Journal of Statistical Physics, Vol. 68, Nos. 1/2, 1992

The Van Vleck Formula, Maslov Theory, and Phase Space Geometry

Robert G. Littlejohn[1]

Received December 13, 1991

The Van Vleck formula is an approximate, semiclassical expression for the quantum propagator. It is the starting point for the derivation of the Gutzwiller trace formula, and through this, a variety of other expansions representing eigenvalues, wave functions, and matrix elements in terms of classical periodic orbits. These are currently among the best and most promising theoretical tools for understanding the asymptotic behavior of quantum systems whose classical analogs are chaotic. Nevertheless, there are currently several questions remaining about the meaning and validity of the Van Vleck formula, such as those involving its behavior for long times. This article surveys an important aspect of the Van Vleck formula, namely, the relationship between it and phase space geometry, as revealed by Maslov's theory of wave asymptotics. The geometrical constructions involved are developed with a minimum of mathematical formalism.

KEY WORDS: Van Vleck formula; Maslov theory; WKB theory; semiclassical mechanics; quantum chaos; Gutzwiller trace formula.

1. INTRODUCTION

The Van Vleck formula is a semiclassical approximation for the usual propagator in quantum mechanics,

$$K(x'', t''; x', t') = \langle x'' | \, U(t'', t') \, | x' \rangle \, \Theta(t'' - t') \tag{1.1}$$

where the singly primed variables (x', t') represent some initial position and time, the doubly primed variables (x'', t'') represent some final position and time, $U(t'', t')$ is the unitary time evolution operator for some quantum system (possibly time-dependent), and Θ is the unit step function. The

[1] Department of Physics, University of California, Berkeley, California 94720.

0022-4715/92/0700-0007$06.50/0

Van Vleck formula is most commonly applied to the nonrelativistic Schrödinger equation for scalar particles, as we shall do here.

The Van Vleck formula is the starting point for a sequence of derivations, approximations, and intuitive leaps which take one from exact quantum expressions to a variety of results expressing energy eigenvalues and their correlations, wave functions, and matrix elements in terms of classical periodic orbits. The first stage in this process is the Gutzwiller trace formula,[1] which expresses the density of states of a quantum system as a sum over the periodic orbits of the corresponding classical system. Various resummation techniques can be applied to the Gutzwiller trace formula, yielding one of the principal bodies of theoretical methods available for the analysis of quantum systems whose classical analogs are chaotic. Considerable progress has been made along these lines in recent years, and new methods have appeared for understanding a variety of systems in atomic, molecular, and nuclear physics.[2] Many of these results are in a sense improvements on the Gutzwiller trace formula, having advantages in terms of their convergence properties. Nevertheless, these results all depend logically on the Gutzwiller trace formula and, through it, on the Van Vleck formula, since direct methods of derivation are not known. Furthermore, there remain important questions concerning the Van Vleck formula itself, such as its long-time validity.[3] Therefore a proper understanding of the Van Vleck formula is more important than ever.

This article surveys an important aspect of this question, namely the relation between the Van Vleck formula and geometrical structures in the classical phase space. This is a subject developed in large measure by Maslov and his co-workers,[4] whose theories have been nicely reviewed in the context of time-independent and scattering problems by Delos.[5] Another clear introduction to this theory has been given by Percival.[6] In this article we will focus primarily on the Van Vleck formula and its derivation, keeping in mind a primary application of this formula, the derivation of the Gutzwiller trace formula. The latter derivation is notorious for its difficulty, and this survey is intended in part to illuminate and remove part of this difficulty. The specific manner in which geometrical ideas can be applied to the Gutzwiller trace formula and to other trace formulas has been explored in considerably more detail in ref. 7.

There exists a substantial body of mathematical literature on wave asymptotics, of which the books by Guillemin and Sternberg[8] and Leray[9] are important examples. Much of this literature is, however, rather technical, and as a result it has been quite decoupled from and has had little impact on semiclassical studies in physical applications. There are even examples of important results known in the physical literature, such as the Gutzwiller trace formula itself, which apparently have been rederived

in the mathematical literature, with even today scarcely an acknowledgment of the parallel (and earlier) development. The books by Maslov and by Maslov and Fedoriuk[4] are relatively more accessible than the other mathematical references in this area, and pay particular attention to the Van Vleck formula. Although they are still rather technical, they are the primary references for this survey.

In this article we will attempt to convey essential geometrical ideas in an intuitive and plausible way, without excessive mathematical formalism on rigorous proofs. The book by Arnold[10] is a standard reference on the background material for this article, and contains in addition considerable material on wave asymptotics in general and the Van Vleck formula in particular. One of the main purpose of the geometrical approach to wave asymptotics is to understand properly the covariance of semiclassical mechanics under the canonical transformations of classical mechanics. This is a subject which is perhaps best known in the physical literature on account of the work of Miller,[11] which explores the self-consistency of semiclassical methods in different representations, and which has had considerable influence. Covariance of semiclassical mechanics under canonical transformations is of particular relevance to the Gutzwiller trace formula, in which transformations are performed (such as from the time representation to the energy representation), and in which the final results (the trace formula) are expressed in terms of canonical invariants, after a derivation which proceeds through representation-dependent intermediaries.

Nowadays it is popular to derive the Van Vleck formula as the semiclassical limit of the Feynman path integral for the propagator.[12] This approach is natural in view of the fundamental role played by path integrals in many areas of physics, and in view of the great physical appeal of path integrals and the physical intuition they provide. But Van Vleck's original derivation[13] was based on a different approach, namely, multidimensional, time-dependent WKB or phase integral theory. The latter approach has a number of advantages over that based on the path integral, such as the fact that it is easily generalized to cover a wide class of wave equations (not just those having the standard kinetic-plus-potential form of quantum mechanics), and the fact that representation covariance is (with proper understanding) essentially built in. It is the WKB approach which we will follow in this survey.

2. BACKGROUND ON THE PROPAGATOR

In this section we present some background on the propagator, setting up the derivation of its semiclassical limit in terms of WKB theory. We do this by expressing the propagator as the solution of an initial value

problem in quantum mechanics, which, when subjected to a semiclassical approximation, leads to an initial value problem in time-dependent, multidimensional WKB theory. The purely quantum mechanical side of this picture is a standard subject in Green's function theory,[14] which we now summarize.

The propagator can be defined as the solution of the inhomogeneous Schrödinger equation,

$$\left(H'' - i\hbar \frac{\partial}{\partial t''}\right) K(x'', t''; x', t') = -i\hbar\, \delta(t'' - t')\, \delta(x'' - x') \tag{2.1}$$

subject to the condition $K = 0$ for $t'' < t'$. Here the double prime on H'' indicates that the quantum operators in the Hamiltonian act on x'', and that any explicit time dependence in the Hamiltonian is evaluated at t''. For the most part, we will not distinguish notationally between the one-dimensional and the multidimensional cases. For example, the symbol x should be interpreted as an f-dimensional vector, where f is the number of degrees of freedom, and similarly for p, etc. Scalar products, where they occur, are usually obvious, as in $p\,dx$, which stands for $\sum p_i\, dx_i$. The distinction between x and q is that the former suggests rectangular coordinates, whereas the latter suggests generalized coordinates on phase space.

Because the inhomogeneous term in Eq. (2.1) vanishes for $t'' \neq t'$, the propagator satisfies the ordinary, homogeneous Schrödinger equation in the double-primed variables for $t'' > t'$ (and also for $t'' < t'$, where K vanishes). Therefore the determination of K for $t'' > t'$ can be viewed as an initial value problem, in which the initial conditions can be taken as the value of K for small positive times, i.e., for $t'' = t' + \varepsilon$, with $\varepsilon \to 0^+$.

To find these initial conditions, we introduce the two-time unitary evolution operator $U(t'', t')$, defined as the operator solution of

$$i\hbar \frac{\partial U(t'', t')}{\partial t''} = H(t'')\, U(t'', t') \tag{2.2}$$

subject to the initial condition $U = \text{Identity}$ at $t'' = t'$.[15] Then we can write K in terms of the x-space matrix elements of U as shown in Eq. (1.1), as follows by directly substituting Eq. (1.1) into Eq. (2.1), and noting that K does indeed vanish for $t'' < t'$. If we now let t'' approach t' from above, then Eq. (1.1) shows that

$$\lim_{\varepsilon \to 0^+} K(x'', t' + \varepsilon; x', t') = \delta(x'' - x') \tag{2.3}$$

We thereby obtain a simple way of thinking about the propagator: for $t'' > t'$, $K(x'', t''; x', t')$ is the solution $\psi(x'', t'')$ of the time-dependent Schrödinger equation, subject to the initial condition $\psi(x'', t'') = \delta(x'' - x')$ at $t'' = t'$.

2.1. The Initial Value Problem in WKB Theory

We now consider the initial value problem in WKB theory, in order to solve for K in the semiclassical approximation. We begin by considering time-dependent WKB theory from a general standpoint.

Suppose we are given an initial wave function of the form

$$\psi(x, t') = \psi_0(x) = A_0(x) \exp[(i/\hbar)\, S_0(x)] \tag{2.4}$$

where for now we suppress the primes on x. The initial amplitude is $A_0(x)$, assumed to be real and positive, and the initial action is $S_0(x)$. This form should properly be regarded as the leading term of an asymptotic expansion in $\hbar$, in which $O(\hbar)$ terms in ψ are neglected. Since the action turns out to be representable in terms of a line integral, we will say that a wave function such as in Eq. (2.4) has the "phase integral" or "WKB form."

We will make the assumption that at a later time $t'' > t'$ the wave function can again be represented in WKB form,

$$\psi(x, t'') = A(x, t'') \exp[(i/\hbar)\, S(x, t'')] \tag{2.5}$$

This assumption may not be valid; as we will discuss more fully below, it turns out that the final wave function can be represented as shown in Eq. (2.5) only for sufficiently short elapsed times $t'' - t'$, after which it must be replaced by a sum of terms of the WKB form. The breakdown of the single-term representation is due to the formation of caustics, and it occurs at times which are classical, i.e., of order $\hbar^0$. For sufficiently longer times, the WKB approach presumably breaks down altogether. For some time now it has been assumed that this breakdown is due to the progressive convolution of manifolds in phase space,[16] which ultimately reach a quantum scale. This assumption has lately been called into question,[3] and it is likely that a more searching analysis would reveal important new insights into the time limitations of WKB theory. Certainly of the two limitations—the short-time limit on the single-term expression, and the longer-time limit on the multiterm expression—the former is more innocuous. We will return to the question of time limitations later; for the moment we will simply proceed with Eq. (2.5) as it stands.

Our immediate problem is to solve for the final amplitude $A(x, t'')$ and action $S(x, t'')$, given the initial values $A_0(x)$, $S_0(x)$. We do this by

substituting Eq. (2.5) into the (double-primed) time-dependent Schrödinger equation, expanding in powers of $\hbar$, and neglecting terms of order $\hbar$ and higher. At lowest order we find the time-dependent Hamilton–Jacobi equation for the action S,

$$H\left(x, \frac{\partial S(x, t'')}{\partial x}, t''\right) + \frac{\partial S(x, t'')}{\partial t''} = 0 \tag{2.6}$$

where the momentum dependence of the Hamiltonian has been replaced by $p = \partial S/\partial x$. This result is verified by standard calculations[15] for Hamiltonians of the form

$$H(x, p, t) = p^2/2m + V(x, t) \tag{2.7}$$

(in any number of degrees of freedom), but it can be justified as well for evolution operators which are quite general functions of (x, p, t) (so long as they have a classical limit).

The result of the expansion at the next order in $\hbar$ is the so-called amplitude transport equation for A, which is conveniently expressed in terms of a quantity ρ, defined by

$$\rho(x, t'') = |A(x, t'')|^2 \tag{2.8}$$

As we shall see later, for small elapsed times, A is real and positive, just like its initial condition A_0, and the absolute value signs are unnecessary. Because of the probabilistic interpretation of the wave function in Eq. (2.4), it is suggestive to interpret ρ as a kind of density of classical particles on configuration space. In terms of ρ, the equation of evolution for A is simply the continuity equation,

$$\frac{\partial \rho(x, t'')}{\partial t''} + \frac{\partial}{\partial x}\left[\rho(x, t'')\, v(x, t'')\right] = 0 \tag{2.9}$$

where in more than one dimension the x derivative is a divergence, and where the velocity field v is given by

$$v(x, t'') = \frac{\partial H(x, p, t'')}{\partial p} \tag{2.10}$$

with p set equal to $\partial S(x, t'')/\partial x$. Since the amplitude transport equation for ρ or A involves S [through the velocity field $v(x, t'')$], it is necessary to solve the Hamilton–Jacobi equation first for S, and then to use this solution in the amplitude transport equation to solve for ρ or A.

