A variational principle in optics

Jacob Rubinstein and Gershon Wolansky

Department of Mathematics, Indiana University, Bloomington, Indiana, 47401, and Department of Mathematics, Technion, Haifa 32000, Israel

Received March 16, 2004; revised manuscript received May 14, 2004; accepted May 27, 2004

We derive a new variational principle in optics. We first formulate the principle for paraxial waves and then generalize it to arbitrary waves. The new principle, unlike the Fermat principle, concerns both the phase and the intensity of the wave. In particular, the principle provides a method for finding the ray mapping between two surfaces in space from information on the wave's intensity there. We show how to apply the new principle to the problem of phase reconstruction from intensity measurements.

OCIS codes: 080.0080, 000.3860.

1. INTRODUCTION

The Fermat principle is one of the pillars of optics. It lies at the foundations of geometrical optics, where it provides a theoretical and computational tool to find ray trajectories and hence the phase of a wave. The principle, though, only concerns rays and provides no information on intensity transport. The main goal of this paper is to derive a new variational principle in optics that relates the phase and the intensity of a wave. The new principle is formulated in terms of the geometrical-optics approximation of the wave equation.

Fermat postulated that a light ray travels between two specified points so as to minimize the action $\int n \mathrm{d}l$, where n is the refraction index of the medium. It was later shown that this principle is equivalent to the eikonal equation. In our setup, we are not given the terminal points of a ray. Instead, we are given two intensity distributions on two planes. Our principle determines both the end points of each ray and the ray trajectory.

One of the promising applications of the new principle is as a means for determining the wave's phase from intensity measurements. We therefore start by recalling in Section 2 the idea of the transport-of-intensity equation (TIE) and curvature sensors. This theory was developed for paraxial waves. Therefore we first formulate our new principle in the paraxial regime and for nonhomogeneous media. The theory is developed in Section 3 in full detail. In particular, we explain there the precise meaning of paraxiality in our approximations. This explanation leads us in a natural way to derive (Section 4) the general form of our principle. Finally, in Section 5 we summarize and discuss our results. We also briefly discuss there extending the principle to include singular solutions and practical aspects such as the numerical solution of the variational problem. The numerical questions will be addressed in more detail, together with simulation results, in a sequel.

2. PHASE-RECONSTRUCTION PROBLEM

A central problem in optics is to determine the phase of a wave. The problem is particularly hard when the phase is not very close to being planar or spherical, and therefore interferometry methods are difficult to apply. The need to find the phase arises in a variety of applications including adaptive optics, astronomy, and ophthalmic optics.

A widely used general phase sensor is the Hartmann–Shack device. It consists of an array of lenslets that convert an incoming beam into spots of light on a detection screen. This sensor has a number of drawbacks: The resolution is limited by the size of lenslets, the location of the spot centroids is hard to determine accurately, and the transformation from the location of the centroids of the spots to the phase gradient is only approximate.

In contrast to phase determination, it is relatively easy to measure the wave's intensity. It is therefore tempting to seek methods for finding the phase from intensity measurements. Indeed, Teague[1] proposed such a phase sensor. His method was further developed by Roddier[2] and others. To explain the idea behind such sensors (sometimes called curvature sensors), we consider a complex-valued wave u in the Fresnel regime where the wave equation in a homogeneous medium is approximated by

$$-i \frac{\partial u}{\partial z} = k u + \frac{1}{2k} \Delta u. \tag{1}$$

Here z is the main direction of propagation, u is the wave function, k is the wave number, and ∇ and Δ denote, respectively, the gradient and Laplacian operators in the plane orthogonal to z. Writing $u = A \exp(ik\phi)$, we obtain for the real and imaginary parts of Eq. (1)

$$-\frac{\partial A}{\partial z} = \frac{1}{2A} \nabla (A^2 \nabla \phi), \tag{2}$$

$$\frac{\partial \phi}{\partial z} = 1 + \frac{1}{k^2 A} \Delta A - \frac{1}{2} |\nabla \phi|^2. \tag{3}$$

The first equation can be written more conveniently as an equation for the intensity $I = A^2$:

$$-\frac{\partial I}{\partial z} = \nabla \cdot (I \nabla \phi). \tag{4}$$

Equation (4) is called the TIE. Teague[1] pointed out that Eq. (4) can be thought of as an elliptic partial differential equation for the phase ϕ in terms of the intensity I. Thus he considered Eq. (4) over some domain D in a plane $z = z_0$ and solved it under prescribed boundary conditions (the Dirichlet problem). We note that, strictly speaking, the phase is $k\phi$, but we shall refer here to ϕ alone also as the phase.

