A weighted least action principle for dispersive waves

Jacob Rubinstein and Gershon Wolansky
Mathematics Department, Indiana University, Bloomington, IN 47405, USA, and
Department of Mathematics, Technion, Haifa 32000, Israel

Abstract. We extend the least action principle to continuum systems. The data for the new
principle consists of the intensity of the wave (or rather the wave action) at two instances of
time. We define an appropriate Lagrangian, and formulate a variational problem in terms of
it. The critical points of the functional are used to determine the wave’s phase. The theory is
applicable to the semiclassical limit of a large class of dispersive wave equations. Associating
the wave equation with a Liouville equation for the Wigner distribution function, we are able
to extend the theory to include singular solutions such as caustics.

1 Introduction

The orbits of mechanical systems consisting of finitely many bodies can be found by specifying
the initial and final positions of these bodies, and minimizing the action among all orbits
connecting those terminal points. We derive a similar theory for waves. The information on
the wave consists of wave action distributions at two instances of time $t_1$ and $t_2$. The goal is
to find the wave’s phase. Our theory is formulated in the semiclassical (physical optics) limit
of the underlying wave equation. The variational principle must now provide not only the
individual orbits (rays) but also their endpoints. We are particularly interested in quantum
mechanics wave functions, but the theory is applicable to a wide family of linear dispersive
waves. Therefore we present a general formulation of the theory, starting from the generic
equations of the physical optics limit of high wave number.

The new principle is formulated and justified in section 2. In section 3 we demonstrate how
to apply it in a variety of examples of physical setups. The first formulation of the weighted
least action principle in section 2 corresponds to smooth solutions. We extend the principle in
section 4 to include singular solutions (caustics). Finally, in section 5 we discuss our results,
point out applications to phase sensors, and comment on numerical methods for solving the
variational problem.
2 Formulation

We study wave equations associated with a Lagrangian. Our principle is based on an appropriate weighted action which is derived from the Lagrangian. We shall consider the semiclassical (physical optics) limit of the wave equation. Therefore we introduce a small positive parameter $\epsilon$ and express the wave function $u$ as $u = a(x, t)e^{i\phi(x, t)/\epsilon}$, where $t$ denotes time and $x$ denotes a point in space. We assume that in the physical optics limit the phase $\phi$ satisfies an eikonal equation (dispersion relation):

$$H(\phi_t, \nabla \phi, x) = 0.$$  

(1)

It is useful to solve (1) for $\phi_t$ in the form $\phi_t = D(x, \nabla \phi)$. In some applications there are several such branches of solutions to (1), and we treat them separately. We assume further that the relation (1) is strictly dispersive, that is, the Hessian matrix of $D(x, \xi)$ with respect to $\xi$ is nonsingular. We also introduce a notion of wave action $I$, defined as $I = a^2 H_{\phi_t}$. The wave action $I$ is assumed to satisfy a transport equation

$$I_t + \nabla \cdot (v I) = 0, \quad v = -\nabla_{\xi} D(x, \nabla \phi).$$  

(2)

The physical optics equations (1)-(2) are generic for a wide class of waves. A few examples will be presented in the next section. The derivation of (1)-(2) from the original wave equation can be done, for instance, by a WKB expansion [3] or by the Whitham’s averaging method [7].

We consider the wave problem (1)-(2) in a time interval $t \in (t_1, t_2)$, where the wave action $I$ is given at the two endpoints of the interval, i.e.

$$I(t_1) = I_1, \quad I(t_2) = I_2.$$  

(3)

We assume that $I_1$ and $I_2$ are nonnegative piecewise continuous functions with the same finite integral over the entire space. Our problem is to use the data on $I$ to determine the wave action $I$ and the phase $\phi$ for all $t$. Phase determination is in general harder than energy or intensity measurement. Therefore the weighted action method provides an attractive tool to obtain the wave’s phase.

To formulate a variational principle for the problem above we need to construct an appropriate action for an orbit. For this purpose we define a convex function $D^*(x, v)$ that is related to $D(x, \nabla \phi)$ through

$$D^*(x, v) = \max_{\nabla \phi} (D(x, \nabla \phi) + \nabla \phi \cdot v).$$  

(4)

We initially assume that $D$ is a concave function with respect to its second argument. We shall comment later on other cases.
The Lagrangian $D^*$ is used to introduce the functional

$$E(\sigma, v) = \int_{t_1}^{t_2} \int \sigma(x, t) D^*(x, v(x, t)) \, dx \, dt. \quad (5)$$

Consider the problem of constrained minimization of $E$:

$$\min_{\sigma, v} E(\sigma, v) \quad \text{subject to} \quad \sigma_t + \nabla \cdot (v\sigma) = 0, \quad \sigma(t_1) = I_1, \quad \sigma(t_2) = I_2. \quad (6)$$