We will spend the next section preparing for the solution of the

Hamilton–Jacobi equation. This is a standard topic in most texts on classical mechanics, although it is poorly explained in many of them. For example, it is almost never mentioned that the existence of global solutions of the Hamilton–Jacobi equation depends critically on whether the classical motion is regular or chaotic. We will be especially interested in a geometrical interpretation of the solutions of the Hamilton–Jacobi equation, which is intimately connected with the concept of Lagrangian manifolds. This is an important concept, without which multidimensional WKB theory and semiclassical mechanics cannot properly be understood. We therefore turn now to a discussion of Lagrangian manifolds and the manner in which they emerge from WKB theory. Once this is completed, the solution of the Hamilton–Jacobi equation will be almost immediate.

3. LAGRANGIAN MANIFOLDS AND THE HAMILTON–JACOBI EQUATION

Let us return to the initial action function $S_0(x)$ and its associated momentum field,

$$p = p_0(x) = \frac{\partial S_0(x)}{\partial x} \tag{3.1}$$

which can be viewed as a vector field on the f-dimensional configuration space at the initial time t'. It is useful to think of this field as representing the initial momenta of a swarm of particles; as we shall see, the initial density of these particles should be interpreted as being $\rho_0(x) = |A_0(x)|^2$. This initial momentum field is also associated with an initial velocity field,

$$v_0(x) = \frac{\partial H(x, p, t')}{\partial p} \tag{3.2}$$

with p set to $\partial S_0(x)/\partial x$. This swarm of particles and the associated vector fields constitute the classical or semiclassical interpretation of the initial action function in configuration space.

When viewed in the $2f$-dimensional phase space, the initial swarm of particles lies on an f-dimensional surface, since Eq. (3.1) constitutes f independent constraints on the $2f$ variables (x, p). See Figs. 1 and 2, in which this surface, denoted L_0, is illustrated for one and two degrees of freedom. For two degrees of freedom, some imagination must be used to visualize the 2-dimensional surface embedded in the 4-dimensional phase space. The surface L_0 is the *graph* of the function $p = p_0(x)$, i.e., the set of points in phase space of the form $(x, p_0(x))$.

The f-dimensional surface created in this manner is not arbitrary, but

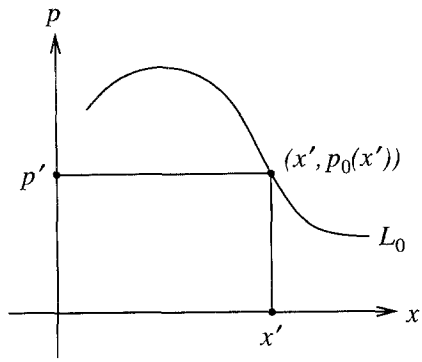

Fig. 1. The initial Lagrangian manifold $p_0(x) = \partial S_0/\partial x$ in one dimension is just a curve.

rather satisfies certain differential constraints, due to the fact that the momentum field $p_0(x)$ is the gradient of a scalar (namely S_0). That is, we have

$$\frac{\partial p_{0i}(x)}{\partial x_j} = \frac{\partial p_{0j}(x)}{\partial x_i} \tag{3.3}$$

for $i, j = 1,..., f$. We will call a momentum field satisfying Eq. (3.3) "curl-free."

We will define a Lagrangian manifold momentarily, but it turns out that any f-dimensional surface created as the graph of a curl-free momentum field is, in fact, a Lagrangian manifold. The converse is not quite true; there exist Lagrangian manifolds for which p cannot be expressed as a

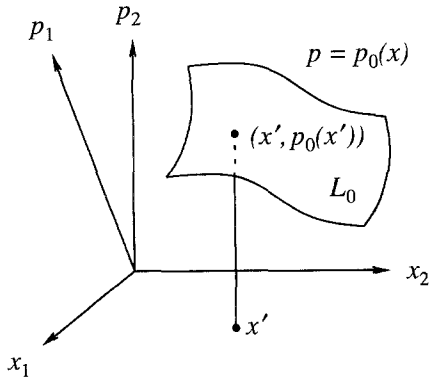

Fig. 2. The initial Lagrangian manifold $p_0(x) = \nabla S_0(x)$ in two degrees of freedom. Note that L_0 is 2-dimensional surface in a phase space of four dimensions.

function of x, or for which the derivatives in Eq. (3.3) diverge or are undefined. Because of these exceptional cases, we cannot take Eq. (3.3) as a definition of a Lagrangian manifold.

The actual definition of a Lagrangian manifold involves the symplectic form, denoted ω, which is an antisymmetric, bilinear operator acting on vectors in phase space.[10] If we let $\delta z_1 = (\delta q_1, \delta p_1)$ and $\delta z_2 = (\delta q_2, \delta p_2)$ be two small displacements in phase space, then the action of the symplectic form on them is defined by

$$\omega(\delta z_1, \delta z_2) = \delta p_1 \cdot \delta q_2 - \delta p_2 \cdot \delta q_1 \tag{3.4}$$

or, in matrix form,

$$\omega(\delta z_1, \delta z_2) = \delta z_1 \cdot \mathbf{J}^{-1} \cdot \delta z_2 \tag{3.5}$$

where $\mathbf{J}$ is the unit symplectic matrix,

$$\mathbf{J} = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \tag{3.6}$$

The matrix $\mathbf{J}$ is antisymmetric and orthogonal, so $\mathbf{J}' = \mathbf{J}^{-1} = -\mathbf{J}$. Note that only two vectors are involved in the definition of the symplectic form, no matter how many dimensions in the phase space. In one degree of freedom, the symplectic form measures the area of the parallelogram spanned by the vectors; in higher degrees of freedom, the symplectic form may be used to define phase space area for 2-dimensional subspaces. The symplectic form is invariant under canonical transformations, in the sense that the value of the right side of Eq. (3.4) is independent of the canonical coordinates used to compute it. Indeed, perhaps the best definition of a canonical transformation is as a transformation which has this property for all vectors $\delta z_1, \delta z_2$.

We now define a Lagrangian manifold as a f-dimensional surface L in the $2f$-dimensional phase space such that at all points (x, p) on L and for all vectors $\delta z_1, \delta z_2$ tangent to L at (x, p), we have (Fig. 3)

$$\omega(\delta z_1, \delta z_2) = 0 \tag{3.7}$$

That is, a Lagrangian manifold is a null surface of dimensionality f, in the sense of the phase space geometry engendered by the symplectic form ω.

Let us examine some of the consequences of this definition. To begin, in one degree of freedom, a Lagrangian manifold must be a 1-dimensional surface, i.e., a curve, in the 2-dimensional phase plane. (See Fig. 4.) Furthermore, since all tangent vectors $\delta z = (\delta x, \delta p)$ at a given point on any curve in a plane are linearly dependent, the antisymmetry of the symplectic

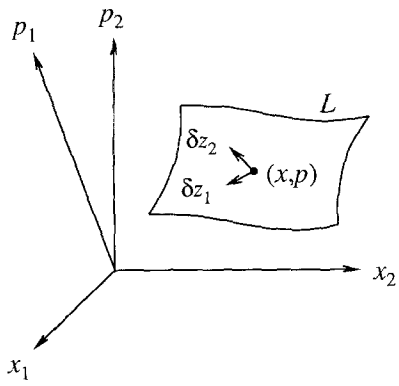

Fig. 3. An f-dimensional surface L in phase space is a Lagrangian manifold if the symplectic area spanned by any two tangent vectors δz_1, δz_2 at any point of the surface is zero. It is a null surface in the symplectic geometry.

form guarantees that $\omega(\delta z_1, \delta z_2)$ always vanishes. Therefore all curves in the 2-dimensional phase plane are Lagrangian manifolds, i.e., the Lagrangian condition imposes no constraints at all in one degree of freedom. The concept of a Lagrangian manifold is really only needed for multidimensional problems.

Another consequence of the definition is that the surfaces $x = \text{const}$ or $p = \text{const}$, in any number of degrees of freedom, are Lagrangian manifolds, because one or the other of the increments δq, δp in Eq. (3.4) is zero. In particular, both configuration space and momentum space, conceived of as subsets of phase space, are Lagrangian manifolds. These statements generalize to any set of canonical coordinates, because the value of the

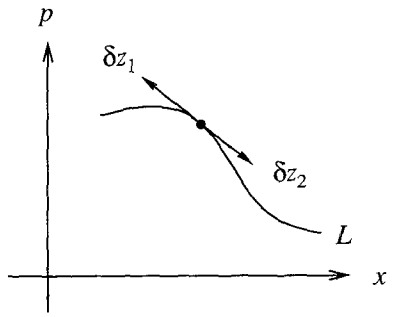

Fig. 4. In one degree of freedom, all tangent vectors δz at a point of a curve are linearly dependent, so all curves in the phase plane are Lagrangian.

symplectic form of Eq. (3.4) is invariant under canonical transformations. Conversely, one can show that every Lagrangian manifold is a constant-q or constant-p surface in some set of canonical coordinates. It is suggestive to think of a Lagrangian manifold as a surface in phase space which is completely "q-ish;" such a surface is also completely "p-ish," because q and p can be interchanged by the canonical transformation $P=q$, $Q=-p$. The point is that a Lagrangian manifold does not, in a sense, have any cross q–p behavior.

A third consequence of the definition is that the graph of any curl-free momentum field, $p=p(x)$, is, in fact, a Lagrangian manifold. To see this, note that any vector $\delta z=(\delta x, \delta p)$ tangent to the surface $p=p(x)$ satisfies a constraint connecting its x and p components, namely,

$$\delta p_i=\sum_j \frac{\partial p_i}{\partial x_j}\,\delta x_j \tag{3.8}$$

Substituting this into Eq. (3.4) and using the symmetry of $\partial p_i/\partial x_j$, we easily find that $\omega(\delta z_1, \delta z_2)=0$.

One of the conditions on a Lagrangian manifold is that it should have dimensionality f, one half of the dimensionality of the phase space. But null surfaces of other dimensionalities also exist, i.e., surfaces whose tangent vectors satisfy Eq. (3.7). Such surfaces are called *isotropic* (the terminology bears no relation to the usual meaning of this word in physics), and are also sometimes important in classical and semiclassical mechanics. There is a limit, however, on the dimensionality of such surfaces: it cannot exceed f, because it turns out that the maximum number of linearly independent vectors which can be pairwise annihilated by the symplectic form is f.

More precisely, if we are given f linearly independent phase space vectors $X_1,\ldots, X_f$ (the subscripts distinguish the vectors, and are not components) such that $\omega(X_k, X_l)=0$ for $k, l=1,\ldots, f$, and another vector Y such that $\omega(Y, X_k)=0$ for $k=1,\ldots, f$, then Y must be a linear combination of the X's. To prove this, consider the two sets of f vectors $(X_1,\ldots, X_f)$ and $(\mathbf{J}X_1,\ldots, \mathbf{J}X_f)$, where $\mathbf{J}$ is given by Eq. (3.6). The first set is linearly independent by hypothesis, and the second set is linearly independent because $\mathbf{J}$ is nonsingular. Further, every vector of the first set is perpendicular to every vector of the second, because $\omega(X_k, X_l)=-X_k\cdot(\mathbf{J}X_l)=0$. Therefore the two sets span perpendicular subspaces of dimensionality f, and together span the $2f$-dimensional space of all possible tangent vectors in phase space. If now $\omega(Y, X_k)=0$, then Y is perpendicular to all the $\mathbf{J}X_k$, and must therefore lie in the subspace spanned by the X_k. This proves the theorem. As a result of this theorem, we can characterize a Lagrangian manifold as an isotropic manifold of maximum dimensionality (namely f).

3.1. Lagrangian Manifolds and Caustics

We have just shown that the graph of a curl-free momentum field is always a Lagrangian manifold in phase space. The converse is not quite true, because many Lagrangian manifolds contain points at which the derivatives in the curl-free condition, Eq. (3.3), are not defined. Such points are associated with caustics, and are therefore important in WKB theory. We will now examine the conditions under which the derivatives of Eq. (3.3) are not defined, and thereby clarify the geometrical meaning of caustics.

