.e. the vector spaces

$$S(t) = \left\{ (y, \eta) : \eta_i = \frac{\partial^2 \psi}{\partial x_i \partial x_j}(x(t))y_j \right\}$$

would be invariant under the flow defined in (4). Using the existence and uniqueness theorems for ordinary differential equations, one can conclude from this that to construct a symmetric solution of (10), $M(t) = \dfrac{\partial\psi}{\partial x_i \partial x_j}(x(t))$, compatible with $\partial\psi(x(t))/\partial x = \xi(t)$, it suffices to choose an n-dimensional subspace of C^{2n}, $S(0)$, containing ϕ_1, such that σ vanishes on $S(0)$, and then propagate $S(0)$ once around γ under the flow defined in (4). This yields subspaces $S(t)$, $0 \leq t \leq T$. One recovers $(\partial^2\psi(x(t))/\partial x_i \partial x_j) = M(t)$ by writing the vectors in $S(t)$ in the form $(y, M(t)y)$. This, of course, assumes that $S(t)$ contains no vectors of the form $(0, \eta)$, $\eta \neq 0$. Our choice of $S(0)$ was $\langle \phi_1, \ldots, \phi_n \rangle_C$, giving $S(t) = \langle \phi_1(t), \ldots, \phi_n(t) \rangle_C$ and we showed earlier that $\sum_{i=1}^{n} a_i \phi_i(t_0) = (0, \eta) \Rightarrow a_i = 0$ for all i.

The particular choice of $S(0)$ used here was made in order to satisfy two additional requirements:

(i) we needed $S(0) = S(T)$, so that $\partial^2\psi/\partial x_i \partial x_j$ would be single-valued on Γ.

(ii) We needed $\operatorname{Im} \partial^2\psi(x(t))/\partial x_i \partial x_j = \operatorname{Im} M(t)$ positive definite on the orthogonal complement of $\dot{x}(t)$ so that $\operatorname{Im} \psi > 0$ off Γ. From (5) and (6) we have

$$\bar{y}^j(t) \cdot \operatorname{Im} M(t) y^i(t) = \delta_{ij} \quad \text{for} \quad i > 1 \qquad (11)$$

and $\operatorname{Im} M(t) y^1(t) = 0$. Thus $\operatorname{Im} M(t)$ is semi-definite with null space spanned by $\dot{x}(t)$. Thus we may assume $\operatorname{Im} \psi(x) > cd^2$ for x near Γ, where d denotes the distance from x to Γ.

This completes the construction of the phase function $\psi(x)$.

2. Construction of the Amplitude

If we let $c_0 = H_0\left(x, \dfrac{\partial \psi}{\partial x}\right) - E_0$, then the coefficients $c_i(x)$ in (9) are given by

$$
\begin{aligned}
c_1 &= \frac{1}{i} \frac{\partial H_0}{\partial \xi_j}\left(x, \frac{\partial \psi}{\partial x}\right) \frac{\partial a_0}{\partial x_j} + \frac{1}{2i} \frac{\partial^2 H_0}{\partial \xi_j \partial \xi_k}\left(x, \frac{\partial \psi}{\partial x}\right) \frac{\partial^2 \psi}{\partial x_j \partial x_k} a_0 \\
&\quad + H_1\left(x, \frac{\partial \psi}{\partial x}\right) a_0 + c_0(x) a_1 \\
&\equiv L a_0 + c_0 a_1 \\
c_s &= L a_{s-1} = g_s + c_0 a_s
\end{aligned}
\tag{12}
$$

where g_s, $s = 2, \ldots$ is a complicated function of ψ, $a_0, \ldots, a_{s-2}$, and the derivatives of these functions. In Section 1, ψ was constructed so that $c_0(x)$ vanishes to third order on Γ, and we will assume this here.

As in Section 1 we want to choose $a_0, \ldots, a_N$ so that the functions $c_i(x)$ vanish on Γ to given orders. We also want to solve recursively, determining a_i before a_{i+1}. Differentiating $c_1(x)$ and equating the derivatives to zero on Γ one sees that the rth order partial derivatives of a_0 must satisfy an inhomogeneous ordinary differential equation along Γ with inhomogeneous terms depending on the derivatives of a_0 and a_1 of order up to $r-1$ and $r-3$ respectively. If we want to solve for the partial derivatives of a_0 on Γ up to order l, it follows that we'll be able to do this before determining a_1, only if a_1 vanishes to order $l-2$ at all points of Γ. Hence, if a_0 vanishes to order l at $x(0)$, it must vanish to order l at all points of Γ.

In view of the preceding remarks we construct the principal amplitude a_0 in the following way. We assume that a_0 vanishes to order l at all points of Γ and a_1 vanishes to order $l-2$ on Γ. Then c_1 will vanish to order $l+1$ on Γ provided the lth order derivatives of a_0 satisfy a *homogeneous* ordinary differential equation along Γ. We do not require that a_0 be single-valued near Γ, but we do require that a_0 and all its derivatives be multiplied by a fixed $\lambda \in C$, when one goes once around Γ. Since the mapping T_l of the data for the lth order partial derivatives of a_0 at $x(0)$ to the data at $x(T)$ that one gets from the ordinary differential equation along Γ is linear, determining the possible values of λ and the corresponding a_0's is equivalent to finding the eigenvalues and eigenvectors for T_l, $l = 0, 1, \ldots$. This turns out to be quite easy to do.

Let $\psi_i(x)$, $i = 2, \ldots, n$ be functions satisfying $\dfrac{\partial \psi_i}{\partial x}(x(t)) = (\operatorname{Im} M(t)) \bar{y}^i(t)$ and $\psi_i(x(0)) = 0$. Since $\dot{x}(t) \cdot \dfrac{\partial \psi_i}{\partial x}(x(t)) = y'(t) \cdot (\operatorname{Im} M(t)) \bar{y}^i(t)) = 0$, it follows that $\psi_i(x(t)) = 0$

for all t. Since $\dfrac{\partial \psi_i}{\partial x}(x(T)) = \bar{\lambda}_i \dfrac{\partial \psi_i}{\partial x}(x(0))$, we cannot make ψ_i single-valued. Instead we define it near Γ so that it and all its derivatives are multiplied by $\bar{\lambda}_i$ when one goes once around Γ. Also let $g(x)$ be a function such that

$$g(x(t)) = (f(x(t)) \det Y(t))^{1/2} e^{i \int_0^t \operatorname{Re} H_1(x(s), \xi(s)) \, ds}.$$

We choose g near Γ so that it and its derivatives are multiplied by $e^{i\Delta}$ when one goes once around Γ.

We claim that if

$$a_0 = g^{-1} \psi_2^{m_2} \ldots \psi_n^{m_n}, \qquad m_2 + \ldots + m_n = l \tag{13}$$

then $c_1(x)$ vanishes to order $l+1$ on Γ. It is possible to give an explanation (not really a proof) of why this works based on the fact that in conventional geometric optics the equation $L a_0 = 0$ [see (12)] is equivalent to the invariance of $a_0^2 f \, dx_1 \wedge \ldots \ldots \wedge dx_n$ under the hamiltonian flow restricted to the surface $(x, d_x \psi)$ (see [15]). Here, however, we will just compute. Modulo terms vanishing on Γ to order $l+1$,

$$c_1(x) = i^{-1} [\mu(x) \cdot \partial a_0 / \partial x + \beta(x) a_0]$$

where $\mu(x(t)) = \dot{x}(t)$, $\partial \mu_j(x(t)) / \partial x_i = (\dot{Y} Y^{-1})_{ij}$ and

$$\beta(x(t)) = 2^{-1} \operatorname{Trace} \dot{Y} Y^{-1} + d f^{1/2}(x(t))/dt + i \operatorname{Re} H_1(x(t), \xi(t)).$$

The formula for β takes this form because the hypothesis that P was "symmetric to second order" implies

$$\operatorname{Im} H_1(x, \xi) = 2^{-1} \partial^2 H_0 / \partial x_i \partial \xi_i - (2f)^{-1} (\partial H_0 / \partial \xi_i)(\partial f / \partial x_i).$$