To carry out the constrained minimization we introduce a Lagrange multiplier $\varphi$ and look for

$$\max_{\varphi} \min_{\sigma, v} \int_{t_1}^{t_2} \int (\sigma D^*(x, v) + \varphi(\sigma_t + \nabla \cdot (\sigma v))) \, dx \, dt. \quad (7)$$

Integrating by parts the term multiplying $\varphi$, we obtain

$$\max_{\varphi} \min_{\sigma, v} \int_{t_1}^{t_2} \int (\sigma (D^*(x, v) - \varphi_t - v \cdot \nabla \varphi)) \, dx \, dt + \int \varphi I_2 dx - \int \varphi I_1 dx. \quad (8)$$

From the definition of $D^*(x, v)$ we have $D^*(x, v) - v \cdot \nabla \varphi \geq D(x, \nabla \varphi)$, where the minimum is achieved if $v = - \nabla \xi D(x, \nabla \varphi)$. Substituting this formula for $v$ into the minimization problem and using (4) we obtain

$$\max_{\varphi} \min_{\sigma} \int_{t_1}^{t_2} \int (\sigma (D(x, \nabla \varphi) - \varphi_t)) \, dx \, dt + \int \varphi I_2 dx - \int \varphi I_1 dx. \quad (9)$$

Now, since we included the constraint on $\sigma$ in the functional through the Lagrange multiplier $\varphi$, then $\sigma$ is not limited to positive values. Therefore if $D(x, \nabla \varphi) - \varphi_t \neq 0$, then $\min_{\sigma} \int (\sigma (D(x, \nabla \varphi) - \varphi_t)) \, dx \, dt = -\infty$. But this will not be the optimal (maximal) value for all $\varphi$. It follows that $\varphi$ must satisfy the eikonal equation. Together with (6), the pair $(\sigma, \varphi)$ is a solution to (1)-(3). Notice in particular that we also obtain the relation $\min_{\sigma, v} E = \max_{\varphi} \int \varphi I_2 dx - \int \varphi I_1 dx$, where the maximization is taken over all functions $\varphi$ that satisfy the eikonal equation.

To state the weighted least action principle we use $D^*$ to associate an orbit $\bar{x}(t)$ with a Lagrangian $L = D^*(\bar{x}, \bar{x}_t)$ and an action $j = \int D^*(\bar{x}, \bar{x}_t) dt$. We also define the optimal action

$$Q(x, y) = \min_{\bar{x}, \bar{x}_t} \int_{t_1}^{t_2} D^*(\bar{x}, \bar{x}_t) dt,$$ \quad (10)

where the minimization is over all orbits connecting two fixed points $(x, t_1)$ and $(y, t_2)$.

The least action principle associated with the wave problem (1)-(3) is based on the notion of optimal transporting maps that was introduced by Monge in the 18th century. We say that a map $T$ transports $I_1$ into $I_2$, if

$$I_1(x) = I_2(T(x))|J(T)|. \quad (11)$$
Here $T(x)$ is a map from a point $x$ at time $t_1$ to a point $T(x)$ at time $t_2$, and $J(T)$ is the Jacobian of this map. The notation $T#I_1 = I_2$ is commonly used to denote this property. We further define the following functional

$$M(I_1, I_2, T) = \int Q(x, T(x))I_1(x)dx.$$  \hfill (12) 

The weighted least action problem I. Find a mapping $\bar{T}$ such that $\bar{T}#I_1 = I_2$, and

$$M(I_1, I_2, \bar{T}) \leq M(I_1, I_2, T) \quad \forall T#I_1 = I_2,$$  \hfill (13) 

where $Q$ is given in (10).

The relation between the variational problem (13) and the wave problem (1)-(3) is described in the following statement:

The weighted least action principle I. The optimal mapping is given by $\bar{T}(x) = \bar{x}(t_2)$, where the flow $\bar{x}$ is defined by $\bar{x}_t = v$, $\bar{x}(t_1) = x$, and $v$ is the minimizer of $E$ under the constraint (6).

In practice, we start from the variational problem (13). Solving it, we obtain the optimal map $\bar{T}$. For each initial point $x$ and its associated map $\bar{T}(x)$, we compute the full orbit through the optimal action $Q$. The orbit, in turn, determines $v$, and then, since the minimization of $E$ under (6) leads to a solution of (1)-(3), we obtain $\phi$ and $I$.