The difficulty with Teague's method is that the values of the phase at the boundary ∂D are not easy to measure. Therefore a number of people suggested alternative algorithms related to Eq. (4) that attempt to resolve this issue.

Roddier[2] proposed to use homogeneous Neumann boundary conditions at the boundary ∂D instead of the Dirichlet conditions but did not justify this proposal. Gureyev and Nugent[3] analyzed more carefully the boundary behavior of the wave u. They pointed out that, in practice, the domain D is determined by the regime where the intensity is positive (essentially the image of the aperture). Then they argued that since I vanishes at ∂D, Eq. (4) is singular and has a unique solution without any boundary condition. Lee and Rubinstein[4] showed that a more delicate analysis of the boundary behavior of I is needed. Indeed, Eq. (4) has a unique (up to an additive constant) solution without any boundary condition only if I vanishes at ∂D at a suitable rate. They also devised numerical methods to solve such equations.

Notice that the TIE is only one half of the Fresnel equation. Clearly, a proper solution must satisfy the other half [Eq. (3)], too. This raises the following question: Suppose we measure the intensity I at *two* planes $z = Z_1$, $z = Z_2$; can we use this information to determine the phase by considering jointly Eqs. (4) and (3)? In fact, a measurement of the intensity at two planes is also required for the TIE, since we need to find not only the intensity I but also its derivative I_z. Computing this derivative requires measuring the intensity at two nearby planes. In the question we posed above, however, the two planes can be arbitrarily located.

A partial answer to our question was given by van Dam and Lane.[5] They realized that if the wave, confined to both observation planes, depends only on one variable and if the rays do not intersect, one can order the initial and terminal points of the rays on the two respective screens such that all successive pairs of rays hold between them the same amount of total intensity. Once the ray end points are known, one can determine the phase slopes and from them the phase itself. Van Dam and Lane also tried to extend this approach to the general two-dimensional case. They proposed to sample the intensity, as in the one-dimensional case, in many orientations and to apply the Radon transform to deduce the phase slopes from the obtained integrals. No justification, however, was given for this method, and it is not clear why it should give a good approximation to the phase slopes for arbitrary intensity distributions on the detection screens.

We shall use the new variational principle to solve the problem we posed. We also present a preliminary analysis of a number of numerical schemes for actually computing the phase. To incorporate Eq. (3) in the analysis, we further express the phase ϕ in the form

$$\phi = z + \psi(x, z), \tag{5}$$

where x denotes a point in the plane R^2 and ψ is the perturbation of the phase about the planar term z. Substituting Eq. (5) into Eq. (3), we obtain for $\psi(x, z)$

$$\frac{\partial \psi}{\partial z} + \frac{1}{2}|\nabla \psi|^2 = \frac{1}{2k^2 A}\Delta A. \tag{6}$$

In the small-wavelength approximation we neglect the term on the right-hand side and replace Eq. (6) with

$$\frac{\partial \psi}{\partial z} + \frac{1}{2}|\nabla \psi|^2 = 0. \tag{7}$$

Our discussion on the phase reconstruction was limited to homogeneous media. The variational principle we shall derive, however, is applicable to arbitrary media. When the refraction index n is not constant, we need to include the term $\frac{1}{2}(n^2 - 1)$ in the right-hand side of Eq. (6). Thus the optical problem we consider consists of the following equations and boundary conditions:

$$\frac{\partial I}{\partial z} + \nabla \cdot (I \nabla \psi) = 0, \tag{8}$$

$$\frac{\partial \psi}{\partial z} + \frac{1}{2}|\nabla \psi|^2 = \frac{1}{2}(n^2 - 1) \coloneqq P(x, z). \tag{9}$$

Here $x \in R^2$, $z \in [Z_1, Z_2]$, $I(z = Z_1, x) = I_1(x)$, and $I(z = Z_2, x) = I_2(x)$, where I_1 and I_2 are two given intensity distributions. Equations (8) and (9) together with the side conditions will be denoted collectively as problem (**Op**). In Section 3 we show that problem (**Op**) can be solved by certain optimization problems.