Since Lagrangian manifolds are f-dimensional, it is always possible to impose f coordinates, say $(u_1,..., u_f)$, on one of them. It is an implication of the word "coordinate" that the u's provide a unique labeling of points on the manifold, which we assume to be smooth. Therefore the two f-dimensional vectors x and p can be regarded as functions of u on the Lagrangian manifold, $x=x(u)$, $p=p(u)$, and these functions are smooth. Now it may or may not happen that the f variables x are locally invertible functions of u; the condition for invertibility is that the determinant of the Jacobian $\partial x/\partial u$ should not vanish. If this condition is satisfied, we can write $u=u(x)$, and then by substitution, $p(x)=p(u(x))$. In this way, p becomes a function of x on the Lagrangian manifold when the determinant of $\partial x/\partial u$ does not vanish. On the other hand, if we approach a point of a Lagrangian manifold at which the determinant of $\partial x/\partial u$ does vanish, then the matrix occurring in Eq. (3.3),

$$\frac{\partial p_i}{\partial x_j}=\sum_k \frac{\partial p_i}{\partial u_k}\frac{\partial u_k}{\partial x_j} \tag{3.9}$$

must behave badly. In one dimension, it can do so only by diverging, since the derivative dp/du must be nonzero when dx/du is zero; in higher dimensions, its behavior may be more complicated, since some of the eigenvalues of $\partial p/\partial u$ may vanish at the same place that some of the eigenvalues of $\partial x/\partial u$ also vanish.

The set of points on a Lagrangian manifold where the determinant of $\partial x/\partial u$ is zero is called the *singular set*, and the points of configuration space which lie directly below them are the *caustic points*. We will use somewhat loose terminology and refer to points of either kind as caustic points. The geometrical meaning of this condition, as illustrated in Fig. 5, is that the projection of the Lagrangian manifold onto configuration space has an edge at a caustic point, or, as is commonly said, the projection is singular there. One can also see that the f-dimensional tangent plane to the Lagrangian manifold at the caustic point is "vertical" in one or more of its

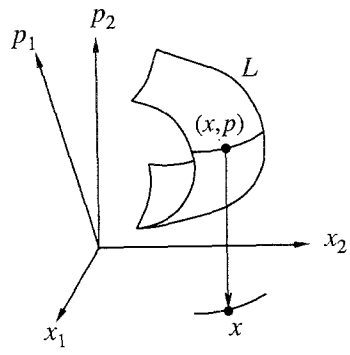

Fig. 5. The singular set is the set of points (x, p) on the Lagrangian manifold at which the tangent plane becomes "vertical" to configuration space in one or more of its directions. A configuration point x below the singular set is a caustic point.

dimensions, i.e., that there exist vectors in this tangent plane which are annihilated when projected onto configuration space.

These facts are fairly clear geometrically, but it is worthwhile to also examine them analytically. A small phase space vector $\delta z = (\delta x, \delta p)$ which is tangent to the Lagrangian manifold can be conveniently described by its displacement δu in the u coordinates, i.e., $\delta x = (\partial x/\partial u)\,\delta u$, $\delta p = (\partial p/\partial u)\,\delta u$. If the matrix $\partial x/\partial u$ is singular, then there exist nonzero displacements δu such that $\delta x = 0$. For these displacements, the phase space vector δz has vanishing components in its x components, and is therefore annihilated upon projection onto configuration space. It is convenient to designate the corank of the matrix $\partial x/\partial u$ as the *order* of the caustic; this is the number of linearly independent null eigenvectors δu, and therefore the number of directions in which the tangent plane is "vertical." Usually caustics will be first order, but they may go as high as the fth order. A caustic of order f is a *focus*.

Note that the definition of the singular set and the order of the caustic are independent of the coordinates u imposed on the Lagrangian manifold, since the rank of a matrix does not change under coordinate transformations.

In one degree of freedom, a caustic is simply a place where the derivative dp/dx of the curve is infinite (see Fig. 6). If the Lagrangian manifold is an orbit of a particle, then the caustic will also be a turning point, i.e., a place where $\dot{x} = 0$; but, as we shall see, in time-dependent problems, the Lagrangian manifold at given time need not be the trace of an orbit, and, in such problems, caustics do not occur at turning points.

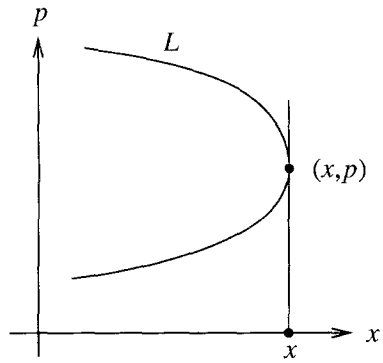

Fig. 6. In one degree of freedom, a caustic occurs where the slope dp/dx of the Lagrangian manifold is infinite. It is not necessarily a turning point.

3.2. Lagrangian Manifolds Project onto Curl-Free Momentum Fields

We will now show that if a region of a Lagrangian manifold which is free of caustics is projected onto configuration space, so that Eq. (3.3) is meaningful, then the resulting momentum field is curl-free. This is the best converse we can make of the result proved above, that the graphs of curl-free momentum fields are always Lagrangian manifolds.

To do this, we apply Stokes' theorem to the symplectic form, integrating it over a 2-dimensional surface in the $2f$-dimensional phase space, and relating the result to a line integral over the boundary (see

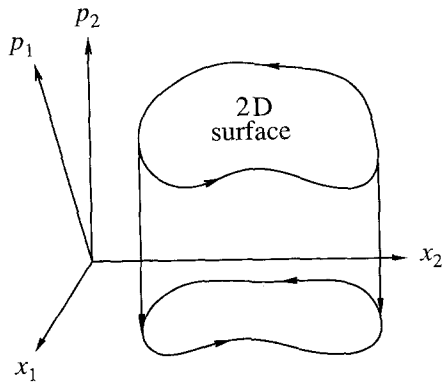

Fig. 7. The integral of the symplectic form over a 2-dimensional surface in phase space is equal to the contour integral of $p\,dx$ around the boundary. The line integral can be viewed either in phase space or via its projection onto configuration space. Note that the integral is always taken over a 2-dimensional surface, regardless of the dimension of the phase space.

Fig. 7). (The surface is 2-dimensional, not f-dimensional; it is generally not a Lagrangian manifold itself.) The result can be written

$$\int_{\text{Surface}} \omega = \oint_{\text{Boundary}} p\, dx \tag{3.10}$$

The meaning of this equation is the following. To interpret the left side, we imagine imposing a coordinate system (α, β) on the 2-dimensional surface, so that the f-dimensional vectors x, p are functions of (α, β) on the surface. Then the meaning of the surface integral in Eq. (3.10) is

$$\int_{\text{Surface}} \omega = \int d\alpha\, d\beta \sum_i \left(\frac{\partial p_i}{\partial \alpha} \frac{\partial x_i}{\partial \beta} - \frac{\partial p_i}{\partial \beta} \frac{\partial x_i}{\partial \alpha} \right) \tag{3.11}$$

One can see that the integrand on the right is simply ω acting on two vectors tangent to the surface,

$$\delta z_1 = \left(\frac{\partial x}{\partial \alpha}, \frac{\partial p}{\partial \alpha} \right), \qquad \delta z_2 = \left(\frac{\partial x}{\partial \beta}, \frac{\partial p}{\partial \beta} \right) \tag{3.12}$$

which is the reason for the notation $\int \omega$ in Eq. (3.10).

The line integral on the right side of Eq. (3.10) can be viewed geometrically either as a line integral in phase space, in which both x and p are functions of some parameter along the boundary, or via the projection of the curve onto configuration space, in which p is a function of position x along the projected curve there.

The proof of Eq. (3.10) proceeds in the usual way for Stokes' theorem, by breaking the region into small pieces, and demonstrating the result for each piece separately. On adding up the pieces, facing boundaries cancel. Equation (3.10) is valid for any closed curve in phase space serving as a boundary for a 2-dimensional surface, so long as x and p are smooth and well-defined in a neighborhood of the surface. In particular, the curve need not be the orbit of a physical system. (One must be careful, however, in applying the theorem to action–angle variables, which have discontinuities.)

To return to Eq. (3.3), we consider a Lagrangian manifold or a region of one which has no caustics, as illustrated in Fig. 8, so that p is a single-valued, smooth function of x in the projection. We construct a closed curve on the Lagrangian manifold in such a way that the 2-dimensional surface bounded by the curve can also be chosen to lie on the Lagrangian manifold, as shown in the figure. Then the surface integral of Eq. (3.10) vanishes, and therefore the line integral of $p\, dx$ around the closed loop in phase space must also vanish. This also applies to the integral of $p(x)\, dx$

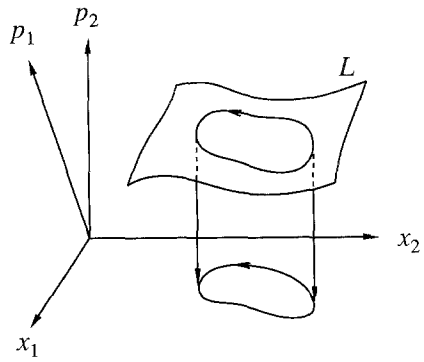

Fig. 8. On integrating $p\,dx$ around the boundary of a 2-dimensional region confined to a Lagrangian manifold, the result is zero. The momentum field $p=p(x)$ resulting from the projection of a Lagrangian manifold onto configuration space is a perfect gradient.

in the configuration space projection. Therefore integrals of $p(x)\,dx$ along open contours in configuration space must be invariant under continuous deformations of path, and must be functions only of the endpoints. This means that there exists a function $S(x)$ such that $p(x)=\partial S(x)/\partial x$, namely,

$$S(x)=\int^x p(x)\,dx \tag{3.13}$$

in which the lower limit is arbitrary. In this way we see that caustic-free regions of a Lagrangian manifold always project onto curl-free momentum fields, which was to be proved.

3.3. Generating Functions of Lagrangian Manifolds

We will call any function $S(x)$ which satisfies $p=p(x)=\partial S/\partial x$ on a Lagrangian manifold a *generating function* of that Lagrangian manifold. This is not universal terminology, but it is convenient and reasonable.

Let us consider to what extent a generating function and a Lagrangian manifold uniquely specify one another. To begin, if we are given a Lagrangian manifold, a generating function is not even defined unless we avoid caustic points. Supposing that a region of the Lagrangian manifold can be found which has no caustic points, then Eq. (3.13) shows that the generating function is unique up to an additive constant. This additive constant is usually associated with phase conventions in semiclassical applications. On the other hand, if the Lagrangian manifold has caustics, such as those illustrated in Figs. 9 and 10, then we can divide the Lagrangian manifold into regions which extend up to the caustics and which are

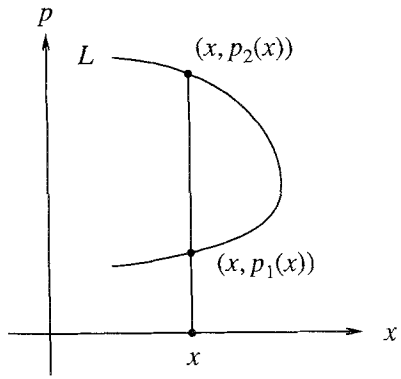

Fig. 9. The existence of caustic points on a Lagrangian manifold usually leads to a multivalued momentum field, $p = p_b(x)$, here illustrated in one degree of freedom for $b = 1, 2$.

separated by the caustics. Then each region corresponds to a distinct branch of the momentum field, $p = p_b(x)$, which is now multivalued, and each branch has its own generating function, $S_b(x)$, where b is a branch index. Each generating function $S_b(x)$ is determined to within its own additive constant, although it is usually convenient to link some or all of the additive constants together by demanding that the different functions $S_b(x)$ approach one another at the caustics dividing the branches. This is equivalent to defining an action function $S(x, p)$ on the Lagrangian manifold itself, as the line integral of $p\,dx$ along a contour confined to the Lagrangian manifold and taken relative to an arbitrary initial point, as shown in Fig. 11. This function then becomes multivalued upon projection, i.e., we have $S_b(x) = S(x, p_b(x))$.

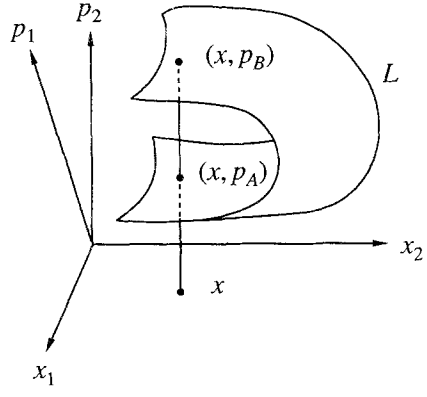

Fig. 10. Multiple branches $p_A(x)$, $p_B(x)$ of the momentum field in two degrees of freedom.