For any matrix function $E(t)$ one has the identity

$$\operatorname{Trace}((dE/dt) E^{-1}) = d \log \det E / dt.$$

Using this identity, we see that $c_1(x)$ vanishes to order $l+1$ on Γ if $\mu(x) \cdot \partial g a_0 / \partial x$ vanishes to order $l+1$ on Γ. Furthermore, when $g a_0 = \psi_2^{m_2} \ldots \psi_n^{m_n}$, $\mu(x) \cdot \partial g a_0 / \partial x$ vanishes to order $l+1$ on Γ, if $\mu(x) \cdot \partial \psi_i / \partial x$ vanishes to order 2 on Γ. The last condition is easy to check:

$$\mu(x(t)) \cdot \frac{\partial \psi_i}{\partial x}(x(t)) = \dot{x}(t) \cdot (\operatorname{Im} M) \bar{y}^i(t) = 0,$$

$$\frac{\partial}{\partial x}\left(\mu(x) \cdot \frac{\partial \psi_i}{dx} \right) \bigg|_{x = x(t)} = (\dot{Y} Y^{-1})^t (\operatorname{Im} M) \bar{y}^i + \frac{d}{dt}(\operatorname{Im} M) \bar{y}^i.$$

Since $y^j \cdot (\operatorname{Im} M) \bar{y}^i = \delta_{ij}$,

$$\dot{Y}^t (\operatorname{Im} M) \bar{y}^i + Y^t d \operatorname{Im} M \bar{y}^i / dt = 0. \tag{14}$$

Thus when a_0 is given by (13) $c_1(x)$ vanishes to order $l+1$ on Γ.

The preceding computations show that the mapping T_l has the eigenvectors

$$\left\{ \frac{\partial \psi_{i_1}}{\partial x}(x(0)) \underset{s}{\bigotimes} \ldots \underset{s}{\bigotimes} \frac{\partial \psi_{i_l}}{\partial x}(x(0)) : i_1 \leq i_2 \leq \ldots \leq i_l \right\}.$$

These vectors are linearly independent and hence span a space of dimension η_l, where η_l is the number of $n-1$ component multi-indices of length l. If a_0 vanishes to order l on Γ, then any partial derivative of a_0 of order l which includes a differentiation in a direction tangent to Γ must vanish. Thus the invariant subspace S_l of the domain of T_l, consisting of partial derivatives of functions vanishing to order l on Γ has dimension η_l, and we see that we have found a basis for S_l consisting of eigenvectors for T_l.

We turn now to the coefficients $c_s(x)$, $s \geqq 2$. As noted earlier if we are to solve recursively a_1 must vanish to order $l-2$ on Γ and, in general, a_s must vanish to order $l-2s$. We want to make c_s vanish to order $l-2s+3$ on Γ. Modulo terms which vanish on Γ to order $l-2s+3$,

$$c_s = \frac{1}{i}\left[\mu(x) \cdot \frac{\partial a_{s-1}}{\partial x} + \beta(x) a_{s-1}\right] - \frac{1}{2}\frac{\partial^2 H_0}{\partial \xi_i \partial \xi_j}\left(x, \frac{\partial \psi}{\partial x}\right)\frac{\partial^2 a_{s-2}}{\partial x_i \partial x_j}.$$

We want to choose a_s so that it is multiplied by the same factor, when one goes once around Γ, that multiplied a_0. The reader can easily convince himself that, if $\prod_{i=2}^{n} \lambda^{m_i} = \prod_{i=2}^{n} \lambda^{p_i}$ where $\sum m_i = l$ and $\sum p_i = l-2s$, the existence of such an a_s imposes compatibility conditions on the inhomogeneous term $\frac{1}{2}\frac{\partial^2 H_0}{\partial \xi_i \partial \xi_j}\left(x, \frac{\partial \psi}{\partial x}\right)\frac{\partial^2 a_{s-1}}{\partial x_i \partial x_j}$. Surprisingly these conditions are always satisfied, and once again we can construct the amplitudes $a_s(x)$ explicitly.

As a preliminary step we note that (14) implies

$$0 = (BY + AMY)^t (\operatorname{Im} M)\bar{y}^j + Y^t \frac{d(\operatorname{Im} M)\bar{y}^j}{dt}.$$

Thus

$$\frac{d(\operatorname{Im} M)\bar{y}^j}{dt} = -B^t(\operatorname{Im} M)\bar{y}^j - MA(\operatorname{Im} M)\bar{y}^j$$

and

$$\frac{d}{dt}\bar{y}^i \cdot \operatorname{Im} M\bar{y}^j = -2\sqrt{-1}\,(\operatorname{Im} M)\bar{y}^i \cdot A(\operatorname{Im} M\bar{y}^j). \tag{15}$$

When $a_0 \ldots g^{-1} \prod_{k=1}^{l} \psi_{i_k}$,

$$\frac{\partial^2 H_0}{\partial \xi_i \partial \xi_j}\left(x, \frac{\partial \psi}{\partial x}\right)\frac{\partial^2 a_0}{\partial x_i \partial x_j} = \frac{1}{g}\sum_{k_1 \neq k_2}\frac{\partial \psi_{i_{k_1}}}{\partial x} \cdot A \frac{\partial \psi_{i_{k_2}}}{\partial x}\prod_{\substack{k=1 \\ k \neq k_1, k_2}}^{l}\psi_{i_k}$$

plus terms vanishing on Γ to order $l-1$. Hence by (15), if we set P equal to the operator $-\frac{1}{4}\frac{\partial}{\partial x} \cdot P^\sim(x)\frac{\partial}{\partial x}$, where the matrix $P^\sim(x(t))$ satisfies

$$P^\sim(x(t))\operatorname{Im} M(t)\bar{y}^i(t) = \bar{y}^i(t), \quad i = 2, \ldots, n,$$

and set $g_1 = Pa_0$, it follows that c_2 vanishes on Γ to order $l-1$. When $a_1 = Pa_0$
and $a_0 = g^{-1} \prod_{k=1}^{l} \psi_{i_k}$,

$$\frac{\partial^2 H_0}{\partial \xi_i \partial \xi_j} \left(x, \frac{\partial \psi}{\partial x} \right) \frac{\partial^2 a_1}{\partial x_i \partial x_j}$$

$$= \sum_{\substack{k_1, k_2, k_3, k_4 \\ \text{distinct}}} \left(\frac{\partial \psi_{i_{k_1}}}{\partial x} \cdot A \frac{\partial \psi_{i_{k_2}}}{\partial x} \right) \left(\frac{\partial \psi_{i_{k_3}}}{\partial x} \cdot P \frac{\partial \psi_{i_{k_4}}}{\partial x} \right) \cdot \prod_{k \neq k_1, \dots, k_4} \psi_{i_k}$$

plus terms vanishing on Γ to order $l-3$. Now one checks that c_3 vanishes to order $l-3$ on Γ, if we set $a_2 = \frac{1}{2} Pa_1 = \frac{1}{2!} P^2 a_0$. Continuing in this way one checks that c_{s+1} vanishes to order $l-2s+1$ on Γ when $a_s = \frac{1}{s} Pa_{s-1} = \frac{1}{s!} P^s a_0$. We stop determining a_s when $l-2s+1$ becomes negative and this determines the index N in the formula (8) for u. Our choice for a_0 implies that a_0 is multiplied by $\bar{\lambda}_2^{m_2} \dots \bar{\lambda}_n^{m_n} e^{-i\Delta}$ when one goes around Γ. Note that this is also true of $a_s = (1/s!) P^s a_0$.

At this point we have constructed

$$u = e^{i\psi/\varepsilon} \left(a_0 + \varepsilon Pa_0 + \dots + \frac{\varepsilon^N}{N!} P^N a_0 \right)$$

so that

$$(P(x, D, \varepsilon) - E_0) u = e^{i\psi/\varepsilon} (c_0 + \varepsilon c_1 + \dots) \tag{16}$$

where c_s vanishes to order $l-2s+3$ on Γ. The final step is to choose ε so that u will be single-valued in a neighborhood of Γ. From the preceding it is clear that this will be the case when

$$\frac{1}{\varepsilon} K - (m_2 \theta_2 + \dots + m_n \theta_n + \Delta) = 2\pi m$$

which is precisely the quantum law (7).