To justify the principle, integrate the phase $\phi$ along an arbitrary orbit $\eta(t)$ and use (4) to obtain

$$\frac{d\phi}{dt}(\eta(t), t) = \phi_t + \nabla\phi \cdot \eta_t = D(\eta, \nabla\phi) + \nabla\phi \cdot \eta_t \leq D^*(\eta, \eta_t).$$  \hfill (14) 

Let $\bar{x}(t)$ be the orbit defined in the principle, and assume that it maps an initial point $\bar{x}(t_1) = x$ into $\bar{x}(t_2) = T(x)$. If we substitute the orbit $\bar{x}(t)$ for $\eta(t)$, then the inequality in (14) becomes an equality. Integrating between $t_1$ and $t_2$ we get

$$\phi(T(x), t_2) - \phi(x, t_1) = \int_{t_1}^{t_2} D^*(\bar{x}, \bar{x}_t)dt.$$  \hfill (15) 

Multiply the last equation by $I_1$, integrate further over $x$, and use the fact that $T#I_1 = I_2$ and the definition of the optimal action $Q$ to derive

$$\int \phi(x, t_2)I_2(x)dx - \int \phi(x, t_1)I_1(x)dx \geq \int Q(x, T(x))I_1(x)dx.$$  \hfill (16) 

On the other hand, we can integrate (14) along the optimal orbit with respect to $Q$ that connects $x$ with $\bar{T}(x)$. Therefore

$$\phi(\bar{T}(x), t_2) - \phi(x, t_1) \leq Q(x, \bar{T}(x)).$$  \hfill (17)
Multiplying the last inequality by $I_1(x)$, and integrating with respect to $x$ using the transporting property of $T$ we obtain

$$
\int \phi(x, t_2) I_2(x) dx - \int \phi(x, t_1) I_1(x) dx \leq \int Q(x, T(x)) I_1(x) dx.
$$

Finally we deduce from (16) and (18) that $T = T$.

Remark 1. In addition to the optimal map $T$ we can use the flow $\bar{x}(t)$ to define temporary maps $T(t_1) = J$ where $J$ is the identity mapping, and $T(t_2) = \bar{T}$. The analysis above implies also that $T(t)\#I_1 = I(t)$ where $I(t)$ is the wave action at time $t$. To close the circle notice that $I(t)$ satisfies the continuity equation (2).

3 Examples

We illustrate the principle by applying it to a number of wave models.

The Klein-Gordon equation. The linear Klein-Gordon equation models elastic vibrations with a restoring force. The wave $u(x, t)$ satisfies

$$
u_{tt} - c^2 \Delta u + \epsilon^{-2} \alpha^2 u = 0,
$$

where $\alpha$ could depend on $x$. The dispersion relation is

$$
\phi_t^2 - c^2 |\nabla \phi|^2 - \alpha^2 = 0.
$$

Therefore we can identify $D(x, \nabla \phi) = -\sqrt{c^2|\nabla \phi|^2 + \alpha^2}$, and hence $v = \frac{c^2 \nabla \phi}{\sqrt{c^2|\nabla \phi|^2 + \alpha^2}}$. The Lagrangian is

$$
D^*(x, v) = \begin{cases}
-\frac{\alpha}{c} \sqrt{c^2 - |v|^2} & |v| \leq c, \\
\infty & |v| > c.
\end{cases}
$$

Remark 2. When we formulated the weighted least action principle we implicitly assumed that for any $I_1$ and $I_2$ there exists candidate mappings for the minimization. The current example indicates that the functional $\int Q(x, T(x)) I_1(x) dx$ may not be finite. The reason for the difficulty is that a mapping might have to connect pairs of points along orbits that require transport with a velocity greater than $c$. We therefore need to exclude such cases in relativistic models.

The magnetic Schrödinger equation. The nondimensional equation for the wave function $\psi$ of a charged particle in a magnetic field is $(i\epsilon \partial_t + \beta \Phi)\psi = (i\epsilon \nabla - \beta A)^2\psi + V\psi$, where $A$
is the magnetic potential, $\Phi$ is the electric potential, $V$ is an additional potential and $\beta$ is a parameter. The dispersion relation is $\phi_t + |\nabla \phi - \beta A|^2 + \beta \Phi + V = 0$. Therefore,

$$D(x, \nabla \phi) = -|\nabla \phi - \beta A|^2 - \beta \Phi - V,$$

which implies $v = 2(\nabla \phi - \beta A)$. The optimal action is

$$Q(x, y) = \min \int_{t_1}^{t_2} \left( \frac{1}{4} |\vec{x}_t|^2 + \beta \vec{x}_t \cdot A - V - \beta \Phi \right) dt.$$  \hfill (23)

Notice that this example includes the case of the nonmagnetic Schrödinger equation (by setting $\beta = 0$) and the case of the Schrödinger equation for a free particle (by setting $V = 0$).

**Monochromatic scalar waves**

Pure harmonic (in time) waves are described by

$$u_{zz} + \Delta u + \epsilon^2 n^2 u = 0,$$ \hfill (24)

where $n$ is the refraction index, and here $\Delta$ is the two dimensional Laplacian. The wave action (called radiance in optics) is given here on two planes $z = z_1$ and $z = z_2$. Notice that the $z$ variable plays here the role of $t$ in the other examples and in the general formulation. In this example $v = \frac{\nabla \phi}{\sqrt{n^2 - |
abla \phi|^2}}$, and $D^*(x, v) = n\sqrt{1 + v^2}$.