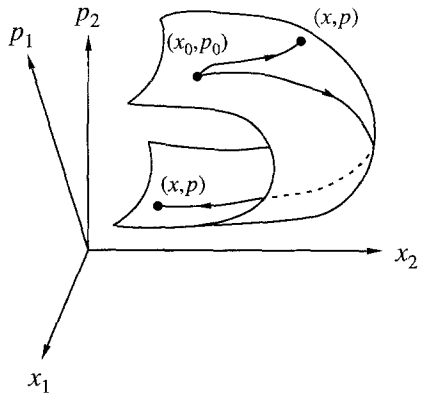

Fig. 11. An action function $S(x, p)$ can be defined at phase space points on a Lagrangian manifold as the integral of $p\,dx$ along a contour confined to the Lagrangian manifold, relative to an initial point (x_0, p_0). This action function becomes multivalued upon projection onto configuration space, but is continuous at caustics.

Even the function $S(x, p)$ on the Lagrangian manifold in phase space will be multiple-valued, if the manifold has a nontrivial topology. This has nothing to do with caustics, but rather the fact that sometimes a given final point of a Lagrangian manifold can be reached by more than one topologically distinct path leading from the given initial point. This occurs with the invariant tori of integrable systems, and also in the neighborhood of unstable periodic orbits of chaotic systems. Lagrangian manifolds which are topologically nontrivial do not normally occur in the Van Vleck formula.

Some Lagrangian manifolds consist entirely of caustic points, such as the surface $x = x_0 = \text{const}$, illustrated in Fig. 12. Such Lagrangian manifolds

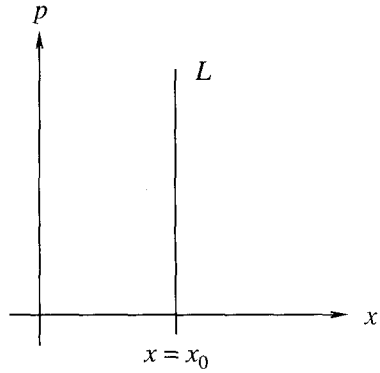

Fig. 12. An example of a Lagrangian manifold which has no x-space generating function $S(x)$, since it consists entirely of caustic points.

do not have a generating function $S(x)$, although we will show momentarily that they do have generating functions with respect to other coordinate systems. These Lagrangian manifolds are important in the semiclassical theory of the propagator, as we shall see.

Incidentally, Eq. (3.3) gives us another reason why all curves in the 2-dimensional phase plane are Lagrangian: it is because all functions $p(x)$ in one dimension are perfect gradients.

Although the definition of a Lagrangian manifold, Eq. (3.7), is formulated in such a way as to be invariant under canonical transformations, the generating function $S(x)$ is specific to the x-representation. Everything we have done with the generating function $S(x)$, however, can be carried over to any generalized coordinate Q, which is part of some canonical coordinate system (Q, P). This will give a new generating function $\tilde{S}(Q) = \int P(Q)\, dQ$, in which P is determined as a function of Q by projecting the Lagrangian manifold onto Q-space. (Even in the case of nonlinear coordinates, we project onto Q-space simply by throwing away the P coordinates.) In general, there will again be caustics, which will occur where the matrix $\partial Q/\partial u$ is singular; however, these singularities will usually not occur at the same points on the Lagrangian manifold at which $\partial x/\partial u$ is singular. That is, the singular set is determined relative to the representation being used.

For example, under the canonical transformation $Q = p$, $P = -x$, we have

$$\tilde{S}(Q) = \int^{Q} P(Q)\, dQ = \tilde{S}(p) = -\int^{p} x(p)\, dp \tag{3.14}$$

Notice that we are now dealing with a vector field $x(p)$ on momentum space, which will satisfy a version of Eq. (3.3) with the roles of x and p swapped. If the initial points for the x- and p-space integrals are the same point (x_0, p_0), as illustrated in Fig. 11, then the old and new actions are related by

$$\tilde{S}(p) = S(x) - xp + x_0 p_0 \tag{3.15}$$

However, $\tilde{S}(p)$ may be defined even when $S(x)$ is not; for example, the Lagrangian manifold of Fig. 12, although it has no configuration-space generating function, does have the perfectly nice momentum-space generating function, $\tilde{S}(p) = -x_0 p$.

The fact that the locations of caustic points are relative to the representation being used is important in WKB theory, for it allows one to avoid caustics by changing representation. For example, in one degree of freedom it is geometrically obvious that configuration-space caustics and

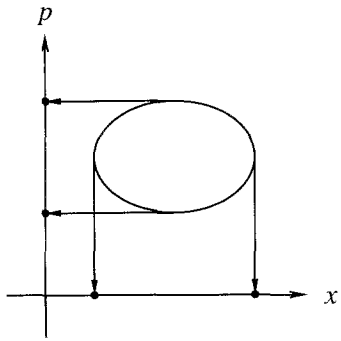

Fig. 13. In one degree of freedom, configuration-space caustics and momentum-space caustics never occur at the same place. Even in higher degrees of freedom, there always exists a representation in which a given point is caustic-free.

momentum-space caustics never occur at the same places on a Lagrangian manifold, as illustrated in Fig. 13. In higher degrees of freedom, this may no longer be true, as for example with the Lagrangian manifold specified by $x_1 = a = \text{const}$, $p_2 = b = \text{const}$ in the (x_1, x_2, p_1, p_2) phase space. Every point of this Lagrangian manifold is simultaneously on a configuration-space caustic and a momentum-space caustic. If, however, we perform a canonical transformation in which $(Q_1, Q_2) = (p_1, x_2)$, then in the Q-representation, the Lagrangian manifold is caustic-free. Maslov has shown[4] that by using a representation involving some mixture of commuting x's and p's, it is always possible to avoid caustics. The proof is notationally awkward, but not difficult. This leads to another fact, also geometrically obvious in one degree of freedom, that every Lagrangian

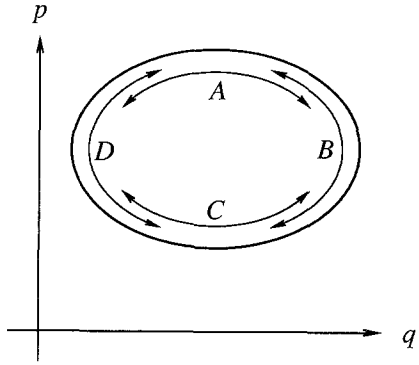

Fig. 14. A Lagrangian manifold can always be covered by overlapping regions such that each region is caustic-free in some representation.

manifold can be covered by overlapping regions, such that every region is caustic-free in some representation obtained from a commuting mixture of x's and p's. See Fig. 14, in which regions A and C are caustic-free in the x-representation, while regions B and D are caustic-free in the p-representation.

3.4. Generating Functions of Canonical Transformations

The generating functions of classical mechanics, which are used to generate canonical transformations, are only a slight generalization of the generating functions of Lagrangian manifolds we have introduced here. We have been discussing the generating function for a specific Lagrangian manifold, such a would arise in WKB theory when dealing with a specific wave function. On the other hand, a whole family of wave functions, such as a complete set of eigenstates of some complete set of commuting observables, will produce a whole family of Lagrangian manifolds, parametrized by some set of parameters $\lambda = (\lambda_1, ..., \lambda_f)$. There will in general be f of these parameters for a system of f degrees of freedom, for reasons explained in ref. 7. In this way, the family of wave functions of WKB form is associated with an action function $S(x, \lambda)$.

The geometrical picture corresponding to $S(x, \lambda)$ is shown in Fig. 15, in which phase space is divided up into an f-parameter family of f-dimensional Lagrangian manifolds. The parameter λ indicates which Lagrangian manifold we are on, and $p(x, \lambda) = \partial S/\partial x$ is the momentum field (possibly multivalued) associated with it. In quantum mechanics, the parameter λ may be restricted to discrete values, but in classical mechanics

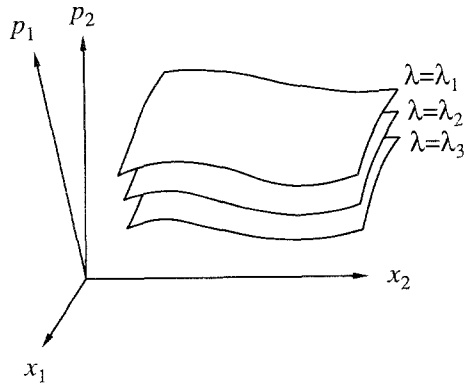

Fig. 15. A foliation of phase space into an f-parameter family of f-dimensional Lagrangian manifolds. Such a foliation corresponds to a canonical transformation, once a parameters λ of the Lagrangian manifolds are chosen and identified with some set of commuting new Q's and P's.

it is allowed to be continuous, so that an entire $2f$-dimensional region of phase space (perhaps all of it) is filled by the family of Lagrangian manifolds. Such a division of a space into a family of lower dimensional surfaces is called a *foliation*; an individual surface itself is called a *leaf*. The parameter λ may be allowed to be an arbitrary label of the Lagrangian manifolds, except that it should provide a unique specification for the members of the family.

The function $S(x, \lambda)$ is essentially one of the generating functions of canonical transformations in classical mechanics. To see this, note first that any given point of phase space (x, p) must lie on one leaf of the foliation, so the function $\lambda = \lambda(x, p)$ is uniquely specified. Therefore the λ's can serve as one half of a new coordinate system on phase space. Further, the Lagrangian condition implies that the Poisson brackets of the λ's among themselves vanish, $\{\lambda_i, \lambda_j\} = 0$, so λ can be identified with either the Q or P of a new canonical coordinate system. The proof of this fact proceeds as follows. Because the Lagrangian manifolds are contour surfaces of all the λ's, we have $X \cdot \nabla \lambda_k = 0$, $k = 1,..., f$, for any vector X tangent to the Lagrangian manifold. (The operator ∇ is a phase space gradient.) This can be written

$$0 = X \cdot \nabla \lambda_k = X \cdot \mathbf{J}^{-1} \mathbf{J} \cdot \nabla \lambda_k = \omega(X, \mathbf{J} \nabla \lambda_k) \tag{3.16}$$

This is true for any of f linearly independent vectors X tangent to the Lagrangian manifold. Therefore the vectors $\mathbf{J} \nabla \lambda_k$, $k = 1,..., f$, must also be tangent to the Lagrangian manifold, because the maximum number of linearly independent vectors which are pairwise annihilated by ω is f. Therefore the symplectic form acting on any pair of the $\mathbf{J} \nabla \lambda_k$'s must vanish. But this can be written in terms of the Poisson bracket,

$$\begin{aligned} \omega(\mathbf{J} \nabla \lambda_k, \mathbf{J} \nabla \lambda_l) &= \nabla \lambda_k \cdot \mathbf{J}^t \mathbf{J}^{-1} \mathbf{J} \cdot \nabla \lambda_l \\ &= -\nabla \lambda_k \cdot \mathbf{J} \cdot \nabla \lambda_l = -\{\lambda_k, \lambda_l\} = 0 \end{aligned} \tag{3.17}$$

It does not matter whether we identify the λ's with the Q's or P's of a new coordinate system, or even with some mixture of commuting Q's and P's. But to be specific, let us take $\lambda = Q$. Then the generating relation $p(x, Q) = \partial S(x, Q)/\partial x$ is clear on the basis of our whole development; the other generating relation $P(x, Q) = -\partial S(x, Q)/\partial Q$ serves as a definition of P, and is designed so as to make the formula for the symplectic form, Eq. (3.4), appear the same in both the old and new coordinates. That is, one demands the equality of the integrals,

$$\oint p \, dx = \oint P \, dQ \tag{3.18}$$

for all closed contours in phase space.

We have constructed here the Goldstein[17]-type $F_1(q, Q)$ generating function; the type $F_2(q, P)$ generating function is really the same, except the labels λ of the Lagrangian manifolds of the foliation are considered to be P's instead of Q's. The other two types, $F_3(p, Q)$, $F_4(p, P)$ are computed by projecting the Lagrangian manifolds of the foliation onto p-space, instead of onto q-space, and thus they bear the same relation to F_1 and F_2 that $S(q)$ does to $\tilde{S}(p)$ in Eq. (3.15). Books on classical mechanics sometimes discuss the fact that not every canonical transformation has a generating function, such as the identity transformation, $Q = q$, $P = p$, which does not have an $F_1(q, Q)$ generating function. The reason for this geometrically is the existence of caustics; for example, the identity transformation corresponds to a family of Lagrangian manifolds $q = Q$, where Q is interpreted as the parameter, which breaks phase space up into a family of vertical lines. See Figs. 12 and 16.