3. Refinement of the Construction

In this section we sketch the improvements made possible by more stringent stability hypotheses on γ. The reader who wishes to check the details of the computations will probably find it helpful to introduce coordinates near Γ such that $n-1$ coordinates are constant on Γ.

If the sequence u_m constructed in Section 2 is to satisfy (3), we will see that the coefficient $c_s(x)$ in (16) must vanish to order $l-2s+2M+2$ on Γ. Differentiating the coefficients $c_s(x)$ and setting the derivatives equal to zero on Γ, we obtain differential equations for $\frac{\partial^\alpha \psi}{\partial x^\alpha}(x(t))$, $2 < |\alpha| \leq 2M+1$, and for $\frac{\partial^\alpha a_s}{\partial x^\alpha}(x(t))$, $l-2s < |\alpha| \leq$ $l-2s+2M+1$, with boundary conditions $\frac{\partial^\alpha \psi}{\partial x^\alpha}(x(0)) = \frac{\partial^\alpha \psi}{\partial x^\alpha}(x(T))$ and $\frac{\partial^\alpha a_s}{\partial x^\alpha}(x(0)) =$ $\beta \frac{\partial^\alpha a_s}{\partial x^\alpha}(x(t))$ where $\beta = \lambda_2^{m_2} \dots \lambda_n^{m_n} e^{i\Delta}$.

One gets a *linear* differential equation for the rth-order derivatives of ψ on Γ with an inhomogeneous term involving the derivatives of ψ up to order $r-1$, when $r > 2$. Since any r^{th}-order derivative involving a differentiation in a direction tangential to Γ can be obtained by differentiating the derivatives $\dfrac{\partial^\alpha \psi}{\partial x^\alpha}(x(t))$, $|\alpha| = r-1$ with respect to t, we can include these derivatives in the inhomogeneous term. When we do this, we get an inhomogeneous equation for the rth-order nontangential derivatives of ψ on Γ, such that the corresponding homogeneous equation is precisely the equation derived in Section 2 for the rth-order derivatives of ga_0 in the case $l = r$. By the results in Section 2 this equation has nontrivial solutions satisfying periodic boundary conditions only if $\lambda_2^{p_2} \ldots \lambda_n^{p_n} = 1$ for some multi-index $(p_2, \ldots, p_n)$ of length r. Thus, by the Fredholm alternative, the $2M$-stability of γ implies we can solve the equations necessary to make $c_0(x)$ vanish to order $l + 2M + 2$ on Γ.

The situation for amplitudes a_s is quite similar. We get a linear differential equation for the rth-order partial derivatives of a_s on Γ with an inhomogeneous term involving the derivatives of a_s up to order $r-1$ and previously determined derivatives of ψ and $a_0, \ldots, a_{s-1}$. The coefficients $b_2, \ldots, b_M$ also appear in the computations. b_{s+1} appears first in the equation for the lth-order derivatives for a_s, where the inhomogeneous term takes the form

$$b_{s+1} \frac{1}{g} \frac{\partial \psi_i^1}{\partial x} \underset{s}{\otimes} \ldots \underset{s}{\otimes} \frac{\partial \psi_{i_1}}{\partial x} + h_{s,l}$$

where $h_{s,l}$ does not depend on b_{s+1}. When we place all the rth-order derivatives of a_s involving tangential differentiations in the inhomogeneous terms, we again get an inhomogeneous equation for the nontangental r^{th}-order derivatives of a_s. This time the corresponding homogeneous equation is the equation derived in Section 2 for the rth-order derivatives of a_0 in the case $l = r$. By the results in Section 2 this equation has nontrivial solutions with the multiplicative period β if and only if

$$\lambda_2^{m_2} \ldots \lambda_n^{m_n} = \lambda_2^{p_2} \ldots \lambda_n^{p_n} \tag{17}$$

for some multi-index $(p_2, \ldots, p_n)$ of length r. If γ is $(l + 2M)$-stable, (17) holds if and only if $(p_2, \ldots, p_n) = (m_2, \ldots, m_n)$. In this case the space nontrivial solution is spanned by

$$v = \frac{1}{g} \frac{\partial \psi_{i_1}}{\partial x} \underset{s}{\otimes} \ldots \underset{s}{\otimes} \frac{\partial \psi_{i_l}}{\partial x}.$$

Thus by the Fredholm alternative, there is a unique choice for b_{s+1} such that the equation for the lth-order partial derivatives of a_s has a solution. This solution is, of course, not unique and this is the first point where we have encountered only nonuniqueness in this construction. However, the nonuniqueness here arises only because if u_m is a sequence satisfying (3) so is $(1 + a\varepsilon_m + \ldots + b\varepsilon_m^M)u_m$.

Summarizing the results of the preceding two paragraphs, given that γ is $(l + 2M)$-stable we can choose $\psi, a_0, a_1, \ldots, a_N$ and $b_2, \ldots, b_M$ so that $c_s(x)$ in (12) vanishes to order $l - 2s + 2M + 2$. Note that now we stop solving equations when $l - 2s + 2M + 2$ becomes negative so that here N is larger than it was in (12).

4. Consequences of the Estimates (2) and (3)

Before discussing the consequences of (2) and (3), we should point out that the sequences $\{u_m\}_{m=1}^{\infty}$ constructed in Sections 2 and 3 do satisfy (2) and (3) respectively. We assume that the u_m have been smoothly cut off to zero outside an m-independent neighborhood of Γ. To check (2) and (3) we introduce new coordinates $w_1, \ldots, w_n$ near Γ so that Γ corresponds to $w_2 = \ldots = w_n = 0$, and then change variables in the integrals defining the norms in (2) and (3) via

$$z_1 = w_1, \qquad z_i = \varepsilon_m^{-1/2} w_i, \qquad i = 2, \ldots, n .$$

Letting $E(\varepsilon) = E_0 + b_2 \varepsilon^2 + \ldots + b_M \varepsilon^M$, one can then check easily that

(i) $\| (P(x, D, \varepsilon_m) - E(\varepsilon_m)) u_m \| < c \varepsilon_m^{n/4 + K/2}$

where $K = l + 3$ for the u_m constructed in Section 2 and $K = l + 2M + 2$ for the improved sequence constructed in Section 3, and

(ii) $\| u_m \| = \varepsilon_m^{n/4 + l/2} c_2(\varepsilon_m)$

where $\lim_{x \to 0+} c_2(x)$ is essentially the L^2-norm of the product of a Gaussian exponential with a nonzero polynomial in the normal variables. Hence $\lim_{x \to 0} c_2(x) > 0$

and (2) and (3) hold.

Let $\mathscr{P}(\varepsilon)$ be any closed extension of $P(x, D, \varepsilon)$ defined on $C_0^{\infty}(R^n)$. Then either $E(\varepsilon_m)$ is in the spectrum of $\mathscr{P}(\varepsilon_m)$ or it follows from (2) and (3) that

$$\| (\mathscr{P}(\varepsilon_m) - E(\varepsilon_m))^{-1} \| \geq (C \varepsilon_m)^{-K} \tag{18}$$

where $K = 3/2$ for the sequence $\{u_m\}$ constructed in Section 2, and $K = M + 1$ for the sequence constructed in Section 3. If $\mathscr{P}(\varepsilon)$ is self-adjoint then (18) implies that the spectrum of $\mathscr{P}(\varepsilon)$ must intersect the intervals $|\lambda - E(\varepsilon_m)| \leq C \varepsilon_m^K$. If one knows that $\mathscr{P}(\varepsilon)$ has only point spectrum in these intervals it follows that $\mathscr{P}(\varepsilon)$ has eigenvalues in each of these intervals. Since integration by parts in the tangential variable z_0 shows

$$|(u_m, \bar{u}_m)| \leq C_r \varepsilon_m^r, \qquad \forall r ,$$

if $\mathscr{P}(\varepsilon)$ has real coefficients then it must have at least 2 eigenvalues, counted by multiplicity in the intervals $|\lambda - E(\varepsilon_m)| \leq 2 C \varepsilon_m^K$, when m is sufficiently large.