**Elastic waves.**

The vibrations of a plate are described by the equation $u_{tt} + \epsilon^2 \gamma^2 \Delta^2 u = 0$. In this example $x$ is a point in the plane. The dispersion relation is $\phi_t^2 - \gamma^2 |\nabla \phi|^4 = 0$. Therefore $v = 2\gamma^2 \nabla \phi$, and $D^*(x, v) = \frac{1}{4\gamma^2} v^2$. This implies the optimal action

$$Q(x, y) = \min \int_{t_1}^{t_2} \frac{1}{4\gamma^2} |\vec{x}_t|^2 dt.$$ \hfill (25)

Notice that, just like in the case of the Schrödinger equation for a free particle, we can write explicitly $Q(x, y) = \frac{1}{4\gamma^2(t_2-t_1)} |x - y|^2$.

### 4 Critical points of the weighted action and singularities

Solutions to wave problems can exhibit singularities such as caustics. At a caustic rays intersect and the amplitude $I$ in the physical optics limit blows up. The global minimizers of $M$, on the other hand, yield smooth solutions. The purpose of this section is to show that critical points of $M$ (other than the global minimizer) can produce singular solutions.
The orbits, or rays, generated from the weighted least action principle do not intersect. To see why, look at a discrete approximation of the problem by finitely many rays, each carrying with it an equal amount of wave action. Consider two rays, $x_i(t), \ i = 1, 2$ connecting $x_i(t_1) = a_i$ and $x_i(t_2) = b_i$. If the rays intersect at some time $\tau \in (t_1, t_2)$, we can swap the orbits between $\tau$ and $t_2$. The new orbits also transport $I_1$ to $I_2$, and the weighted action $\int Q(x, T(x))I_1(x)$ is unchanged by the swapping. In general, however, the newly generated orbits do not minimize the action $j$ among all orbits connecting $a_1$ to $b_2$, and $a_2$ to $b_1$. Thus the map $T$ cannot be optimal. In particular the optimal map $\bar{T}$ cannot describe caustics. Therefore it is possible for the wave problem (1)-(3) to have solutions that are not captured by (13).

It follows that in order to capture caustics we must look for local minimizers or other critical points of $M(I_1, I_2, T)$. Before doing so, however, we must give an adequate meaning to singular solutions, since the physical optics approximation (1)-(2) breaks down at caustics. An elegant way of constructing wave approximations beyond caustics is the Wigner transform. The idea is to replace the wave $u$ with the Wigner distribution function (WDF) and to consider its evolution in phase space. Taking into account the natural $\epsilon$ scale in our wave equations, we define the WDF of the wave $u(x, t)$ by

$$W(x, \xi, t) = \frac{1}{2\pi} \int u(x - \frac{\epsilon}{2} y, t)\bar{u}(x + \frac{\epsilon}{2} y, t)e^{i\xi y} \, dy.$$ \hspace{1cm} (26)

To simplify the presentation in this section we shall assume that the underlying wave equation has the form

$$i\epsilon u_t = D(x, \epsilon \nabla)u.$$ \hspace{1cm} (27)

This includes the Schrödinger equation, the parabolic (Fresnel) approximation for the Helmholtz equation and other examples. Extending the approach in this section to other wave models is somewhat delicate. Note that in the case of (27) the wave action $I$ can be identified with the intensity $a^2$.

We recall some basic properties of the WDF in the context of transport past singularities (see for example [5] and other references therein for details). In the semiclassical limit $W$ is transported in phase space by the Liouville equation

$$W_t + \{D, W\} = 0,$$ \hspace{1cm} (28)

where $\{D, W\}$ denotes the Poisson brackets of $D$ and $W$. Also, in the limit $\epsilon \to 0$, the WDF of an initial wave with a fast oscillating phase $u(x, 0) = a_0(x)e^{i\phi_0(x)/\epsilon}$ becomes the Wigner measure $W(x, \xi, 0) = I_0(x)\delta(\xi - \nabla \phi_0(x))$, where $I_0 = a_0^2$. The Liouville equation is integrated along the characteristic curves defined by the Hamiltonian system

$$\bar{x}_i(x, \xi, t) = -\nabla_\xi D(\bar{x}, \bar{\xi}), \hspace{0.5cm} \bar{x}(x, \xi, 0) = x,$$ \hspace{1cm} (29)
\[ \xi_t(x, \xi, t) = \nabla_x D(\bar{x}, \bar{\xi}), \quad \xi(x, \xi, 0) = \xi, \quad (30) \]

The WDF is constant along the characteristics. Therefore we can express it at an arbitrary point \((x, \xi, t)\) in terms of the initial conditions

\[ W(x, \xi, t) = I_0(\bar{x}(x, \xi, -t))\delta(R(x, \xi, t)), \quad R(x, \xi, t) = \xi(x, \xi, t) - \nabla\phi_0(\bar{x}(x, \xi, -t)). \quad (31) \]