A really satisfactory geometrical interpretation of generating functions and their role in semiclassical mechanics would involve a more symmetrical treatment of the old and new variables than we have taken here. From a classical standpoint, this involves working in a kind of doubled phase space, where the differential of the action is $p\,dq - P\,dQ$, as discussed in Abraham and Marsden.[18] From a quantum or semiclassical standpoint, it is necessary to recognize that a wavefunction $\psi(q)$, which is parametrized by parameters $\lambda = Q$, might better be written $\psi(q, Q)$ or even $\langle q \mid Q \rangle$, and that it really satisfies two Schrödinger equations, one in the q variables, and one in the Q variables. This point of view has been developed in an elegant analysis by Miller,[11] and is discussed more fully in ref. 7.

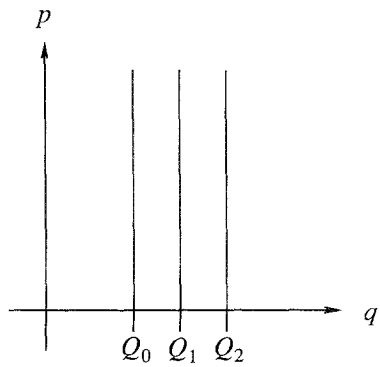

Fig. 16. The identity canonical transformation $Q = q$, $P = p$ is associated with a foliation of phase space into vertical Lagrangian manifolds, labeled by the values of Q.

3.5. Solving the Hamilton–Jacobi Equation

Let us now see how the theory of Lagrangian manifolds can be applied to the Hamilton–Jacobi equation. Consider Fig. 17, in which an initial Lagrangian manifold L' is obtained from the initial action $S(x', t') = S_0(x')$ by

$$p_0(x') = p(x', t') = \frac{\partial S(x', t')}{\partial x'} \tag{3.19}$$

We assume the initial Lagrangian manifold is free of caustics, so that $p_0(x')$ is a single-valued momentum field. Now we let each point of the initial Lagrangian manifold, such as (x', p') in the figure, flow according to Hamilton's equations, thereby mapping the initial Lagrangian manifold L' into a final, f-dimensional manifold L''. It turns out, as we will show momentarily, that L'' is also Lagrangian. Therefore, if L'' is free of caustics, it also has a generating function, which we may write as $S(x'', t'')$. For many problems the final Lagrangian manifold will be free of caustics for short elapsed times $t'' - t'$, because only small changes in topology will occur in small times. If, however, L' has tails which extend to infinity, then it may be that caustics will develop far out on the tails in arbitrarily short times. For simplicity, we will assume for now that L'' has no caustics.

It is then a reasonable guess that $S(x'', t'')$, with suitably chosen additive constant, is the solution of the Hamilton–Jacobi equation, Eq. (2.6). This is true, but there are several steps involved in the proof.

The first is to show that Lagrangian manifolds are always mapped into other Lagrangian manifolds under the time flows generated by

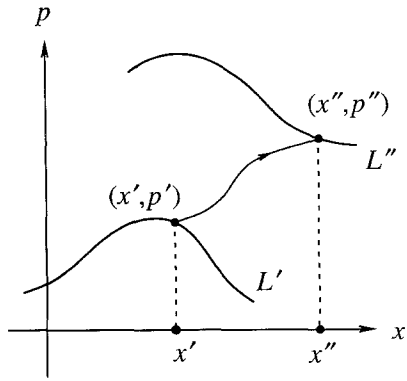

Fig. 17. An initial Lagrangian manifold L' at time t' is mapped into a final Lagrangian manifold L'' at time t'' by following orbits. The solution of the time-dependent Hamilton–Jacobi equation is a generating function of the final Lagrangian manifold.

Hamilton's equations. This follows immediately from a basic fact about the symplectic form, illustrated in Fig. 18. We consider an initial condition (x', p') in phase space, and an orbit taking it to (x'', p'') in elapsed time $t'' - t'$. Two small displacements $\delta z_1'$, $\delta z_2'$, going from (x', p') to initial conditions for two nearby orbits, are mapped under the flow to final displacements $\delta z_1''$, $\delta z_2''$. We use the linearized equations of motion to describe the evolution of the displacements. It then turns out that

$$\omega(\delta z_1', \delta z_2') = \omega(\delta z_1'', \delta z_2'') \tag{3.20}$$

giving us a sense in which the symplectic form is conserved in time. This is equivalent to the fact that the solution to Hamilton's equations, expressing the final q's and p's as functions of the initial q's and p's at fixed time, constitute a canonical transformation. If now the initial conditions (x', p') lie on an initial Lagrangian manifold, and if $\delta z_1'$, $\delta z_2'$ are tangent to it, then the left side of Eq. (3.20) vanishes. But then $\delta z_1''$, $\delta z_2''$ are tangent to the final manifold, which must therefore also be Lagrangian.

Therefore L'' has a generating function, determined to within an additive constant. Not just any constant will do, however, because $S(x'', t'')$ must satisfy both the Hamilton–Jacobi equation and the initial conditions. It turns out that the solution we want is given by

$$S(x'', t'') = S(x', t') + R(x'', t''; x', t') \tag{3.21}$$

where R is Hamilton's principal function, i.e., the line integral of $p\,dx - H\,dt$ along an orbit connecting (x', t') with (x'', t'').

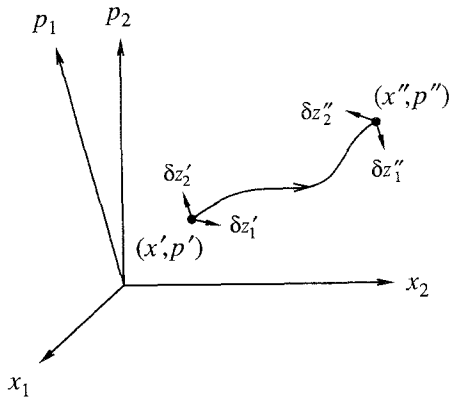

Fig. 18. When a pair of small displacements is mapped along an orbit by the linearized equations of motion, the symplectic form acting on them is constant in time.

To prove this, we must first be careful about the functional dependences of the symbols we have used. The times t', t'' can be specified independently, as can the final point x'' at which we wish to evaluate the action. Given these, however, the initial point x' is determined. The rule for the function $x' = x'(t', t'', x'')$ is the following. Given x'', we find the momentum p'' which lies on the final Lagrangian manifold above x''. We then follow the orbit ending at (x'', p'') at time t'' backward to (x', p') at time t'. The x' value of this initial condition is then the function $x'(t', t'', x'')$. Finding the final momentum p'', given x'', is easy to say and is clear geometrically, but in practice one usually must search initial conditions (x', p') for an orbit which ends at coordinate x'' at time t''.

Direct substitution now verifies that $S(x'', t'')$ from Eq. (3.21) does satisfy the Hamilton–Jacobi equation. First we differentiate with respect to x'', obtaining

$$\frac{\partial S(x'', t'')}{\partial x''} = \frac{\partial S(x', t')}{\partial x'} \frac{\partial x'}{\partial x''} + \frac{\partial R}{\partial x'} \frac{\partial x'}{\partial x''} + \frac{\partial R}{\partial x''} \tag{3.22}$$

But, because $\partial S(x', t')/\partial x' = p'$ and $\partial R/\partial x' = -p'$, the first two terms cancel, while the last term gives p''. Therefore

$$\frac{\partial S(x'', t'')}{\partial x''} = p'' \tag{3.23}$$

This shows that $S(x'', t'')$ defined by Eq. (3.21), which must in any case generate some Lagrangian manifold, in fact generates the final Lagrangian manifold which is the image of the initial Lagrangian manifold under the time evolution.

Similarly, differentiating Eq. (3.21) with respect to t'' yields

$$\frac{\partial S(x'', t'')}{\partial t''} = -H(x'', p'', t'') \tag{3.24}$$

with two terms again canceling. This shows that $S(x'', t'')$ actually does satisfy the time-dependent Hamilton–Jacobi equation.

Finally, we note that as $t'' \to t'$, the R term in Eq. (3.21) goes to zero and $S(x'', t'') \to S(x', t')$, because the integral of $L\,dt = p\,dq - H\,dt$, where L is the Lagrangian, along a zero-length orbit is zero. Therefore $S(x'', t'')$ satisfies the required initial conditions, and it is the solution we seek.

An interesting interpretation of this solution is obtained by endowing the particles of our swarm, introduced earlier, with an action S in addition to their position and momentum (x, p). The initial action is $S(x', t')$,

and as time evolves, each particle accumulates action by integrating $L\,dt = p\,dq - H\,dt$. The action $S(x'', t'')$ at a final point x'' is then simply the action of whatever particle ends up there.

4. AMPLITUDE TRANSPORT, CAUSTICS, AND THE MASLOV INDEX

To complete the solution of the initial value problem in WKB theory, we must now solve the amplitude transport equation, Eq. (2.9). This is just the continuity equation, which we will interpret as representing conservation of particles of density $\rho = |A|^2$. Thus, we can immediately write $\rho(x'', t'')\,dx'' = \rho(x', t')\,dx'$, as illustrated in Fig. 19, or, by taking square roots,

$$A(x'', t'') = A(x', t') \left| \det \frac{\partial x'}{\partial x''} \right|^{1/2} \tag{4.1}$$

The expression $\partial x'/\partial x''$ is the derivative of the function $x'(t', t'', x'')$ discussed below Eq. (3.21); it is neither a derivative at fixed p' nor at fixed p'', since as x'' changes, both p' and p'' move along their respective Lagrangian manifolds. Collecting things, we can now write the solution to the initial value problem in WKB theory in the form

$$\psi(x'', t'') = A(x', t') \left| \det \frac{\partial x'}{\partial x''} \right|^{1/2} \exp\left\{ \frac{i}{\hbar} [S(x', t') + R(x'', t''; x', t')] \right\} \tag{4.2}$$

For short elapsed times, the absolute value signs in Eqs. (4.1) and (4.2) are not necessary, since at $t = t'$, we have $\det \partial x'/\partial x'' = +1$. For such

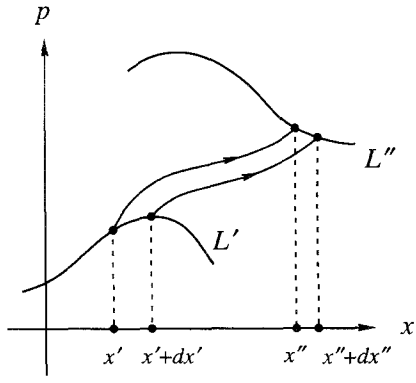

Fig. 19. The number of particles between x' and $x' + dx'$ is the same as between x'' and $x'' + dx''$. This leads to an immediate solution of the amplitude transport equation.

times the amplitude $A(x'', t'')$ remains real and positive. At longer times, the determinant may change sign—not, as it turns out, by passing through zero, but by diverging. To examine this question, it is better to look at the inverse matrix, $\partial x''/\partial x'$, which is always finite. To see this, note first that x' uniquely labels points on the initial Lagrangian manifold, since we assume it has no caustics. Therefore we can treat x' as coordinates on L', and identify x' with the u coordinates introduced in Eq. (3.9). By carrying this coordinate system along with the Lagrangian manifold as it evolves, x' can also be used as coordinates at any later time. Then the finiteness of $\partial x''/\partial x'$ is merely a reflection of the fact that L'' is smooth and well behaved (as we assume).

On the other hand, the quantity $\det \partial x''/\partial x'$ may vanish, signaling a caustic of L'', precisely as in Eq. (3.9). This is illustrated in Fig. 20, where orbit a has reached a caustic of L'' at the final time t''. In the same diagram, orbit b has not yet reached a caustic at the final time, whereas orbit c has already passed through one; the respective derivatives $\partial x''/\partial x'$ (in one degree of freedom) for orbits b, a, c are positive, zero, and negative.

A common misconception about caustics is that they are somehow a property of a single orbit, taken out of context. In fact, a caustic is only determined by a family of orbits, which are always associated with a Lagrangian manifold. For example, the same orbit a in Fig. 20 would not be at a caustic at time t'' if the initial Lagrangian manifold were modified [leaving (x', p') alone].

Equation (4.2) shows that the WKB expression for the final wave function $\psi(x'', t'')$ diverges when x'' is at a caustic. Since nonlinear partial differential equations typically develop singularities in finite time, this

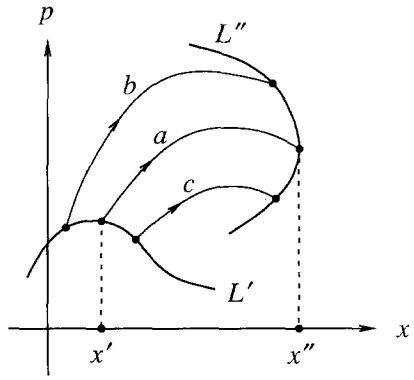

Fig. 20. Orbit a, starting at x', has reached a caustic at x''. Orbit b has not yet reached a caustic at the final time, while orbit c has passed one.

behavior is not surprising. The divergence represents a nonuniformity in the variables (x'', t'') of the expansion of ψ in $\hbar$, i.e., although the error in ψ goes to zero like $\hbar$ as $\hbar \to 0$ for fixed (x'', t''), it goes to ∞ for fixed $\hbar$ as (x'', t'') approaches a caustic.