5. Applications

The results described here have applications in differential geometry, where they yield eigenvalues of the Laplace-Beltrami operator, in geometric optics, where they yield "bouncing ball" waves and "ducted" waves, and in quantum mechanics. The applications to differential geometry require replacing R^n by a compact manifold, and in geometric optics one replaces R^n by a subset of R^n with smooth boundary and considers rays reflected off the boundary. Since the applications to differential geometry and geometric optics require these modifications in the presentation and were already discussed in [11], we will restrict ourselves to quantum mechanics here.

In quantum mechanics $P(x, D, \varepsilon)$ is the Schrödinger hamiltonian of a system with ε equal to Planck's constant $\hbar$. $H_0(x, \xi)$ is the classical hamiltonian for the same system, and the limit $\varepsilon \to 0$ is known as the classical limit. $P(x, D, \varepsilon)$ is self-adjoint with respect to the inner product on $L^2(R^n)$. $P(x, D, \varepsilon)$ is generally bounded below with pure point spectrum in the complement of a closed half-line $[a, \infty)$.

To apply the results described here we need to find a stable periodic orbit γ of energy E_0 for the classical hamiltonian, such that the system is never completely at rest at any point on γ, i.e. such that $\dot{x} \neq 0$ on γ. Then, as was noted in the introduction there is a smooth family of stable periodic orbits $\gamma(E)$ with $\gamma(E_0) = \gamma$ for E in an interval I_0 about E_0, and the quantum law (7) defines

$$\varepsilon_m(E) = K(E)(2\pi m + m_2\theta_2(E) + \ldots + m_n\theta_n(E) + \Delta(E))^{-1}$$

as a smooth function of E on I_0 for m large. If I_0 is contained in $(-\infty, a)$, then (2) implies $P(x, D, \varepsilon_m(E))$ has a bound state energy level in the interval $|E - \lambda| < Cm^{-3/2}$. If λ is $(2M + l)$-stable, the analogous statement with $3/2$ replaced by $M + 1$ and I_0 replaced by a possibly smaller interval follows from (3). If $E_0 > a$, we cannot make any assertions about the existence of bound states. In this case (2) shows that for m large there is something like "spectral concentration" at E_0, see [8].

In [2] Arnol'd proves, modulo a "transversality" hypothesis, that even when $P(x, D, \varepsilon)$ has pure point spectrum, in general one cannot expect the quasimodes u_m to be close to true eigenfunctions of the operator $P(x, D, \varepsilon_m(E))$. However, if we set $w_m(x, t)$ equal to the solution to the Schrödinger equation

$$\frac{1}{i} \frac{\partial w}{\partial t} = P(x, D, \varepsilon_m(E))w$$

with initial data $u_m(x)$, then

$$\sup_{0 \leq t \leq T} \|w_m(t, \cdot) - e^{iEt}u_m(\cdot)\| \leq Cm^{-3/2}T. \tag{19}$$

Hence Arnol'd's term "quasimodes" for the sequence u_m. The proof of (19) is given in [4] and [15].

In Arnol'd's example the orbit γ can be made $(l + 2M)$-stable, but the symmetry of the example is such that for all $\alpha > 0$ the intervals

$$|\lambda - (E_0 + b_2\varepsilon_m^2 + \ldots + b_M\varepsilon_m^M)| < \alpha\varepsilon_m^M$$

contain at least 3 eigenvalues of $P(x, D, \varepsilon_m)$, counted by multiplicity, for m sufficiently large. If for some α these intervals contained at most one eigenvalue (and no continuous spectrum), it would follow directly from (3) and the spectral theorem that $P(x, D, \varepsilon_m)$ had eigenfunctions ϕ_m such that $\|\phi_m - \|u_m\|^{-1}u_m\| \leq \frac{2C}{\alpha}\varepsilon_m$ for m large.

We conclude by discussing the relation of these results to those that may be obtained by the method of Keller and Maslov. The typical situation where both methods are applicable is the case that H_0 is completely integrable. Hence we assume that there are functions $I_2(x, \xi), \ldots, I_n(x, \xi)$ such that the Poisson brackets $\{I_i, I_j\}$ and $\{H_0, I_j\}$ vanish for all i, j, and that the surfaces $S(E, c_2, \ldots, c_m) = \{(x, \xi) : H_0 = E, I_2 = c_2, \ldots, I_n = c_n\}$ are smooth n-tori for $(E, c_2, \ldots, c_m)$ in open set

$\mathcal{O}$ in R^n. In this situation (see [5]) one can construct functions u concentraced near the projection of $S(E, c_2, ..., c_n)$ into x-space such that

$$\|P(x, D, \varepsilon)u - Eu\| \|u\|^{-1} \leq C\varepsilon^{3/2} \tag{20}$$

whenever

$$\int_{\gamma_i} \xi \cdot dx = 2\pi \left(m_i + \frac{I_i}{4} \right), \quad m_i \in Z$$

where γ_i, $i = 1, ..., n$, are generators of the torus $S(E, c_2, ..., c_n)$ and the integer I_i is the Keller-Maslov-Arnol'd-Hörmander index of γ_i. A sharper form of (20) analogous to (3) holds under more restrictive hypotheses (see [4]).

Suppose that, for some $(E_0, \hat{c}_2, ..., \hat{c}_n) \in \bar{\mathcal{O}}$, $S(E_0, \hat{c}_2, ..., \hat{c}_n)$ is a topological circle on which $\partial H_0/\partial \xi$ does not vanish. Then $S(E_0, \hat{c}_2, ..., \hat{c}_n)$ will be a periodic orbit γ with a projection Γ in x-space. To get quasimodes u_m analogous to those construction in Sections 1–3, using the Keller-Maslov construction, one whould pick $(\varepsilon_m, c_2, ..., c_n)$ so that on $S(E_0, c_2, ..., c_n)$

$$\int_{\gamma_1} \xi \cdot dx = 2\pi \left(m + \frac{I_1}{4} \right) \varepsilon_m$$

and

$$\int_{\gamma_i} \xi \cdot dx = 2\pi \left(m_i + \frac{I_i}{4} \right) \varepsilon_m, \quad i = 2, ..., n$$

and let m tend to infinity while $(m_2, ..., m_n)$ remains fixed. If this makes $(c_2, ..., c_n)$ converge to $(\hat{c}_2, ..., \hat{c}_n)$, then the Keller-Maslov construction produces a sequence of functions u_m which are small outside shrinking neighborhoods of Γ. This is closely related to the procedure used in [17] to deduce the quantum law (7). As a rigorous procedure, it suffers from the difficulty that it is not clear that the constant C in (20) will be independent of m as the curvatures of $S(E_0, c_2(m), ..., c_n(m))$ tend to ∞—as they must when m tends to infinity—but it does lead to the correct quantum law. The method of Sections 1–3 is in a sense opposite to this, since it builds out from γ instead of contracting to it.