3. VARIATIONAL PROBLEM I: THE PARAXIAL LIMIT

Consider two planes P_1: $z = Z_1$ and P_2: $z = Z_2$. Let I_1 and I_2 be two nonnegative functions given on P_1 and P_2. Optically, the functions I_1 and I_2 are the measured intensities; mathematically, however, we can consider them arbitrary density functions. We assume that the intensities are normalized to 1, and that they have finite second moments:

$$\int I_1(x)\mathrm{d}x = \int I_2(x)\mathrm{d}x = 1,$$

$$\int x^2 I_i(x)\mathrm{d}x < \infty, \qquad i = 1, 2. \tag{10}$$

We recall from geometrical optics that if a point $x \in P_1$ is mapped by a ray into a point $y \in P_2$, if the refraction index near P_1 and P_2 is the same, and if the ray is approximately orthogonal to the planes, then the intensities are related by[6]

$$I_1(x) = I_2(T(x))J(T). \tag{11}$$

Here $T(x)$ is the ray mapping from P_1 to P_2, and $J(T)$ is the Jacobian of this mapping. We shall say that a mapping T satisfying the relation (11) transports I_1 to I_2. We use the formal notation

$$T_{\#}I_1 = I_2. \tag{12}$$

More generally, a mapping T (not necessarily continuous) transports I_1 to I_2 if and only if

$$\int \zeta(T(x))I_1(x)\mathrm{d}x = \int \zeta(x)I_2(x)\mathrm{d}x, \qquad \forall \zeta \in C_0(R^2), \tag{13}$$

where $C_0(R^2)$ is the space of all continuous functions in the plane with compact support. Our first variational principle, denoted by problem (**Mp**), is the following:

Find a map $\bar{T}$ such that $\bar{T}_{\#}I_1 = I_2$ and

$$\begin{aligned} M(I_1, I_2, \bar{T}) &:= \int Q(x, \bar{T}(x))I_1(x)\mathrm{d}x \\ &\leq \int Q(x, T(x))I_1(x)\mathrm{d}x, \qquad \forall T_{\#}I_1 = I_2, \end{aligned} \tag{14}$$

where the action Q is given by

$$\begin{aligned} Q(x, y) &:= Q(x, y, Z_1, Z_2) \\ &\equiv \min \int_{Z_1}^{Z_2} \left\{ \frac{1}{2}\left|\frac{\mathrm{d}x}{\mathrm{d}z}\right|^2 + P[x(z), z] \right\} \mathrm{d}z \end{aligned} \tag{15}$$

and where the minimization is among all orbits $x(z)$ such that $x(Z_1) = x$, $x(Z_2) = y$.

In the homogeneous case ($P \equiv 0$) the action reduces to

$$Q(x, y) = \frac{|x - y|^2}{2(Z_2 - Z_1)}.$$

In this case our variational principle (**Mp**) becomes the quadratic Monge problem (**Mpq**):

Find a map $\bar{T}$ such that $\bar{T}_{\#}I_1 = I_2$ and

$$\int |\bar{T}(x) - x|^2 I_1(x)\mathrm{d}x \leq \int |T(x) - x|^2 I_1(x)\mathrm{d}x, \qquad \forall T_{\#}I_1 = I_2. \tag{16}$$

We shall show that the optimal mapping $\bar{T}$ is the ray mapping of the optical problem. For this purpose, we relate problem (**Op**) to problem (**Mp**) through several additional equivalent optimization problems. We start by introducing a new problem, denoted by (**Wp**), and prove that its solution is the pair (I, ψ) that solves (**Op**).

Theorem 1. Let $\sigma = \sigma(x, z) \geq 0$ and $v = v(x, z) \in R^2$ be solutions of the following optimization problem:

$$\inf_{\sigma,v} W(I_1, I_2; P) = \inf_{\sigma,v} \int_{Z_1}^{Z_2} \int \left(\frac{1}{2}\sigma|v|^2 + P\sigma\right) \mathrm{d}x\mathrm{d}z, \tag{17}$$

subject to the constraints

$$\frac{\partial \sigma}{\partial z} + \nabla \cdot (\sigma v) = 0, \qquad Z_1 \leq z \leq Z_2,$$

$$\sigma(x, Z_i) = I_i(x), \qquad i = 1, 2. \tag{18}$$

Then

$$\sigma = I, \qquad v = \nabla\psi, \tag{19}$$

where I and ψ solve (**Op**).

Proof. Recall that any vector field in R^2 is the orthogonal sum of a gradient $\nabla\varphi$ and a vector field w such that $\nabla \cdot (\sigma(x, z)w) = 0$. This decomposition holds for any z. Setting $v = \nabla\varphi + w$, we obtain

$$\int \sigma(x, z)|v(x, z)|^2\mathrm{d}x = \int \sigma(x, z)|\nabla\varphi(x, z)|^2\mathrm{d}x + \int \sigma(x, z)|w(x, z)|^2\mathrm{d}x.$$

Clearly, for any candidate σ, the choice $w = 0$ reduces the energy W without affecting constraint (18). Therefore the optimal choice for v must be of the form $v = \nabla\varphi$ for some potential φ.