It can be shown that the equation \(R = 0\) can be solved for \(\xi\) in the form \(\xi = \nabla\phi(x, t)\), where, except for the points where \(R\) is degenerate, \(\phi\) is locally a solution of the eikonal equation (1). If \(R = 0\) has just one solution, then \(W(x, \xi, t)\) can be written as

\[ W = I(x, t)\delta(\xi - \nabla\phi(x, t)), \quad (32) \]

where \(I(x, t) = I_0(\bar{x}(x, \nabla\phi, -t))\) satisfies the transport equation (2). When \(R = 0\) has more than one solution, \(W\) can be written as a sum of solutions like (32), each of them corresponding to a zero of \(R\). The existence of a plurality of solutions for \(R = 0\) is associated with caustics. In such regions we have a coexistence of several waves. The caustic’s boundary is defined in the \((x, t)\) space by the joint conditions \(R = 0, |\nabla_x R|(x, \xi, t) = 0\).

We are now ready to generalize the weighted least action principle so that it would yield also singular solutions. We do it by associating this principle with the Wigner formulation. Since the Wigner picture involves a flow in phase space, it is useful to formulate also the weighted least action principle in this sense. Recall from the analysis in section 2 that, in addition to obtaining the optimal mapping \(\bar{T}\), we also obtain the entire orbit connecting \(x\) and \(\bar{T}\) through the minimization of the action \(Q\). Therefore we can reformulate the weighted action problem as:

**The weighted least action problem II.** Find a mapping \(\bar{T}\) and an initial velocity field \(\xi(x)\) such that \(\bar{T}\) is a critical point of the functional \(M(I_1, I_2, T) = \int Q(x, T(x))I_1(x)dx\) among all maps \(T\) satisfying \(T\#I_1 = I_2\), and where \(\xi(x)\) is the velocity field of the orbits at point \(x\) and time \(t_1\).

Our first step is to characterize the manifold \((x, \xi(x))\) directly from the property that a map \(T\) is a critical point of \(M\). This requires computing the first variation of \(M\) with respect to \(T\). We compute the first variation (and other objects later on) using a general principle that applies to transporting maps. Let \(T^0\) be a mapping that transport \(I_1\) to \(I_2\), i.e. \(T^0\#I_1 = I_2\). It is known [2] that any other mapping \(T\) that transports \(I_1\) to \(I_2\) can be decomposed into the form \(T = T^0 \circ S\), where \(S\) is an \(I_1\)-preserving mapping \((S\#I_1 = I_1)\). We shall study the functional \(M\) by expressing it as a special case of a more general formulation due to Kantorovich [6]. The
idea is to use the critical mapping $\bar{T}$ to define a density $\bar{\lambda}(x, y)$ according to

$$\bar{\lambda}(x, y) = I_1(x)\delta(y - \bar{T}(x)).$$  \hfill{(33)}

Now, applying the decomposition principle we represent any other mapping $T$ that transports $I_1$ to $I_2$ as $T = \bar{T} \circ S$, and define an associated density

$$\lambda^S(x, y) = I_1(Sx)\delta(y - \bar{T}(Sx)).$$ \hfill{(34)}

Instead of searching for critical points of $M$ with respect to $T$, we search for the critical points with respect to $S$ of the functional

$$K(\lambda) = \int \int Q(x, y)\lambda^S(x, y)dx\,dy,$$ \hfill{(35)}

where the density $\lambda^S$ is of the form (34).

We consider a flow of densities $\lambda^{S\tau}(x, y)$ for $\tau$ near 0 defined by

$$\lambda^{S\tau} = (S^\tau \times \mathcal{J})\#\bar{\lambda}.$$ \hfill{(36)}

Here $\bar{\lambda} = \lambda^{S0}$, $S^\tau$ is a flow of $I_1$-preserving mappings such that $S^0 = \mathcal{J}$, and we do not change the $y$ component. Applying a change of variables we write

$$K^\tau = \int \int Q(x, y)\lambda^{S\tau}(x, y)dx\,dy = \int \int Q(S^\tau(x), y)\bar{\lambda}(x, y)dx\,dy.$$ \hfill{(37)}

The advantage of using the Kantorovich formulation is that we can obtain now explicit information on $S^0$. We differentiate (37) with respect to $\tau$ and set $\tau = 0$:

$$\frac{d}{d\tau}K^\tau|_{\tau=0} = \int \int \nabla_x Q(x, y)\frac{d}{dt}S^0(x)\bar{\lambda}(x, y)dx\,dy.$$ \hfill{(38)}

Since $S^\tau$ is $I_1$-preserving, its associated velocity field $\eta^\tau = \eta(S^\tau(x)) = \frac{d}{dt}S^\tau$ satisfies $\nabla \cdot (I_1\eta^\tau) = 0$. We substitute the formula for $\eta$ into (38), and recall the special form (33) of a minimizer to obtain

$$\frac{d}{d\tau}K^\tau|_{\tau=0} = \int I_1(x)\nabla_x Q(x, \bar{T}(x))\eta^0(x)dx.$$ \hfill{(39)}