Appendix A

Let $\langle v, w \rangle = -\frac{1}{2i}\sigma(v, \bar{w})$. Since $\mathcal{T}$ preserves σ, we have $\langle \mathcal{T}v, \mathcal{T}w \rangle = \langle v, w \rangle$. If $\langle , \rangle$ is definite on a generalized eigenspace S_λ of $\mathcal{T}$, it follows that $\mathcal{T}v = \lambda v$ for all $v \in S_\lambda$ and $|\lambda| = 1$. Since $v \in S_\lambda$ if and only if $\bar{v} \in S_{\bar{\lambda}}$, if $\langle , \rangle$ is positive definite on S_λ, it must be negative definite on $S_{\bar{\lambda}}$. From these considerations we see that if γ is stable in the sense of Definition 2', we can choose $\phi_2, ..., \phi_n$ and $\lambda_2, ..., \lambda_n$ so that $\mathcal{T}\phi_i = \lambda_i\phi_i$, $\sigma(\phi_i, \bar{\phi}_i) = -2\sqrt{-1}$, and $\sigma(\phi_i, \bar{\phi}_j) = 0$ when $i \neq j$ and $\lambda_i = \lambda_j$. It also follows that $\lambda_i \neq \bar{\lambda}_j$ for $2 \leq i, j \leq n$, and hence one can verify that the rest of the relations (5) and (6) hold. Since these relations are all that was used in Sections 1, 2 and 4, it follows that approximate eigenfunctions satisfying (2) when $E = E_0$ can be constructed by the method presented there.

When we vary E, the mapping $\mathcal{T}$ varies smoothly, but in general its eigenvectors do not. Nonetheless the following is true. If we let $\tilde{S}$ denote the span of $\phi_2(E_0),\dots,\phi_n(E_0)$, then $\langle\,,\,\rangle$ is positive definite on $\tilde{S}$. Since $\lambda_i(E_0)\neq\bar{\lambda}_j(E_0)$, it follows that for E near E_0 there is a smoothly varying subspace, $\tilde{S}(E)$, invariant under $\mathcal{T}(E)$ with $\tilde{S}(E_0)=\tilde{S}$, and, letting $|\ |$ denote the standard norm on C^{2n},

$$|v|^2\geqq\langle v,v\rangle\geqq\delta|v|^2,\qquad \delta>0 \tag{A.1}$$

on $\tilde{S}(E)$. Hence, given E near E_0, we see that the eigenvalues of $\mathcal{T}(E)$ on $\tilde{S}(E)$, $\lambda_2(E),\dots,\lambda_n(E)$ satisfy $|\lambda_i(E)|=1$, $\lambda_i(E)\neq\bar{\lambda}_j(E)$. Hence, we can choose $\phi_2(E),\dots,\phi_n(E)$ satisfying (5) and (6). These ϕ_i may vary wildly with E, however, in view of (A.1), we can use the Gram-Schmidt procedure to choose a smoothly varying basis for $\tilde{S}(E),\varrho_2(E),\dots,\varrho_n(E)$ which is orthonormal with respect to $\langle\,,\,\rangle$. Since $\phi_2(E),\dots,\phi_n(E)$ is also orthonormal, the $(n-1)\times(n-1)$ matrix $U(E)$ mapping coordinates with respect to the basis $\phi_2(E),\dots,\phi_n(E)$ to coordinates with respect to the basis $\varrho_2(E),\dots,\varrho_n(E)$ is unitary.

Next we propagate the vectors $\varrho_2(E),\dots,\varrho_n(E)$ along γ under the flow (4), getting $\varrho_2(t,E),\dots,\varrho_n(t,E)$, and let $S(t,E)$ be the span of $\{\phi_1(t,E),\varrho_2(t,E),\dots,\varrho_n(t,E)\}$. Since the projection $P:(y,\eta)\to y$ is nonsingular on $S(t,E_0)$ for $0\leqq t\leqq T$, it follows that P is nonsingular on $S(t,E)$ for $0\leqq t\leqq T(E)$ and E near E_0. The matrix $Y(t,E)$ used in Sections 1 and 2 can be written as

$$\begin{pmatrix}1 & 0\\ 0 & U^T(E)\end{pmatrix}P(\phi_1(t,E)\varrho_2(t,E)\dots\varrho_n(t,E))\,.$$

Thus we see that $Y(t,E)$ is uniformly bounded in t, E together with its inverse. Moreover, letting $\tilde{P}$ denote the map $\tilde{P}:(y,\eta)\to\eta$, we see the matrix $M(t,E)$ from Section 1 satisfies

$$M(t,E)Pv=\tilde{P}v,\qquad \forall v\in S(t,E)\,. \tag{A.2}$$

Taking (A.2) as the definition of $M(t,E)$, it follows that $M(t,E)$ depends smoothly on E.

From the observations of the preceding paragraph we conclude:

(i) The phase function ψ can be chosen as a smooth function of (x,E) for E near E_0.

(ii) $|\det Y(t,E)|$ is a smooth nonvanishing function of (t,E) for E near E_0.

(iii) The lengths of the vectors $\dfrac{\psi_2}{\partial x}(x(t))$, $\dfrac{\partial\psi_n}{\partial x}(y(t))$ are bounded above and below for $0\leqq t\leqq T(E)$ and E near E_0.

(i)–(iii) Suffice for the proof that the constant in (2) can be chosen independent of E for E in a neighborhood of E_0.

Appendix B

In [6] Gelfand and Lidskii introduce a class of "strongly stable" matrices. For our purposes the set of strongly stable matrices, S, is most conveniently defined

as the set of matrices $G^{-1}R(\theta)G$ where $G \in Sp(2n, R)$ and

$$R(\theta)=R(\theta_1,\ldots,\theta_n)=\begin{pmatrix} \cos\theta_1 & & O & \sin\theta_1 & & O \\ & \ddots & & & \ddots & \\ O & & \cos\theta_n & O & & \sin\theta_n \\ \hline -\sin\theta_1 & & O & \cos\theta_1 & & O \\ & \ddots & & & \ddots & \\ O & & -\sin\theta_n & O & & \cos\theta_n \end{pmatrix}$$

where $0 < |\theta_i| < \pi$ and $\theta_i \neq -\theta_j$. Given a curve, $Y(t)$, $0 \le t \le 1$, in $Sp(2n, R)$ with $Y(0)=I$ and $Y(1)$ in S, Gelfand and Lidskii assign an index to $Y(t)$. Then they prove that, given two such curves $Y_1(t)$ and $Y_2(t)$ with $Y_1(1)$ and $Y_2(1)$ in the same connected component of S, there is a continuous deformation $Y(t,\alpha)$, $0 \le t \le 1$, $0 \le \alpha \le 1$ with $Y(t,0)=Y_1(t)$, $Y(t,1)=Y_2(t)$, $Y(0,\alpha)=I$ and $Y(1,\alpha)\in S$ if and only if $Y_1(t)$ and $Y_2(t)$ have the same index.

To adapt this procedure to our situation we define a set of matrices Q—for quasi-stable. A matrix belongs to Q if it has the form $GR(a,\theta)G^{-1}$ where $G \in Sp(2n, R)$ and

$$R(a,\theta)=\begin{pmatrix} 1 & 0 & & a & 0 & \\ & \cos\theta_2 & O & & \sin\theta_2 & O \\ & & \ddots & 0 & & \ddots \\ & O & \cos\theta_n & & O & \sin\theta_n \\ \hline 0 & 0 & & 1 & 0 & \\ & -\sin\theta_2 & O & & \cos\theta_2 & O \\ & & \ddots & 0 & & \ddots \\ & O & -\sin\theta_n & & O & \cos\theta_n \end{pmatrix}$$

where $0 < |\theta_i| < \pi$ and $\theta_i \neq -\theta_j$. We make this definition because we want to assign an index to the following curve. Let

$$z(t)=\begin{bmatrix} | & | & & | & | & | & & | \\ \phi_1(t) & \text{Re}\,\phi_2(t) & \ldots & \text{Re}\,\phi_n(t) & \tilde{\phi}_1(t) & \text{Im}\,\phi_1(t) & \ldots & \text{Im}\,\phi_n(t) \\ | & | & & | & | & | & & | \end{bmatrix}. \tag{B.1}$$

Since $\sigma(\phi_1, \tilde{\phi}_1) = 1$, $\sigma(\operatorname{Re}\phi_i, \operatorname{Re}\phi_j) = \sigma(\operatorname{Im}\phi_i, \operatorname{Im}\phi_j) = \sigma(\phi_i, \phi_1) = \sigma(\bar{\phi}_i, \tilde{\phi}_1) = 0$ and $\sigma(\operatorname{Re}\phi_i, \operatorname{Im}\phi_j) = \delta_{ij}$ for $2 \leq i, j \leq n$, we see that $Z(t) \in \operatorname{Sp}(2n, R)$. Moreover, defining $Y_0(t) = Z(t)Z(0)^{-1}$ — we see that $Y_0(0) = I$ and $Y_0(T) \in Q$.