To further characterize φ, we equate to zero the first variation of the energy in Eq. (17), taking into account constraints (18). We therefore write $\sigma = I + \epsilon\alpha$, $\varphi = \psi + \epsilon\beta$, where I and $v = \nabla\psi$ solve constraint (18) and ϵ is a small positive number. Substituting σ and φ into constraints (18), we obtain

$$\frac{\partial \alpha}{\partial z} + \nabla \cdot (\alpha\nabla\psi) + \nabla \cdot (I\nabla\beta) = O(\epsilon). \tag{20}$$

The first variation of the energy is

$$\delta W = \epsilon \int\int \left(\frac{1}{2}\alpha|\nabla\psi|^2 + I\nabla\psi \cdot \nabla\beta + \alpha P\right)\mathrm{d}x\mathrm{d}z + O(\epsilon^2). \tag{21}$$

Integrating the second term in the integrand by parts with respect to the x variable, we obtain for any $z \in [Z_1, Z_2]$:

$$\int I\nabla\psi \cdot \nabla\beta\mathrm{d}x = -\int \psi\nabla \cdot (I\nabla\beta)\mathrm{d}x.$$

Integrating now on $R^2 \times [Z_1, Z_2]$, using Eq. (20), and then performing another integration by parts, we get

$$\begin{aligned} \int\int I\nabla\psi \cdot \nabla\beta\mathrm{d}x\mathrm{d}z &= \int\int \left[\frac{\partial\alpha}{\partial z} + \nabla \cdot (\alpha\nabla\psi)\right]\psi\mathrm{d}x\mathrm{d}z + O(\epsilon) \\ &= -\int\int \alpha\left(\frac{\partial\psi}{\partial z} + |\nabla\psi|^2\right)\mathrm{d}x\mathrm{d}z + O(\epsilon), \end{aligned}$$

where we used the fact that constraints (18) imply $\alpha(x, Z_1) = \alpha(x, Z_2) = 0$. Substituting this equation into Eq. (21) and equating the first variation to zero, we obtain that ψ solves Eq. (9), and then constraint (18) imply that I solves Eq. (8), as required.

We proceed to show that the infimum of the functional W equals M defined in expression (14). Let (I, ψ) be the solution of problem (**Wp**). We use the phase ψ, i.e., the solution to the Hamilton–Jacobi equation (9), to define the following flow:

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}z} = \nabla\psi(\bar{x}(z), z), \qquad \bar{x}(Z_1) = x. \tag{22}$$

The flow (22) induces a mapping

$$T^z_{Z_1}(x) \coloneqq \bar{x}(z). \tag{23}$$

Proposition 2. The mapping (23) transports I_1 to $I(x, z)$.

Proof. We define

$$\tilde{I}(x, z) = J(T^z_{Z_1})I(T^z_{Z_1}(x), z). \tag{24}$$

A standard result in the theory of ordinary differential equations states that the Jacobian j of the mapping τ induced by the flow generated by an equation of the form $\mathrm{d}x/\mathrm{d}t = f(x)$ satisfies the identity $\mathrm{d}j/\mathrm{d}t = j\nabla\cdot f$. Applying this identity to the flow (22), we find

$$\frac{\mathrm{d}J(T^z)}{\mathrm{d}z} = J(T^z)\Delta\psi(T^z, z). \tag{25}$$

Therefore

$$\frac{\partial\tilde{I}}{\partial z} = J(T^z_{Z_1})\left(I\Delta\psi + \nabla I\cdot\nabla\psi + \frac{\partial I}{\partial z}\right)_{\left(T^z_{Z_1}, z\right)} = 0,$$

where the last equality follows from the assumption that I satisfies Eq. (8). Since $\tilde{I}$ does not depend on z, we can write $\tilde{I}(x, z) = \tilde{I}(x, Z_1) = I_1$. Replacing $\tilde{I}(x, z)$ in Eq. (24) with I_1 we obtain

$$I_1(x) = J(T^z_{Z_1})I(T^z_{Z_1}(x), z), \tag{26}$$

which, on recalling Eq. (11), proves our assertion.

We are now ready to state the main result of this section.