What is more surprising is that an almost obvious continuation of the solution through the divergence is valid (in the sense of an asymptotic expansion in $\hbar$), even though the solution near the caustic is not. The obvious part is to take the prescription surrounding Eq. (3.21) literally, even for orbits which have passed through caustics; the only change is that the function $x'(t', t'', x'')$ is now multiple-valued, corresponding to possibly several orbits which reach x'' in the allowed time from the initial Lagrangian manifold (but with different values of x', p', and p''). This is illustrated in Fig. 21.

As for the amplitude, let us assume that $\det \partial x''/\partial x'$, which was initially $+1$, passes through zero and goes negative at the caustics. Then it is logical to interpret the square root in Eq. (4.1) as giving an imaginary result after the first caustic has been passed; by convention, we will force $A(x'', t'')$ to be positive, absorbing the imaginary unit into a phase factor. The only part that is not obvious is whether the phase factor should be $-i$ or $+i$; we will write it as $\exp(-i\kappa\pi/2)$, with κ an integer yet to be determined. As more caustics are passed along an orbit, the phase factor will accumulate, always being representable in the form $\exp(-i\kappa\pi/2)$ for some integer κ. The integer κ is variously called the *Morse index* or *Maslov index*; the distinction will be discussed below.

Finally, we interpret the multiple orbits arriving at the same final x''

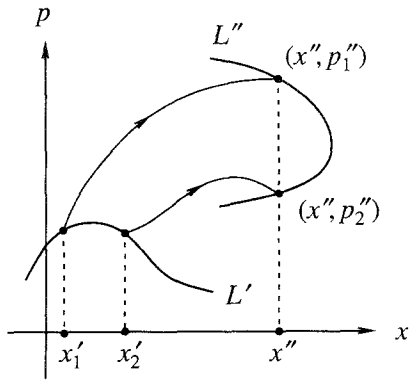

Fig. 21. After caustics develop, the quantities x', p', and p'', all considered functions of (t', t'', x''), become multiple-valued, corresponding to distinct orbits arriving at (x'', t''). Two branches are shown, $b = 1, 2$.

point as contributing independent wavelets to the solution, all of which are to be added together. The result is

$$\psi(x'', t'') = \sum_b A_b(x'', t'') \exp\left[\frac{i}{\hbar} S_b(x'', t'') - i\kappa_b \frac{\pi}{2}\right] \tag{4.3}$$

where b is the branch of the function $x'(t', t'', x'')$, and where A_b, S_b, and κ_b are determined separately for each branch [now the absolute value signs in Eq. (4.1) are necessary].

4.1. Momentum-Space Wave Functions

The plausibility arguments leading to Eq. (4.3) can be strengthened, and a rule obtained for computing the indices κ, by considering momentum-space wave functions. We begin with a configuration-space wave function consisting of a single term of phase integral form,

$$\psi(x) = A(x) \exp\left[\frac{i}{\hbar} S(x)\right] \tag{4.4}$$

corresponding to a Lagrangian manifold which is free of caustics in configuration space. We suppose further that the Lagrangian manifold is free of momentum space caustics, i.e., that either x or p can be used as coordinates on the Lagrangian manifold, and that x and p are invertible functions of each other. This is illustrated in Fig. 22. We will write

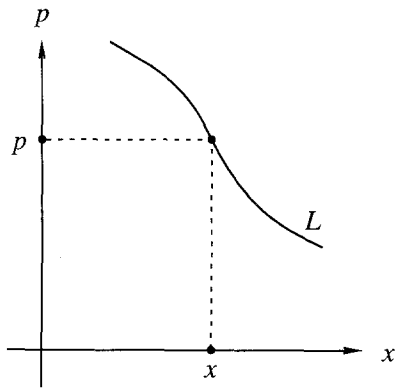

Fig. 22. A Lagrangian manifold which has neither x-space caustics nor p-space caustics. The stationary phase evaluation of the Fourier transform converts a single branch of a WKB wave function in the x-representation into another single branch in the p-representation.

$p = p_L(x)$ and $x = x_L(p)$ for these functions, using the subscript L to distinguish functions from values.

Then the momentum-space wave function is given by the Fourier transform,

$$\phi(p) = \frac{1}{(2\pi\hbar)^{f/2}} \int d^f x\, A(x) \exp\left\{\frac{i}{\hbar}[S(x) - xp]\right\} \tag{4.5}$$

We evaluate this integral by the stationary phase approximation, because by so doing we obtain an approximation for ϕ which is of the same order in $\hbar$ as that we started with for ψ. Note that p is a parameter of the integral, so the stationary phase points are x values depending on p; they are the roots of $p = \partial S(x)/\partial x = p_L(x)$. But by our assumptions, there is only one such root, $x = x_L(p)$. The stationary phase evaluation of the integral also involves the symmetric matrix $\mathbf{M}(x) = \partial^2 S(x)/\partial x\, \partial x = \partial p_L(x)/\partial x$; this matrix is finite and nonsingular by our assumption of the absence of caustics, so its eigenvalues λ are real and nonzero. The integral then gives

$$\phi(p) = e^{i\alpha\pi/4} A(x)\, |\det \mathbf{M}(x)|^{-1/2} \exp\left\{\frac{i}{\hbar}[S(x) - xp]\right\} \tag{4.6}$$

where α is an integer given by the *index of inertia* of $\mathbf{M}$, defined as the number of positive eigenvalues minus the number of negative eigenvalues. On the right side of Eq. (4.6), x is understood to mean $x_L(p)$; A, $\mathbf{M}$, and S depend on x, but α does not, since if $\det \mathbf{M}$ neither vanishes nor diverges, its index of inertia does not change. That is, regarded as a function of position on the Lagrangian manifold, α is constant.

Notice that like $\psi(x)$, $\phi(p)$ also has phase integral form. The momentum-space action is given by

$$\tilde{S}(p) = S(x) - xp \tag{4.7}$$

so that it is the momentum-space generating function of the Lagrangian manifold, precisely as in Eq. (3.15). The momentum-space amplitude $\tilde{A}(p)$ is conveniently expressed in terms of a momentum-space density $\tilde{\rho}(p) = A(p)^2$ given by

$$\tilde{\rho}(p) = \rho(x) \left|\det \frac{\partial x}{\partial p}\right| \tag{4.8}$$

exactly as we would expect for the transformation of a density under a change of variables. (Here we have written $\mathbf{M}^{-1} = \partial x/\partial p$.) Only the overall phase $\exp(i\alpha\pi/4)$ of Eq. (4.6) might not have been predicted on classical

grounds, and it is dependent on the usual phase conventions used for the Fourier transform in quantum mechanics.

A significant fact about these calculations is that the stationary phase approximation is a semiclassical approximation of the same order in $\hbar$ as that inherent in the WKB form of the wave functions, and that it is also interpreted geometrically in terms of Lagrangian manifolds in phase space and their projections. More generally, any quantum mechanical operation involving integrals or sums can be evaluated by the stationary phase approximation, and the error is always of order $\hbar$. However, the error is generally nonuniform in other parameters, such as position, time, quantum numbers, etc.

Let us now modify this calculation by allowing the Lagrangian manifold to have a momentum-space caustic (but still without any configuration-space caustics), as shown in Fig. 23. Now the stationary phase evaluation of the integral in Eq. (4.5) produces two stationary phase points x_1, x_2 for a given value of p, as shown in the figure, corresponding to the now double-valued function $x_L(p)$. The integral is now expressed as a sum of two terms, each of the WKB form shown in Eq. (4.6).

The index of inertia of $\mathbf{M}$, regarded as a function on the Lagrangian manifold, is no longer constant, but rather changes discontinuously at the momentum-space caustic, where $\mathbf{M} = \partial p/\partial x$ has vanishing determinant. The number of eigenvalues of $\mathbf{M}$ which vanish at the caustic is the order of the caustic, and each eigenvalue which passes through zero at the caustic will change the index of inertia of $\mathbf{M}$ by either $+2$ or -2, depending on the direction of the change. Therefore the relative phase shift between the two

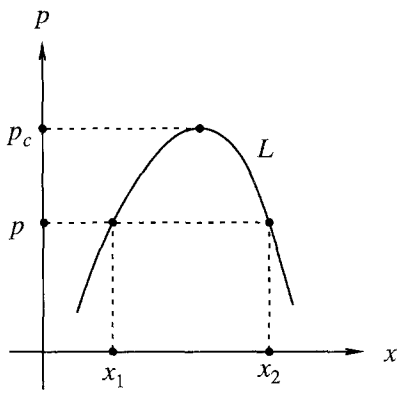

Fig. 23. A Lagrangian manifold with no x-space caustics and one p-space caustic, at momentum value p_c. A single-branch wave function in WKB form in the x-representation corresponds to two branches in the p-representation, with a relative phase shift which is an integral multiple of $\pi/2$.

branches due to the indices of inertia will have the form $\exp(-i\kappa\pi/2)$, for some integer κ. Altogether, the wave function can be written in the form

$$\phi(p) = e^{i\alpha_0\pi/4} \sum_b \tilde{A}_b(p) \exp\left[\frac{i}{\hbar}\tilde{S}_b(p) - i\kappa_b \frac{\pi}{2}\right] \tag{4.9}$$

where α_0 is the index of inertia of one of the branches, whose κ is zero; where κ_b for the other branch represents the relative phase shift due to the change in the index of inertia across the caustic; and where $\tilde{S}_b$ and $\tilde{A}_b$ are defined for each branch by Eqs. (4.7) and (4.8) [with $\tilde{A}_b = (\tilde{\rho}_b)^{1/2}$.] For example, in the case illustrated in Fig. 23 for one degree of freedom, we find $\alpha_0 = 1$, $\kappa_1 = 0$, and $\kappa_2 = 1$.

4.2. The Maslov Index

Our calculation of momentum-space wave functions has shown that a WKB wave function consisting of a single branch in one representation may result in multiple branches in another representation, exactly as we have hypothesized in Eq. (4.3) for the solution of the time-dependent problem in phase integral theory, and that the branches differ from one another by phase shifts which are integral multiples of $\pi/2$. We can now turn this process around, and argue that if we believe the multiple branch form for the final wave function $\psi(x'', t'')$ shown in Eq. (4.3), then the relative phase shifts between the branches can be determined by switching to another representation in a region straddling the x-space caustic, such that the region is caustic-free in the new representation. We then simply demand that the transformed wave function be continuous over the region in question. A rigorous justification of this prescription may be found in the books by Maslov and by Maslov and Fedoriuk[4] and will not be given here, but the idea is certainly compelling and plausible. Here we will simply proceed to work out the details of the prescription.

We will use the initial position x' as coordinate on the final Lagrangian manifold L'', as discussed earlier, so the matrices $\partial x''/\partial x'$ and $\partial p''/\partial p'$ are smooth and finite on the L''. The configuration-space caustics, which separate the branches of Eq. (4.3), are the places on the Lagrangian manifold where $\det(\partial x''/\partial x')$ vanishes. We will assume that these caustics are surfaces of dimensionality $f-1$ on the f-dimensional Lagrangian manifold, thereby separating two f-dimensional regions of the manifold. (This assumption is typically valid, but not always; we might have a case such as illustrated in Fig. 12, or we might find that the singular set has kinks or portions of dimensionality less than $f-1$. See Arnold[10] for more details on this point.) We will also assume for simplicity that momentum-

space caustics do not coincide with configuration-space caustics, so that $\det(\partial p''/\partial x')$ is nonzero in a region straddling the configuration-space caustic. Later we will relax this assumption, but for now it means that $\partial x''/\partial p'' = (\partial x''/\partial x')(\partial p''/\partial x')^{-1}$ is finite and well behaved, and that caustics are equally well signalled by the vanishing of $\det(\partial x''/\partial p'')$ as by that of $\det(\partial x''/\partial x')$.