Continuing to imitate [6], we assign an index to a curve $Y(t)$ in $\operatorname{Sp}(2n, R)$ with $Y(0) = I$ and $Y(1) \in Q$ as follows. By assumption there is a G such that $Y(1) = GR(a, \theta)G^{-1}$ and we define $\tilde{Y}(t) = G^{-1}Y(t)G$. Taking the polar decomposition of $\tilde{Y}(t)$ we have $\tilde{Y}(t) = P(t)0(t)$. It follows from the uniqueness of polar decomposition that $P(t)$ and $0(t)$ are in $\operatorname{Sp}(2n, R)$, and hence

$$0(t) = \begin{pmatrix} 0_1(t) & 0_2(t) \\ -0_2(t) & 0_1(t) \end{pmatrix}$$

where $0_1(t) + i0_2(t) = U(t)$ is unitary. We extend $\tilde{Y}(t)$ to a closed curve by defining $\tilde{Y}(t) = R((2-t)a, (2-t)\theta)$ for $1 \leq t \leq 2$. Then the index of $Y(t)$ is defined as

$$n_Y = \frac{1}{2\pi} \int\limits_0^2 d(\arg \det U(t)).$$

From this definition one computes that

$$n_Y = \frac{1}{2\pi} \int\limits_0^1 d(\arg \det U(t)) - \frac{1}{2\pi} \sum\limits_{j=2}^n \theta_j - \frac{1}{2\pi} f(a). \tag{B.2}$$

Since $e^{i\theta_2}, \ldots, e^{i\theta_n}$ are the eigenvalues of $Y(1)$ belonging to eigenspaces on which $\langle , \rangle$ (see Appendix A) is positive, we see $\sum\limits_{j=2}^n \theta_j$ is uniquely determined by $Y(1)$. It is not yet clear, however, that n_Y is independent of the choice of G and a.

There are now two tasks remaining. We need to show that two curves with end points in the same component of Q can be deformed into each other through curves beginning at I and ending in Q if and only if their indices are the same— which shows that n_Y is independent of G and a. We also need to show that $n_{Y_0} = p$, where p is the integer which appeared in formula (7'). The first objective requires the following lemma, which plays the role of Lemma 4 in [6].

Lemma B.1. *Given a closed curve $\gamma(t)$, $0 \leq t \leq 1$ in Q, $\gamma(t)$ is contractible in $\operatorname{Sp}(2n, R)$.*

Proof. The only difficulty in this proof is presented by the fact that whereas each point on $\gamma(t)$ has a representation $GR(a, \theta)G^{-1}$, it is not possible to choose these representations depending continuously on t (this was apparently overlooked in [6]). However, if we let $S_1(t)$, $S_+(t)$ and $S_-(t)$ denote the (generalized) eigenspace of $\gamma(t)$ belonging to eigenvalue 1, the span of the eigenvectors of $\gamma(t)$ belonging to $e^{i\theta_j(t)}$, $j = 2, \ldots, n$, and the span of the eigenvectors of $\gamma(t)$ belonging to $e^{-i\theta_j(t)}$, $j = 2, \ldots, n$, it follows from the condition $\theta_i \neq -\theta_j$ in the definition of Q that these spaces depend continuously on t. Since $\langle , \rangle$ is positive definite on $S_+(t)$, we can choose an orthonormal basis $\phi_2(t), \ldots, \phi_n(t)$ with respect to $\langle , \rangle$ for $S_+(t)$ depending continuously on t with $\phi_j(0) = \phi_j(1)$ by the Gram-Schmidt procedure. We pick a continuous nonzero real vector in $S_1(t)$, $\phi_1(t)$. Since σ is nondegenerate on $S_1(t)$, there is a unique $\tilde{\phi}_1(t)$ such that $\sigma(\phi_1(t), \tilde{\phi}_1(t)) = 1$ and hence $\tilde{\phi}_1(t)$ is continuous.

The vectors $\bar{\phi}_2(t), \ldots, \bar{\phi}_n(t)$ form a basis for $S_-(t)$. Defining $G(t)$ to be the symplectic matrix we have already written down in (B.1), we see

$$G^{-1}(t)\gamma(t)G(t) = \begin{pmatrix} a_1(t) & 0 & a_2(t) & 0 \\ 0 & A_1(t) & 0 & A_2(t) \\ a_3(t) & 0 & a_4(t) & 0 \\ 0 & -A_2(t) & 0 & A_1(t) \end{pmatrix}$$

where

$$a(t) = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \quad \text{and} \quad A(t) = \begin{pmatrix} A_1(t) & A_2(t) \\ -A_2(t) & A_1(t) \end{pmatrix}$$

are 2×2 and $(2n-2) \times (2n-2)$ symplectic respectively. Hence $A(t)$ is orthogonal. The eigenvalues of $a(t)$ are both 1 and the eigenvalues of $U_0(t) = A_1(t) + iA_2(t)$ are $e^{i\theta_2(t)}, \ldots, e^{i\theta_n(t)}$ and hence never 1.

The set of real 2×2 matrices with both eigenvalues equal to 1 is equal to

$$\left\{ \begin{bmatrix} a_1 & a_2 \\ \dfrac{(2-a_1)a_1-1}{a_2} & 2-a_1 \end{bmatrix} : a_1 \in R, a_2 \neq 0 \right\} \cup \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

and this set is contractible. Thus we can deform $a(t)$ through symplectic matrices to the identity. A closed curve in the group $U(n)$ of $n \times n$ unitary matrices is contractible if and only if $\int_0^1 d(\arg \det U(t)) = 0$, see [6, p. 164]. Since $(1-\alpha)U_0(t) - \alpha I$, $0 \leq \alpha \leq 1$ is a deformation of $U_0(t)$ to $-I$ in $GL(n, C)$ it follows that

$$\int_0^1 d(\arg \det U_0(t)) = 0.$$

Hence we can deform $U_0(t)$ to the identity through unitary matrices. This gives a deformation of $A(t)$ through symplectic matrices to the identity and completes the proof of this lemma.

We also require the following lemma

Lemma B.2. *Two matrices $R(a_1, \theta_1)$ and $R(a_2, \theta_2)$ lying in the same component of Q can be joined by a curve $\gamma(t)$, $0 \leq t \leq 2$ lying in Q with the following property: writing $\gamma(t)$ in polar decomposition and forming the matrix $U_\gamma(t)$ as we did earlier,*

$$\int_0^2 d(\arg \det U_\gamma(t)) = \sum_{i=2}^n \theta_{i,2} - \theta_{i,1} + f(a_2) - f(a_1). \tag{B.3}$$

Here $f(a)$ is the function in formula (B.2), i.e.

$$f(a) = \int_0^1 d(\arg u(t))$$

where $u(t)$ is the 1×1 unitary matrix arising from the polar decomposition of
$\begin{pmatrix} 1 & ta \\ 0 & 1 \end{pmatrix}$.