Finally, we observe that the direction $\eta^0$ is arbitrary except for the constraint $\nabla \cdot (I_1\eta^0) = 0$. Equating the first variation $\frac{d}{d\tau}K^\tau|_{\tau=0}$ to zero, we obtain that $\nabla_x Q(x, T(x))$ is orthogonal to all the divergence-free vectors $I_1(x)\eta(x)$. Therefore there exists a potential $\phi$ such that

$$\nabla_x Q(x, T(x)) = \nabla_x \phi(x).$$ \hfill{(40)}
The function $\nabla_x Q$ can be computed directly from the definition (10). Together with (40) we get

$$\nabla_x Q(x, \bar{T}(x)) = -\nabla_\xi D^*(\bar{x}, \bar{x})|_{t=t_1} = \nabla_x \phi(x).$$

(41)

Solving the last equation for $\bar{x}_t$ we obtain

$$\bar{x}_t|_{t=t_1} = -\nabla_\xi D(x, \nabla_x \phi).$$

(42)

Recall the Hamiltonian flow (29)-(30), where the initial time is now $t_1$. In light of equation (42), we consider the flow (29)-(30) with $\xi = \nabla_x \phi(x)$.

It is natural at this point to construct an initial condition for a Wigner function (or, rather, measure) that is the semiclassical limit of a wave with intensity $I_1$ and phase $\phi$:

$$W_0(x, \xi) = I_1(x) \delta(\xi - \nabla_x \phi).$$

(43)

Clearly $\int W_0(x, \xi)d\xi = I_1(x)$. We propagate $W$ from its initial value $W(x, \xi, t_1) = W_0(x, \xi)$ according to the Liouville equation (28) until $t = t_2$. We would like to prove the relation

$$\int W(x, \xi, t_2)d\xi = I_2(x).$$

(44)

For this purpose we consider the mapping

$$(x'(x, \xi), \xi'(x, \xi)) = (\bar{x}(x, \xi, t_2), \bar{\xi}(x, \xi, t_2))$$

(45)

induced by the Hamiltonian flow (29)-(30). Notice that (29)-(30) are the Hamilton equations associated with the action $Q$. We are able to switch back and forth between the Hamiltonian initial value problem and the Lagrangian boundary value problem since we have identified the initial condition $\xi = \nabla_x \phi$ associated with the orbit connecting $x$ and $\bar{T}(x)$. In particular a critical mapping $\bar{T}$ of $M$ induces the transformation

$$x'(x, \nabla \phi) = \bar{T}(x).$$

(46)

We assert that

$$W(x, \xi, t_1) := W_0(x, \xi) := I_1(x) \delta(\xi - \nabla_x \phi) = I_1(x) \delta(x'(x, \xi) - \bar{T}(x))|x'_\xi|,$$

(47)

where we use $|x'_\xi|$ to denote the absolute value of the Jacobian of the transformation $x'(x, \xi)$ for fixed $x$ (and similar notation in the sequel). To prove (47), multiply $W_0$ and the last term in (47) by a test function $\Psi(x, \xi)$ and integrate with respect to $x$ and $\xi$. For $W_0$ we obtain (see (43))

$$\int \int \Psi(x, \xi) I_1(x) \delta(\xi - \nabla_x \phi) dx d\xi = \int \Psi(x, \nabla_x \phi) I_1(x) dx.$$

(48)
For the last term in (47) we write
\[
\int \int \Psi(x, \xi) I_1(x) \delta(x'(x, \xi) - \bar{T}(x)) |x'_\xi| dx d\xi = \int \int \Psi(x, \nabla_x \phi) I_1(x) \delta(x' - \bar{T}(x)) dx dx',
\]
where we used the characterization \( \xi(x') = \bar{T}(x), x = \nabla_x \phi \) that was found before for the manifold \((x, \xi)\) associated with the critical Monge mapping \( \bar{T} \). The last integral equals \( \int \Psi(x, \nabla_x \phi) I_1(x) dx_x \), which together with (48) establishes (47).

We now integrate
\[
\int W(x', \xi', t_2) d\xi' = \int W(x(x', \xi'), \xi(x', \xi'), t_1) d\xi' = \int I_1(x(x', \xi')) \delta(x' - \bar{T}(x(x', \xi'))) |x'_\xi| d\xi',
\]
where (47) is used in the second equality. To proceed we need the identity
\[
|x'_\xi| = |x_{\xi'}|
\]
that is valid for all symplectic mappings (45). Since we could not find a reference for (51) we prove it in the appendix. Substituting (51) into (50) one gets
\[
\int W(x', \xi', t_2) d\xi' = \int I_1(x) \delta(x' - \bar{T}(x)) dx.
\]
Finally, changing variables from \( x \) to \( \bar{T} \), using \( x' = \bar{T}(x) \) and recalling the relation (11) we obtain (44).