We concentrate on two branches, say $b = 1, 2$, of Eq. (4.3). On performing the stationary phase evaluation of the Fourier transform to obtain $\phi_b(p'', t'')$ for each of the branches, we find, first, that the amplitude $\tilde{A}_b(p'', t'')$ is continuous across the caustic, even though $A_b(x'', t'')$ is not. This follows from combining the result of the stationary phase evaluation of the integral with Eq. (4.1),

$$\tilde{A}_b(p'', t'') = A_b(x'', t'') \left| \frac{\partial x''}{\partial p''} \right|^{1/2} = A(x', t') \left| \frac{\partial x'}{\partial p''} \right|^{1/2} \tag{4.10}$$

Second, the phases of the two branches of $\phi_b(p'', t'')$ can be written $\tilde{S}_b(p'') - i\kappa_b \pi/2 + i\alpha_b \pi/4$, where $\tilde{S}_b(p'') = S_b(x'') - x''p''$, where κ_b represents the indices in Eq. (4.3), presumed as yet unknown, and where κ_b is the index of inertia of $\partial x''/\partial p'' = (\partial^2 S(x'')/\partial x''\, \partial x'')^{-1}$. Since $\tilde{S}_b(p'')$ is automatically continuous across the caustic, demanding equality of momentum-space phases requires that

$$\kappa_2 = \kappa_1 + \frac{\alpha_2 - \alpha_1}{2} \tag{4.11}$$

Thus, we can say that the change in κ across a caustic is the number of eigenvalues of $\partial x''/\partial p''$ which change from negative to positive at the caustic, minus the number which change the other way. The bounds on $\Delta\kappa$ are therefore

$$-n \leqslant \Delta\kappa \leqslant +n \tag{4.12}$$

where n is the order of the caustic.

This $\Delta\kappa$ links two branches of $\psi(x'', t'')$ together. By extending the process, it is possible to associate a total $\Delta\kappa$ measured between the endpoints of a curve segment on a Lagrangian manifold, which may cross a number of caustics and straddle several branches. We assume that the result of this computation does not change if the path is continuously deformed, while being confined to the Lagrangian manifold. This is a consistency requirement, discussed by Maslov and by Maslov and Fedoriuk[4] and by Arnold.[19]

If it should happen that momentum-space caustics coincide with con-

figuration-space caustics, then we must use a mixed representation Q, composed of some commuting mixture of the original x's and p's. Then the transformation to the Q-representation involves a Fourier transform in a smaller number of variables than f. Otherwise, all the arguments above are repeated, with only minor changes.

The quantity $\Delta\kappa$, associated with a directed curve segment on a Lagrangian manifold, is the *Maslov index* of that segment. Notice that the curve does not have to be the orbit of a physical system; indeed, for the time-dependent problem we have considered, it usually is not. Notice also that the Maslov index really only depends on the geometry of the Lagrangian manifold in phase space, and on its projection onto configuration space; the origin of the Lagrangian manifold in some dynamical problem is irrelevant. The Maslov index is, however, dependent on the representation (x in this case) being used. Although the Maslov index does not change under continuous deformations of the curve, it may change under discontinuous ones; for example, on an invariant torus of an integrable system, there may be more than one topologically distinct path joining two given branches, and the Maslov indices of these two paths are not necessarily the same. Thus, the Maslov index is not simply a relative property of two branches, but depends also on the path which links them.

4.3. The Morse Index

The computation of the Maslov index we have just described allows us to determine the relative phase space shifts between the branches of the wave function $\psi(x'', t'')$ at a fixed time, but it does not directly allow us to determine the absolute phase of any branch. A slight modification of the method, however, will remedy this shortcoming. That is, we simply repeat the process of patching our way through caustics by switching to the momentum (or other appropriate) representation, not along a curve on the Lagrangian manifold, but rather along an orbit. For example, when an orbit is at an x-space caustic, as in Fig. 24, we transform to (say) the momentum representation in some time interval around the caustic, and over some interval on the Lagrangian manifold straddling the caustic. Continuity of the transformed wave function then leads to the same rule as above, formulated in terms of the eigenvalues λ of the matrix,

$$\Lambda(t) = \left[\frac{\partial^2 S(x)}{\partial x\, \partial x}\right]^{-1} = \frac{\partial x}{\partial p} \tag{4.13}$$

and the direction in which they pass through zero. The only difference is that now Λ is considered a function of time along an orbit. Since κ is

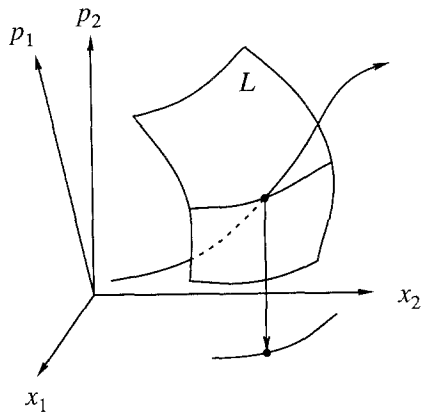

Fig. 24. An orbit is passing through a caustic. The behavior of the eigenvalues of $\Lambda = \partial x/\partial p$, considered as a function of time along the orbit, determine the jump $\Delta\kappa$ in the index κ.

known to be zero at $t'' = t'$, the changes $\Delta\kappa$ along an orbit determined by this rule allow us to determine the final κ_b's for all the branches.

This method for computing κ_b can be simplified in the case of Hamiltonians which have the form of kinetic plus potential energies, with or without magnetic fields. To do this, we require an equation of evolution for Λ along an orbit. Let $\delta z = (\delta x, \delta p)$ be a small displacement vector tangent to the Lagrangian manifold, whose base is on the orbit of interest and whose tip is carried along by the flow, as in Fig. 25. The equations of

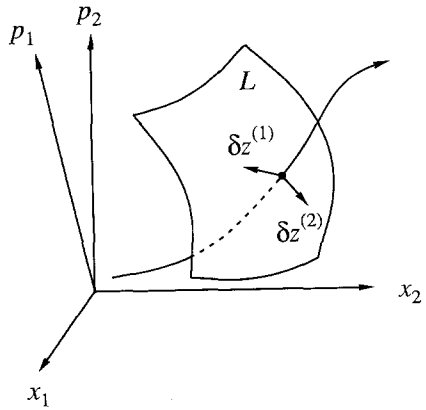

Fig. 25. Small vectors δz tangent to a Lagrangian manifold are transported along with an orbit. The properties of these vectors determine the locations of the caustics and the jumps in the indices κ.

evolution of δz are just the linearized equations of motion obtained by replacing $x(t)$, $p(t)$ in Hamilton's equations by $x(t)+\delta x(t)$, $p(t)+\delta p(t)$ and expanding to first order in small quantities. The result is

$$\begin{aligned} \delta\dot{x} &= \mathbf{H}_{px}\,\delta x + \mathbf{H}_{pp}\,\delta p \\ \delta\dot{p} &= -\mathbf{H}_{xx}\,\delta x - \mathbf{H}_{xp}\,\delta p \end{aligned} \tag{4.14}$$

where $\mathbf{H}_{px}$ is the $f\times f$ matrix $\partial^2 H(x, p)/\partial p\,\partial x$, etc., evaluated as a function of time along the orbit. Notice that since $\delta z=(\delta x, \delta p)$ is tangent to the Lagrangian manifold, its x and p components are related by $\delta x(t)=\mathbf{\Lambda}(t)\,\delta p(t)$.

Let us repeat this process for a collection of f linearly independent vectors $\delta z^{(1)},\ldots,\delta z^{(f)}$, tangent to the Lagrangian manifold, as in Fig. 25, so that each of these vectors obeys Eqs. (4.14). Let us also stack the x and p components of these f vectors columnwise to form two $f\times f$ matrices $\mathbf{A}$, $\mathbf{C}$ given by $A_{ik}=\delta x_i^{(k)}$, $C_{ik}=\delta p_i^{(k)}$, so that $\mathbf{A}(t)=\mathbf{\Lambda}(t)\,\mathbf{C}(t)$. Then the equations of evolution of $\mathbf{A}$ and $\mathbf{C}$ are just Eqs. (4.14), with δx and δp replaced by $\mathbf{A}$ and $\mathbf{C}$, respectively.

Assuming for simplicity that x-space caustics do not coincide with p-space caustics, the matrix $\mathbf{C}$ will be nonsingular in the region around the x-space caustic, through which we wish to patch. Then in this region we can write $\mathbf{\Lambda}=\mathbf{A}\mathbf{C}^{-1}$, or

$$\dot{\mathbf{\Lambda}}=\dot{\mathbf{A}}\mathbf{C}^{-1}-\mathbf{A}\mathbf{C}^{-1}\dot{\mathbf{C}}\mathbf{C}^{-1}=\mathbf{H}_{pp}+\mathbf{H}_{px}\mathbf{\Lambda}+\mathbf{H}_{xp}+\mathbf{\Lambda}\mathbf{H}_{xx}\mathbf{\Lambda} \tag{4.15}$$

This is the desired evolution equation for $\mathbf{\Lambda}$.

We can now convert this into an equation for the evolution of the eigenvalues λ, whose vanishing signals a caustic. Let $\mathbf{e}$ be a normalized, real eigenvector of $\mathbf{\Lambda}$, such that $\mathbf{\Lambda}\mathbf{e}=\lambda\mathbf{e}$. Then we have $\dot{\lambda}=\mathbf{e}\cdot\dot{\mathbf{\Lambda}}\cdot\mathbf{e}$ (the terms involving $\mathbf{e}\cdot\dot{\mathbf{e}}$ cancel since $\mathbf{e}$ is normalized). Therefore

$$\dot{\lambda}=\mathbf{e}\cdot\mathbf{H}_{pp}\cdot\mathbf{e}+2\lambda(\mathbf{e}\cdot\mathbf{H}_{px}\cdot\mathbf{e})+\lambda^2(\mathbf{e}\cdot\mathbf{H}_{xx}\cdot\mathbf{e}) \tag{4.16}$$

From this we see that when $\lambda=0$, $\dot{\lambda}>0$, since the matrix $\mathbf{H}_{pp}$, which is the inverse mass tensor, is positive definite. Therefore the index κ, considered as a function of time along an orbit, always increases at caustics, since all the vanishing eigenvalues of $\mathbf{\Lambda}$ pass through zero from negative to positive values. The amount of the jump is equal to the order of the caustic. (We need not worry about running out of negative eigenvalues of $\mathbf{\Lambda}$ as we proceed along an orbit; the analysis above is only applicable to the region near the x-space caustics, where we assume there are no p-space caustics. In between x-space caustics, however, there may well be p-space caustics, at which some eigenvalues of $\mathbf{\Lambda}$ will change sign by diverging.)

Altogether, for kinetic-plus-potential systems we have a simple rule, which says that the index κ is simply the count of caustics (multiplicities included) encountered by an orbit passing from the initial to the final Lagrangian manifolds. This rule, and its relation to the positive-definiteness of the mass tensor, has been discovered and rediscovered a number of times; it is discussed by Maslov,[4] Pechukas,[21] and Levit *et al.*[20] The index κ obtained by counting caustics is properly called the *Morse index.* In the present context, the Morse index can be viewed as a special case of the Maslov index if we enlarge our phase space to include time and energy as conjugate variables. It then turns out that the Morse index is the Maslov index of a curve segment (namely the orbit itself) which lies on a Lagrangian manifold in the extended phase space.

By whatever name, the Maslov prescription for computing the index by demanding continuity in transformed wave functions is more general than counting caustics because it applies to Hamiltonians in which the matrix $\mathbf{H}_{pp}$ is not positive definite. It also applies to any Hamiltonian in representations other than the x-representation; for example, if we were interested in solving the initial value problem of phase integral theory in the momentum representation, then all the arguments above would be repeated, with the roles of x and p swapped. Then the change in sign of the eigenvalues would be determined by the matrix $\mathbf{H}_{xx}$, which in general is not definite. The Maslov index also applies to curve segments which are not orbits, as one often requires in quantization problems. For such segments, the index does not necessarily increase at caustics, even for kinetic-plus-potential problems in the x-representation. Further discussion and algorithms for computing the Maslov index can be found in refs. 22 and 23.

5. THE VAN VLECK FORMULA

Let us now apply time-dependent WKB theory to the initial wave function $\psi_0(x) = \delta(x - x')$ at time $t = t'$, as discussed in Section 2, in order to find a WKB expression for the propagator at a later time. We immediately encounter a minor problem, namely that the initial wave function does not have the WKB form shown in Eq. (2.4), so we cannot identify an initial amplitude and phase. The problem is easily circumvented, however; we simply work in momentum space, in which the initial wave function is

$$\phi_0(p) = \frac{1}{(2\pi\hbar)^{f/2}} \exp\left(-\frac{ipx'}{\hbar}\right) \tag{5.1}$$

Therefore the initial amplitude and phase are $\tilde{A}_0(p) = (2\pi\hbar)^{-f/2}$ and $\tilde{S}_0(p) = -px'$, and the WKB representation is exact. Working in momentum space for small times is a natural approach in the spirit of Maslov's theory. It is an approach which is also taken by Ozorio de Almeida.[24]

For a fixed value of x', the initial Lagrangian manifold appears as in Fig. 12, and we see that the failure of the WKB form in the x-representation is due to the fact that the initial Lagrangian manifold in the x-representation consists entirely of caustic points. It is, in fact, a perfect focus (a caustic of order f), as the wave function $\psi_0(x) = \delta(x - x')$ indicates.