Proof. Since $R(a_1, \theta_1)$ and $R(a_2, \theta_2)$ lie in the same component of Q if and only if $R(0, \theta_1)$ and $R(0, \theta_2)$ do, it will suffice to show we can join $R(0, \theta_1)$ to $R(0, \theta_2)$ with a curve $\gamma(t)$ in Q satisfying

$$\int_0^2 d(\arg \det U_\gamma(t)) = \sum_{i=2}^n \theta_{i,2} - \theta_{i,1}.$$

As is pointed out in [6] to lie in the same component of Q, $R(0, \theta_1)$ and $R(O, \theta_2)$ must have the same "signature". For any matrix in Q, the signature is an $(n-1)$-tuple of $+$'s and $-$'s with a $+$ in the j-th place if the j-th largest θ (in absolute value) is positive and a minus if it is negative. From this definition it is clear that if $R(0, \theta_1)$ and $R(0, \theta_2)$ belong to the same component of Q we can choose $\theta(t)$, $0 \le t \le 1$ such that $R(0, \theta(t)) \in Q$, $\theta_1 = \theta(0)$ and $\theta(1)$ is just a rearrangement of θ_2. Hence there is a $G \in \mathrm{Sp}(2n, R) \cap O(2n)$ such that $GR(0, \theta(1))G^{-1} = R(0, \theta_2)$. Since $\mathrm{Sp}(2n, R) \cap O(2n)$ is connected we can choose $G(t)$ in $\mathrm{Sp}(2n, R) \cap O(2n)$ such that $G(1) = I$, $G(2) = G$. Then, defining

$$\gamma(t) = \begin{cases} R(0, \theta(t)), & 0 \le t \le 1 \\ G(t)R(0, \theta(1))G(t)^{-1}, & 1 \le t \le 2 \end{cases}$$

the conclusion of this lemma follows.

Now, given two curves $Y_1(t)$, $Y_2(t)$, $0 \le t \le 1$, connecting the identity and points in the same component of Q with $n_{Y_1} = n_{Y_2}$, we wish to show $Y_1(t)$ can be deformed into $Y_2(t)$ through curves with end points in Q. Clearly, if $Y_1(1) = G_1 R_1 G_1^{-1}$ and $Y_2(1) = G_2 R_2 G_2^{-1}$, it will suffice to deform $\tilde{Y}_1(t) = G_1^{-1} Y_1(t) G_1$ into $\tilde{Y}_2(t) = G_2^{-1} Y_2(t) G_2$. By Lemma B.2 we can choose $\gamma(t)$ in G, satisfying (B.3) with $\gamma(1) = R_1$ and $\gamma(2) = R_2$. To show that $\tilde{Y}_1(t)$ can be deformed to $\tilde{Y}_2(t)$ through curves connecting the identity to points in Q, it suffices to show that the curve $\Gamma(t)$ given by $\tilde{Y}_1(t)$ for $0 \le t \le 1$, $\gamma(t-1)$ for $1 \le t \le 3$ and $\tilde{Y}_2(4-t)$, $3 \le t \le 4$ is contractible in $\mathrm{Sp}(2n, R)$. In [6, pp. 164–165] it is shown that a curve is contractible in $\mathrm{Sp}(2n, R)$ if and only if the corresponding curve $U(t)$ satisfies

$$0 = \int d(\arg \det U(t)).$$

We have for the curve $\Gamma(t)$

$$\int_0^3 d(\arg \det U_\Gamma(t)) = \left[2\pi n_{Y_1} + \sum_{j=2}^n \theta_{j,1} + f(a_1) \right]$$

$$+ \left[f(a_2) - f(a_1) + \sum_{j=2}^n \theta_{j,2} - \theta_{j,1} \right] - \left[2\pi n_{Y_2} + \sum_{j=2}^n \theta_{j,2} + f(a_2) \right].$$

This vanishes because we assumed $n_{Y_1} = n_{Y_2}$.

Conversely suppose $Y_1(t)$ can be deformed into $Y_2(t)$. Then, assuming $Y_1(1) = G_1 R_1 G_1^{-1}$ and $Y_2(1) = G_2 R_2 G_2^{-1}$, since $\mathrm{Sp}(2n, R)$ is connected, it follows that $\tilde{Y}_1(t) = G_1^{-1} Y_1(t) G_1$ can be deformed into $\tilde{Y}_2(t) = G_2^{-1} Y_2(t) G_2$. The end points of the curves in this deformation $Y(1, \alpha)$, $0 \le \alpha \le 1$ form a curve in Q connecting R_1

and R_2. Since the closed curve $\Gamma(t)$, given by $Y^{\sim}_1(t)$ for $0 \leq t \leq 1$, $Y(1, t-1)$ for $1 \leq t \leq 2$, and $Y^{\sim}_2(3-t)$ for $2 \leq t \leq 3$ is contractible by hypothesis, it follows from (B.3) that $n_{Y_1} = n_{Y_2}$ if

$$\int_1^2 d(\arg \det U_\Gamma(t)) = \left[f(a_2) - f(a_1) + \sum_{j=2}^n \theta_{j,2} - \theta_{j,1} \right].$$

In other words, if we let $\gamma(t)$ be the curve joining R_1 and R_2 given to us by Lemma B.2, it suffices to show $\gamma_0(t)$ given by $Y(1, t)$ for $0 \leq t \leq 1$, and $\gamma(3-t)$ for $1 \leq t \leq 3$ is contractible. However, $\gamma_0(t)$ is contractible by Lemma B.1.

Now we are left with the problem of showing $n_{Y_0} = p$. To begin we extend the curve $Y_0(t)$ by defining $Y_0(t) = Z(0)R((T+1-t)a, \theta)Z(0)^{-1}$ for $T \leq t \leq T+1$, where $Z(T) = Z(0)R(a, \theta)$. This extension does not change the index of $Y_0(t)$. Hence, setting $Y^{\sim}_0(t) = Z(0)^{-1} Y_0(t)Z(0)$, making the polar decomposition $Y^{\sim}_0(t) = P(t)O(t)$, and setting $U(t) = O_1(t) + iO_2(t)$, we have

$$\int_0^{T+1} d(\arg \det U(t)) = 2\pi n_{Y_0} + \sum_{j=2}^n \theta_j. \tag{B.4}$$

We will now show that, if we set

$$Y_0(t)Z(0) = \begin{pmatrix} Z_1(t) & Z_2(t) \\ Z_3(t) & Z_4(t) \end{pmatrix} = Z(0) Y^{\sim}_0(t),$$

then

$$\int_0^{T+1} d(\arg \det U(t)) = \int_0^{T+1} d(\arg \det(Z_1(t) + iZ_2(t))). \tag{B.5}$$

This requires three steps. First, letting $Z(0)$ have the polar decomposition $P_0 O_0$, we observe that $O_0 Y^{\sim}_0(t) = O_0 P(t)O_0^T O_0 O(t)$. Hence $O_0 Y^{\sim}_0(t)$ has the polar decomposition $P^{\sim}(t)O^{\sim}(t)$ where $O^{\sim}(t) = O_0 O(t)$, and the corresponding $U^{\sim}(t)$ equals $U_0 U(t)$. Thus we conclude

$$\int_0^{T+1} d(\arg \det U(t)) = \int_0^{T+1} d(\arg \det U^{\sim}(t)). \tag{B.6}$$

Secondly, we set

$$O_0 Y^{\sim}_0(t) = \begin{pmatrix} Z^{\sim}_1(t) & Z^{\sim}_2(t) \\ Z^{\sim}_3(t) & Z^{\sim}_4(t) \end{pmatrix}, \qquad P^{\sim}(t) = \begin{pmatrix} P^{\sim}_1 & P^{\sim}_2 \\ P^{\sim}_2 & P^{\sim}_3 \end{pmatrix}$$

and observe that

$$Z^{\sim}_1(t) + iZ^{\sim}_2(t) = (P^{\sim}_1(t) + iP^{\sim}_2(t))U^{\sim}(t).$$

Hence

$$\det(Z^{\sim}_1 + iZ^{\sim}_2) = \det P^{\sim}_1 \det(1 + iP^{\sim -1}_1 P^{\sim}_2) \det U^{\sim}(t).$$

Moreover, since $O_0 Y^{\sim}_0(T+1) = O_0 R(0, \theta)$, we see $P^{\sim}(0) = P^{\sim}(T+1) = I$. Also, since $P^{\sim}(t)$ is a positive symplectic matrix, $P^{\sim -1}_1$ exists and $P^{\sim -1}_1 P^{\sim}_2$ is symmetric. Now

we claim $I + iP_1^{\sim -1}(t)P_2(t)$ is a closed contractible curve in $\mathrm{GL}(n, C)$. Suppose for $\alpha \in [0, 1]$

$$v + i\alpha P_1^{\sim -1}(t)P_2^{\sim}(t)v = 0 .$$

Then $|v|^2 + i\alpha \bar{v} \cdot P_1^{\sim -1}(t)P_2^{\sim}(t)v = 0$ which implies $v = 0$. Thus

$$\int_0^{T+1} d(\arg \det(1 + iP_1^{\sim -1}P_2^{\sim}) = 0$$

and we conclude

$$\int_0^{T+1} d(\arg \det(Z_1^{\sim}(t) + iZ_2^{\sim}(t)) = \int_0^{T+1} d(\arg \det U^{\sim}(t)) . \tag{B.7}$$