We have thus derived

**The weighted least action principle II.** Let \( \bar{T} \) be a critical point of \( M \), and let \((x, \xi(x))\) be the manifold of initial points and initial velocities of the orbits associated with \( \bar{T} \). Then

(i) \( \xi(x) \) is a gradient of some phase function \( \phi \).

(ii) The density \( I_1 \) and phase \( \phi \) define a wave function whose semiclassical Wigner function \( W(x, \xi, t) \) solves the Liouville equation (28) with the initial condition
\[
W(x, \xi, t_1) = I_1(x) \delta(\xi - \nabla_x \phi).
\]
Moreover, \( W \) satisfies (44).

The next natural question is whether there exist any critical points for \( M \) except for the global minimizer that we considered in section 2. We shall now answer this question affirmatively. Moreover, we argue that (under the caveat of Remark 2 above) for every pair of densities \( I_1 \) and \( I_2 \) there exists a local minimizer \( \bar{T} \) such that the associated wave flow (1)-(3) must pass through a singularity.
The idea is to constrain further the mapping $T$ in a way that will force singularities. Specifically, we consider the problem

$$\min_T \int Q(x, T(x))I_1(x)dx,$$  \hspace{1cm} (53)

where $T$ is constrained by

$$T\#I_1 = I_2, \quad J(T) < 0.$$  \hspace{1cm} (54)

A solution to the problem (53)-(54) must involve singularities. To see why, consider the evolution of the temporary mapping $T(t)$ from $T(t_1) = J$ to $T(t_2) = \bar{T}$. Since $J(J) = 1$, it follows that at some $t$ we have $J(T(t)) = 0$, i.e. the mapping is singular and the physical optics approximation (1)-(2) is not valid anymore. This singularity should not worry us, as we already switched over to the Wigner representation. To establish the existence of a solution to the minimization problem (53)-(54), we need to verify that the feasible set is nonempty, and then check that a local minimum is indeed obtained inside the constraint set.

We first prove that there exist maps $T$ satisfying the constraints (54). Let $\bar{T}$ be the optimal mapping from the weighted least action problem $I$. To generate a mapping $T^c$ that transports $I_1$ to $I_2$ and has a negative Jacobian, we write $T^c = \bar{T} \circ S$. It remains to show that for every $I_1$ there exists an $I_1$-preserving mapping $S$ with a negative Jacobian. For this purpose we consider the standard initial value problem where the wave is given by an amplitude $a_0(x) = \sqrt{I_1}$ and a phase $\phi_0(x)$. We choose a phase such that the ray mapping will have a negative Jacobian between times $t_1$ and $t_2$. For instance, in the case of the Schrödinger equation for a free particle, the eikonal equation is $\phi_t + \frac{1}{2}|\nabla \phi|^2 = 0$, and the rays are straight lines. We thus select here a phase such that the point $(x, y, z)$ will be mapped into the point $(-x, y, z)$. Transporting in this way the initial wave action $I_1$, we obtain at time $t_2$ a wave action $I_2^c$, and we denote the induced ray mapping by $T_2^c$. We can then solve for the global minimizer of the weighted least action problem corresponding to $I_1$ and $I_2^c$, and obtain a smooth ray mapping $\bar{T}_2^c$. Thanks to the decomposition theorem $\bar{T}_2^c = T^2 \circ S^2$ for an $I_1$-preserving mapping $S^2$, where $T^2$ is the solution of the weighted least action problem I for the wave actions $I_1$ and $I_2$. But since $T^2$ has a negative Jacobian, while $\bar{T}_2^c$ has a positive Jacobian, than $S^2$ must have a negative Jacobian, and therefore we can construct a constrained mapping $T^c$ as we wanted, and the feasible set for (54) is indeed not empty.

Suppose we start with a mapping $T$ that satisfies (54). We then continuously deform $T$ while maintaining the constraint that $T$ transports $I_1$ into $I_2$ so as to reduce the energy $\int Q(x, T)I_1dx$. We need to show that such a flow will not carry us out of the region where $J(T) < 0$. Notice that $|J(T)| = I_1(x)/I_2(T(x))$. Therefore $J$ cannot vanish as long as $I_1$ is strictly positive. Moreover, the last condition on $|J|$ implies that if one eigenvalue of $J$
approaches zero, some other eigenvalue becomes very large, and hence $Q(x, T(x))$ will be large too. Therefore an energy decreasing flow of deformations that starts in the region $J(T) < 0$ will always stay there. In fact, one can even write down explicit gradient flows for the energy $\int Q(x, T(x))I_1(x)dx$ ([1], [4]). We therefore expect such a flow to converge to a local minimizer satisfying (54).