For short times, we expect the regions of the initial Lagrangian manifold containing small initial momenta to have evolved only a small amount, so that no momentum-space caustics will have developed. This is illustrated in Fig. 26 for the double-well oscillator. Notice that in the figure, the final Lagrangian manifold is almost a straight line, as was the initial Lagrangian manifold. This is because for short times and small initial momenta, the evolution is dominated by the kinetic energy, which generates linear transformations on phase space. This point will be examined more carefully in a moment.

For longer times, we expect momentum-space caustics to develop. These are illustrated in Fig. 27, in which up to five momentum-space branches are visible. The stretching and kneading of the Lagrangian manifold which is evident in Fig. 27 is leading to structures which have been called "whorls and tendrils" by Berry *et al.*[16] Actually, even for short times, there may be multiple branches, because initial conditions at large

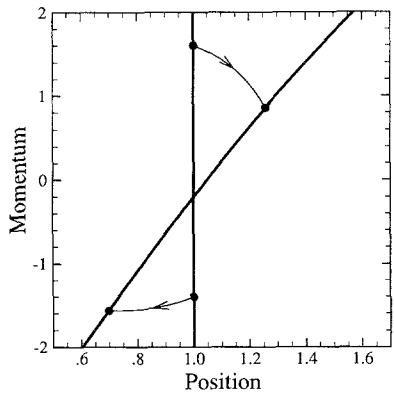

Fig. 26. The initial Lagrangian manifold $x' = 1.0$ at $t' = 0$ and its evolved image at $t'' = 0.2$ for the 1-dimensional double-well oscillator with potential $V(x) = x^4 - x^2$. Two orbits are shown.

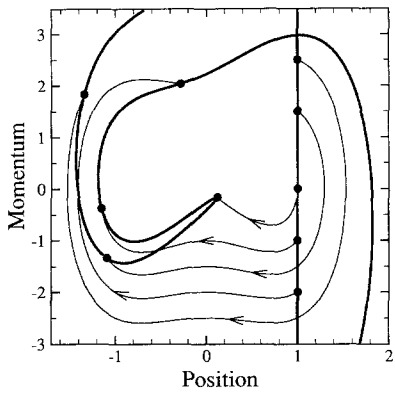

Fig. 27. The initial and final Lagrangian manifolds for the same system at $t'' = 2.0$. Five orbits are shown.

momenta may lead to convolutions in a short time. For example, the Lagrangian manifold in Fig. 26 would show arbitrarily many momentum-space caustics if the diagram were enlarged, because high-momentum particles in a sharply rising potential can bounce many times in a small elapsed time.

We will refer to the one branch illustrated in Fig. 26 as the "principal branch," which exists within bounds on momentum and time which are classical, i.e., of order $\hbar^0$. This branch has a Maslov index in momentum space of $\kappa = 0$, because for short times the orbits do not encounter any momentum-space caustics. This is shown clearly in Fig. 26. The principal branch also has a single-valued projection onto configuration space (except at $t = t'$), so it corresponds to a single branch in the configuration-space wave function as well.

Using the WKB theory leading to Eq. (4.3), but with the replacements $x \to p$, $p \to -x$, we can immediately write down the final momentum-space wave function,

$$\phi(p'', t'') = \sum_b \frac{1}{(2\pi\hbar)^{f/2}} \left| \det \frac{\partial p'}{\partial p''} \right|^{1/2} \times \exp\left\{ \frac{i}{\hbar} [-p'x' + \tilde{R}(p'', t''; p', t')] - i\kappa \frac{\pi}{2} \right\} \tag{5.2}$$

where the initial amplitude and action have been taken from Eq. (5.1). In this equation, x' is regarded as a fixed parameter of the WKB problem.

To find $\psi(x'', t'')$, which is the propagator $K(x'', t''; x', t')$, we simply

Fourier transform Eq. (5.2), using the stationary phase approximation. We note that the momentum-space action $\tilde{R}$ can be written

$$\tilde{R}(p'', t''; p', t') = \int_{\text{orbit}} -x\, dp - H\, dt = R(x'', t''; x', t') - x''p'' + x'p' \tag{5.3}$$

which allows the final phase to be written in configuration-space terms. We find

$$\psi(x'', t'') = \sum_b \frac{e^{i\alpha\pi/4}}{(2\pi\hbar)^{f/2}} \left|\det \frac{\partial p'}{\partial p''}\right|^{1/2} \left|\det \frac{\partial^2 \tilde{R}}{\partial p''\, \partial p'}\right|^{-1/2} \times \exp\left[\frac{i}{\hbar} R(x'', t''; x', t') - i\kappa \frac{\pi}{2}\right] \tag{5.4}$$

where α is the index of inertia of $\partial^2 \tilde{R}/\partial p''\, \partial p' = -\partial x''/\partial p''$. Note that the determinant factors can be combined to give $\det(\partial p'/\partial x'')$.

Assuming we have a kinetic-plus-potential Hamiltonian, we can evaluate the index of inertia explicitly for the principal branch. We simply expand the solution of Hamilton's equations to first order in $\tau = t'' - t'$, to get

$$\begin{aligned} x'' &= x' + \frac{p'}{m}\tau + O(\tau^2) \\ p'' &= p' + F(x', t')\tau + O(\tau^2) \end{aligned} \tag{5.5}$$

where $F(x, t)$ is the force, and then we eliminate p' to solve for x'' as a function of p'', treating x', t', t'' as parameters. This gives

$$x'' = x' + \frac{p''}{m}\tau + O(\tau^2) \tag{5.6}$$

which is the equation $x'' = x''(p'')$ of the final Lagrangian manifold in the momentum projection. The fact that it is a straight line through $O(\tau)$ explains the appearance of Fig. 26; to this order, it does not depend on the potential. From Eq. (5.6) it follows that

$$\frac{\partial \tilde{R}}{\partial p_i''\, \partial p_j'} = -\frac{\partial x_i''}{\partial p_j'} = -\delta_{ij}\frac{\tau}{m} + O(\tau^2) \tag{5.7}$$

so that all the eigenvalues of $\partial^2 \tilde{R}/\partial p''\, \partial p'$ are negative for short times.

Therefore $\alpha = -f$ for the principal branch, and we can write

$$K(x'', t''; x', t') = \frac{1}{(2\pi i\hbar)^{f/2}} \sum_b \left|\frac{\partial p_b'}{\partial x''}\right|^{1/2} \times \exp\left[\frac{i}{\hbar} R_b(x'', t''; x', t') - i\kappa_b \frac{\pi}{2}\right] \tag{5.8}$$

where $i^{f/2}$ means $e^{if\pi/4}$, and where κ_b is now the Maslov index in configuration space relative to the principal branch, for which $\kappa = 0$. It is also the count, multiplicities included, of configuration-space caustics encountered along an orbit, but not including the order-f caustic at $t = t'$. This follows because any given orbit is on the principal branch for short enough elapsed times, when $\kappa = 0$, and because afterward, the configuration space caustics can be counted as in our discussion of the Morse index in Section 4. The configuration-space caustics of the propagator are properly called *conjugate points*, because they are places where there is achieved at least a partial refocusing of the initially perfectly focused collection of particles.

Equation (5.8) is the Van Vleck formula. It is exact for Hamiltonians which are at most a quadratic polynomial in x and p, such as the free particle, the harmonic oscillator, a particle in a constant magnetic field, and the components of angular momentum, e.g., $L_z = xp_y - yp_x$, regarded as an evolution operator. These Hamiltonians generate linear transformations on phase space, so there is always at most one branch. For other Hamiltonians, the Van Vleck formula is only the leading term in an expansion in $\hbar$, and has $O(\hbar)$ corrections.

If x' is regarded as a variable, then the geometrical picture corresponding to the Van Vleck formula at $t = t'$ is that of a foliation of phase space into Lagrangian manifolds as shown in Fig. 16, representing the identity canonical transformation. At later time $t = t''$, the foliation is the image of that in Fig. 16 under the flow; now the canonical transformation is that taking final coordinates (x'', p'') into initial coordinates (x', p'). This is a time-dependent canonical transformation, for which the F_1-type generating function is Hamilton's principal function $R(x'', t''; x', t')$, with the final variables considered "old" and the initial ones considered "new." In quantum mechanics, this canonical transformation corresponds to the unitary change of basis, taking one from the complete set of continuous states $|x''\rangle$ at the final time into the evolved images of the initial states $U(t'', t')\,|x'\rangle$. Of course, the propagator is just the component matrix of this unitary transformation, with continuous indices.

6. CONCLUSIONS

This survey of the geometrical structure of the Van Vleck formula has served several purposes. It has provided an introduction to Maslov's theory, which is particularly elegant in the case of time-dependent problems; it has clarified the relationship between the Van Vleck formula and phase space geometry, which is not only necessary for a proper understanding of the Van Vleck formula, but also for exploring such questions

as its long-time limitations; and it has provided a necessary introduction to the geometrical structure of the Gutzwiller trace formula. The latter subject is developed in considerable detail in ref. 7, in which it is shown how the Gutzwiller trace formula is obtained by projecting one Lagrangian manifold onto another in a doubled phase space, and in which numerous other geometrical features of trace formulas are explored.

ACKNOWLEDGMENTS

I would like to thank Martin Gutzwiller for many interesting discussions, and Jim Morehead for assistance with the figures.

This work was supported by the Office of Energy Research, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract number DE-AC03-76SF00098, and by the National Science Foundation under grant number NSF-PYI-84-51276.

REFERENCES

1. M. C. Gutzwiller, *J. Math. Phys.* **8**:1979 (1967); **10**:1004 (1969); **11**:1791 (1970); **12**:343 (1971).
2. G. S. Ezra, K. Richter, G. Tanner, and D. Wintgen, *J. Phys. B* **24**:L413 (1991); M. Sieber and F. Steiner, *Phys. Rev. Lett.* **67**:1941 (1991); G. Tanner, P. Scherer, E. B. Bogomolny, B. Eckhardt, and D. Wintgen, *Phys. Rev. Lett.* **67**:2410 (1991).
3. Steven Tomsovic and Eric J. Heller, *Phys. Rev. Lett.* **67**:664 (1991).
4. V. P. Maslov, *Théorie des Perturbations et Méthodes Asymptotiques* (Dunod, Paris, 1972); V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston, 1981).
5. J. B. Delos, *Adv. Chem. Phys.* **65**:161 (1986).
6. I. C. Percival, *Adv. Chem. Phys.* **36**:1 (1977).
7. R. G. Littlejohn, *J. Math. Phys.* **31**:2952 (1990).
8. V. Guillemin and S. Sternberg, *Geometric Asymptotics* (American Mathematical Society, Providence, Rhode Island, 1977).
9. J. Leray, *Lagrangian Analysis and Quantum Mechanics* (MIT, Cambridge, Massachusetts, 1981).
10. V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
11. W. H. Miller, *Adv. Chem. Phys.* **25**:69 (1974).
12. L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
13. J. H. Van Vleck, *Proc. Natl. Acad. Sci. USA* **14**:178 (1928).
14. E. N. Economu, *Green's Functions in Quantum Physics* (Springer-Verlag, New York, 1990).
15. Albert Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1966).
16. M. V. Berry, N. L. Balazs, M. Tabor, and A. Voros, *Ann. Phys.* **122**:26 (1979).
17. H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Massachusetts, 1980).
18. R. Abraham and J. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reading, Massachusetts, 1978).
19. V. I. Arnold, *Funct. Anal. Appl.* **1**:1 (1967).

20. S. Levit, K. Möhring, and U. Smilansky, *Ann. Phys.* **114**:223 (1978); S. Levit, K. Möhring, and U. Smilansky, *Ann. Phys.* **114**:223 (1978).
21. Philip Pechukas, *Phys. Rev.* **181**:166, 174 (1969).
22. R. G. Littlejohn and J. M. Robbins, *Phys. Rev. A* **36**:2953 (1987); S. Creagh, R. G. Littlejohn, and J. M. Robbins, *Phys. Rev. A* **42**:1907 (1990); J. M. Robbins, *Nonlinearity* **4**:343 (1991).
23. B. Eckhardt and D. Wintgen, *J. Phys. A* **24**:4335 (1991).
24. Alfredo M. Ozorio de Almeida, *Hamiltonian Systems: Chaos and Quantization* (Cambridge University Press, Cambridge, 1988).