Third, setting $P_0^{-1} = \begin{pmatrix} P_1 & P_2 \\ P_2^t & P_3 \end{pmatrix}$ we have

$$Z_1^{\sim}(t) + iZ_2^{\sim}(t) = P_1(Z_1(t) + iZ_2(t)) + P_2(Z_3(t) + iZ_4(t)) .$$

Thus

$$\det(Z_1(t) + iZ_2(t)) = \det P_1 \det(1 + P_1^{-1}P_2 M(t)) \det(Z_1(t) + iZ_2(t))$$

where

$$M(t) = (Z_3(t) + iZ_4(t))(Z_1(t) + iZ_2(t))^{-1} .$$

Since $O_0 Y_0(t)$ is symplectic, it follows that $M(t)$ is symmetric with a positive imaginary part. Also, since

$$\begin{pmatrix} Z_1(T+1) & Z_2(T+1) \\ Z_3(T+1) & Z_4(T+1) \end{pmatrix} = \begin{pmatrix} Z_1(0) & Z_2(0) \\ Z_3(0) & Z_4(0) \end{pmatrix} R(0, \theta) ,$$

it follows that $M(T+1) = M(0)$. Now we claim $I + P_1^{-1}P_2 M(t)$, $0 \le t \le T+1$ is a contractible closed curve in $\mathrm{GL}(n, C)$. Suppose for $\alpha \in [0, 1]$

$$v + \alpha P_1^{-1}P_2 M(t)v = 0 .$$

Then $M(t)v \cdot \bar{v} + \alpha M(t)v \cdot P_1^{-1}P_2 \overline{M(t)v} = 0$ and we conclude $\mathrm{Im}\, M(t)v \cdot \bar{v} = 0$ which implies $v = 0$. Thus

$$\int_0^{T+1} d(\arg \det(1 + P_1^{-1}P_2 M(t))) = 0$$

and we have

$$\int_0^{T+1} d(\arg \det(Z_1 + iZ_2) = \int_0^{T+1} d(\arg \det(Z_1^{\sim} + iZ_2^{\sim}) . \tag{B.8}$$

Combining (B.6)–(B.8) gives (B.5).

Unfortunately we are still not quite finished. Combining (B.4) and (B.5) we have

$$2\pi n_{Y_0} + \sum_{j=2}^n \theta_j = \int_0^{T+1} d(\arg \det(Z_1(t) + iZ_2(t))) .$$

From (7') we have

$$2\pi p + \sum_{j=2}^{n} \theta_j = \int_0^T d(\arg \det W(t))$$

where $W(t)$ is the matrix $Z_1(t) + iZ_2(t)$ with the first column replaced by $y^1(t)$. To correct this final discrepancy we extend $Z_1(t) + iZ_2(t)$ to a closed curve by defining $T+1 \leq t \leq T+2$

$$Z_1(t) + iZ_2(t) = (Z_1(0) + iZ_2(0)) \begin{bmatrix} 1 & & & & O \\ & e^{i(T+2-t)\theta_2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ O & & & & e^{i(T+2-t)\theta_n} \end{bmatrix}.$$

Then

$$\int_0^{T+2} d(\arg \det(Z_1(t) + iZ_2(t))) = 2\pi n_{Y_0}.$$

Now let $W(t,s)$ be the matrix with first column $y^1(t) + is\tilde{y}^1(t)$ for $0 \leq t \leq T$, first column $y^1(0) + is(\tilde{y}^1(0) + (T+1-t)ay^1(0))$ for $T \leq t \leq T+1$, first column $y^1(0) + is\tilde{y}^1(0)$ for $T+1 \leq t \leq T+2$. The proof given in the introduction to show that $Y(t)$ was nonsingular applies here to show $T(t,s)$ is nonsingular for $0 \leq t \leq T+2$, $0 \leq s \leq 1$. Thus

$$2\pi n_{Y_0} = \int_0^{T+2} d(\arg \det W(t,0))$$

$$= \left[2\pi p + \sum_{j=2}^{n} \theta_j \right] + [0] + \left[-\sum_{j=2}^{n} \theta_j \right].$$

Thus we have $n_{Y_0} = p$ and are finished.

References

1. Arnaud, J. A.: Hamiltonian theory of beam mode propagation. Progress in Optics XI. Ed. E. Wolf. North Holland, pp. 249—304, 1973
2. Arnol'd, V. I.: Modes and quasimodes. Funct. Anal. Appl. **6**, 94—101 (1972)
3. Babich, V. M., Lazutkin, V. F.: Eigenfunctions concentrated near a closed geodesic, pp. 9—18. Topics in Mathematical Physics, Vol. 2. Ed. M. S. Birman. New York: Consultants' Bureau 1968
4. Colin de Verdière, Y.: Quasi-modes des variétés Riemanniennes. I. To appear
5. Duistermaat, J. J.: Oscillatory integrals, Lagrange immersions and unfolding of singularities. Comm. Pure App. Math. **27**, 207—281 (1974)
6. Gelfand, I. M., Lidskii, V. B.: On the structure of the regions of stability of linear canonical systems of differential equations with periodic coefficients. A.M.S. Transl. (Series 2) **8**, 143—181 (1958)
7. Guillemin, V.: Symplectic spinors and partial differential equations. C.N.R.S. Symposium on Symplectic Geometry and Mathematical Physics, Aix-en-Provence, 1974
8. Guillemin, V., Weinstein, A.: Eigenvalues associated with a closed geodesic. Bull. A.M.S. **82**, 92—94 (1976)

9. Hörmander, L.: On the existence and regularity of solutions of linear pseudo-differential equations. Enseignement Math. **17**, 99—103 (1971)
10. Howland, J.: Spectral Concentration and Virtual Poles, I, II. Amer. J. Math. **91**, 1106—1126 (1969); A.M.S. Trans. **162**, 141—156 (1971)
11. Keller, J. B.: Corrected Bohr-Sommerfeld quantum conditions for nonseparable systems. Ann. Phys. **4**, 180—188 (1958)
12. Krein, M. G.: Generalization of certain investigations of A. M. Lyapunov on linear differential equations with periodic coefficients. Dokl. Akad. Nauk SSSR (N.S.) **73**, 445—448 (1950) (Russian)
13. Maslov, V. P.: Theory of Perturbations and Asymptotic Methods (in Russian), Moscow State University, Moscow: 1965 translated as Théorie des Perturbations et Méthodes Asymptotiques. Paris: Dunod, Gauthier-Villars 1972
14. Moser, J.: On the generalization of a theorem of A. Liapunoff. Comm. Pure Appl. Math. **XI**, 257—271 (1958)
15. Ralston, J.: Construction of approximate eigenfunctions of the Laplacian concentrated near stable closed geodesics. To appear in J. Diff. Geometry
16. Siegel, C. L., Moser, J.: Lectures on Celestial Mechanics. Berlin, Heidelberg, New York: Springer 1971
17. Voros, A.: The WKB-Maslov method for nonseparable systems. C.N.R.S. Symposium on Symplectic Geometry and Mathematical Physics, Aix-en-Provence, 1974, and, Semiclassical Approximations. Ann. Inst. Henri Poincaré **24**, 31—90 (1976)
18. Weinstein, A.: On Maslov's quantization condition, Symposium on Fourier Integral Operators, Nice, 1974. Berlin, Heidelberg, New York: Springer 1974
19. Weyl, H.: The Classical Groups. Princeton, N. J.: Princeton University Press 1946