5 Discussion

Waves are traditionally studied by prescribing their initial state and in some cases also their initial velocity. We proposed here a different theory in which the wave’s action is given at two instances of time. The principle has simple heuristics, say in the quantum mechanical case. In this example the wave action is the square of the amplitude of the wave function and thus it describes the probability of the particle to be at a point $x$. The information on the intensities $I_1$ and $I_2$ can be interpreted as the distribution of the location of the particle at $t_1$ and $t_2$. The weighted action is then the weighted sum of all the actions with respect to all the possible locations of the particle. The mass transportation constraint means that we only allow a minimizer that transports a realization of particles into another realization of particles.

The global minimizer of the weighted action functional corresponds to smooth solutions of the wave problem. We showed that critical points of that functional are associated with the semiclassical Wigner formulation of the wave equation. Therefore such critical points can generate solutions that pass through caustics.

We assume implicitly that the orbits are sufficiently smooth. Indeed smoothness was proved ([2], [6]) in the case of a free particle $L = |\bar{x}_t|^2$, and recently ([8]) also for the action $L = |\bar{x}_t|^2 + V(\bar{x})$ that corresponds to the Schrödinger equation for a particle under a general potential.

The implicit eikonal equation (1) may have more than one explicit solution $D$. Our formulation applies to all the concave solutions. If $D$ is convex, we replace the max in (4) by min, and then the minimization of $W(7)$, $Q(10)$ and $M(13)$, etc. should be replaced by maximization.

The theory we presented required the wave equation to be dispersive. If the equation is not dispersive, we cannot eliminate $\nabla \phi$ from $v$ and evaluate $D^*$. In particular, our theory does not apply to the classical wave equation $u_{tt} - c^2 \Delta u = 0$. The root for the difficulty is that the original problem (1)-(3) is not well posed for the wave equation. There exist wave actions $I_1$ and $I_2$ for which (1)-(3) has no solution at all. For example, in a one-dimensional setup $I_2$ must be a shift of $I_1$, and therefore there are no optimal transporting mapping $T$ for arbitrary pairs $(I_1, I_2)$. 
The variational approach we presented provides a useful tool for solving the wave problem (1)-(3). One can start with an arbitrary initial mapping $T$ that transports $I_1$ to $I_2$, and evolve it along a gradient flow of the functional $M$, while preserving the constraint(s) on $T$. In particular, a promising application of the new principle is a means for a phase sensor. In [4] we considered in detail the example of monochromatic waves in a medium with an arbitrary refraction index. We showed that under this principle one can measure the wave’s radiance on two different planes in space and use this information to determine the wave’s phase. Solving the variational problem (13) in general, and developing a phase sensor in particular, requires efficient numerical solvers. One approach is to use the Kantrorovich formulation (35), which convert the minimization of $M$ into a linear programming problem. A more direct numerical approach [1], [4] is to formulate, as explained above, a gradient flow for the functional $\int Q(x, T(x))I_1(x)$.

References


Appendix: Proof of (51)

Consider the symplectic mapping (45). Differentiating \( x'(x, \xi) \) with respect to \( \xi' \) we find

\[
0 = x'_\xi' = x'_x x'_\xi + x'_\xi x'_\xi'.
\]  \hspace{1cm} (55)

We shall now show that

\[
x'_x = -(\xi'_\xi')^t,
\]  \hspace{1cm} (56)

that together with (55) will imply (51).

To prove (56) we use the Hamilton characteristic function (mixed eikonal) \( S(x, \xi') \) and the associated generating functions \( \Xi(x, \xi') \) and \( X'(x, \xi') \). By definition we have

\[
X'(x, \xi') = S_{\xi'} = x'(x, \Xi(x, \xi')),
\]

\[
\Xi(x, \xi') = -S_x = \xi(X'(x, \xi'), \xi').
\]  \hspace{1cm} (57)

Differentiating the first equation of the pair (57) by \( x \) and the second equation of the pair by \( \xi' \) we obtain

\[
x'_x = X'_x - x'_\xi \Xi_x,
\]

\[
\xi'_\xi' = \Xi'_\xi' - \xi'_x X'_\xi'.
\]  \hspace{1cm} (58)

From (57) we can also write

\[
x'_\xi \Xi_{\xi'} = X'_{\xi'} \Rightarrow x'_\xi = X'_{\xi'} \left( \Xi_{\xi'} \right)^{-1},
\]

\[
\xi'_x X'_x = \Xi_x \Rightarrow \xi'_x = \Xi_x \left( X'_x \right)^{-1}.
\]  \hspace{1cm} (59)

Putting (57), (58) and (59) all together we get

\[
x'_x = X'_x - X'_{\xi'} \left( \Xi_{\xi'} \right)^{-1} \Xi_x = S_{\xi'x} - S_{\xi' \xi'} \left( S_{\xi'x} \right)^{-1} S_{xx},
\]

\[
\xi'_\xi' = \Xi'_\xi' - \Xi_x \left( X'_x \right)^{-1} X'_\xi' = -S_{x \xi' \xi'} + S_{xx} \left( S_{x \xi'} \right)^{-1} S_{\xi' \xi'},
\]  \hspace{1cm} (60)

which, in turn, implies (56).