Theorem 3. The mapping $T = T^{z=Z_2}_{Z_1}$, where $T^z_{Z_1}$ is defined in Eq. (23), is the optimal mapping $\bar{T}$, i.e.,

$$\bar{T} = T^{Z_2}_{Z_1}. \tag{27}$$

In addition,

$$\inf_{\sigma,v} W(I_1, I_2; P) = \int Q(x, \bar{T}(x))I_1(x)\mathrm{d}x.$$

Proof. Let (I, ψ) be a solution to the problem (**Wp**). Integrating the Hamilton–Jacobi equation (9) along an arbitrary orbit $x = \xi(z)$, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}\psi(\xi(z), z) &= \nabla\psi\cdot\frac{\mathrm{d}\xi}{\mathrm{d}z} + \frac{\partial\psi}{\partial z}\\ &= -\frac{1}{2}\left|\frac{\mathrm{d}\xi}{\mathrm{d}z} - \nabla\psi\right|^2 + \frac{1}{2}\left|\frac{\mathrm{d}\xi}{\mathrm{d}z}\right|^2\\ &\quad + P(\xi(z), z).\end{aligned} \tag{28}$$

We first use identity (28) for the special case $\xi = \bar{x}$. Integrating Eq. (28) from Z_1 to Z_2, we write

$$\begin{aligned}&\psi(T^{Z_2}_{Z_1}(x), Z_2) - \psi(x, Z_1)\\ &\quad = \int_{Z_1}^{Z_2}\left[\frac{1}{2}\left|\frac{\mathrm{d}\bar{x}}{\mathrm{d}z}\right|^2 + P(\bar{x}(z), z)\right]\mathrm{d}z \geqslant Q(x, T(x)).\end{aligned} \tag{29}$$

Thanks to expression (29) and to the conclusion we derived above that T transports I_1 to I_2, we can write

$$\begin{aligned}&\int \psi(x, Z_2)I_2(x)\mathrm{d}x - \psi(x, Z_1)I_1(x)\mathrm{d}x\\ &\qquad \geqslant \int Q(x, T(x))I_1(x).\end{aligned} \tag{30}$$

We now show that the left-hand side of expression (30) is nothing but $\inf_{\sigma,v} W(I_1, I_2; P)$ in a disguised form. We thus calculate

$$\begin{aligned}E(\psi, I_1, I_2) &\coloneqq \int \psi(x, Z_2)I_2(x)\mathrm{d}x - \psi(x, Z_1)I_1(x)\mathrm{d}x\\ &= \int_{Z_1}^{Z_2}\int \partial_z(\psi(x, z)I(x, z))\mathrm{d}x\,\mathrm{d}z\\ &= \int_{Z_1}^{Z_2}\int\left(\frac{\partial I}{\partial z}\psi + I\frac{\partial\psi}{\partial z}\right)\mathrm{d}x\,\mathrm{d}z.\end{aligned}$$

Using Eqs. (8) and (9) and then integrating by parts, we find that the last expression equals

$$\begin{aligned}&= \int_{Z_1}^{Z_2}\int\left[-\psi\nabla\cdot(I\nabla\psi) - \frac{1}{2}I|\nabla\psi|^2 + IP\right]\mathrm{d}x\,\mathrm{d}z\\ &= \int_{Z_1}^{Z_2}\int\left(\frac{1}{2}I|\nabla\psi|^2 + IP\right)\mathrm{d}x\,\mathrm{d}z = W(I_1, I_2; P).\end{aligned}$$

We therefore obtain from expression (30)

$$\begin{aligned}W(I_1, I_2; P) &\geqslant \int Q(x, T(x))I_1(x)\mathrm{d}x\\ &\geqslant \int Q(x, \bar{T}(x))I_1(x)\mathrm{d}x.\end{aligned} \tag{31}$$

To complete the proof, we shall establish now the reverse inequality in expression (31). For this purpose we prove the following:

Proposition 4. Let T be any mapping satisfying $T_{\#}I_1 = I_2$ and let ζ be any function satisfying the Hamilton–Jacobi equation (9). Then

$$E(\zeta, I_1, I_2) \leqslant \int Q(x, T(x))I_1(x)\mathrm{d}x. \tag{32}$$

In particular,

$$\begin{aligned}E(\psi, I_1, I_2) = \max_{\zeta}\Bigg[&\int \zeta(x, Z_2)I_2(x)\mathrm{d}x\\ &- \int \zeta(x, Z_1)I_1(x)\mathrm{d}x\Bigg],\end{aligned} \tag{33}$$

where the maximum is taken over all functions $\zeta = \zeta(x, z)$, which satisfy

$$\frac{\partial\zeta}{\partial z} + \frac{1}{2}|\nabla_x\zeta|^2 \leqslant P(x, z).$$

Assuming this proposition, we can substitute the function ψ that solves Eq. (9) for ζ in the left-hand side of expression (32), substitute the optimal mapping $\bar{T}$ for T in the right-hand side, and conclude that the two inequalities in expression (31) must, in fact, be equalities; this establishes Theorem 3.

To prove Proposition 4, we return to the integration formula (28). This formula holds for any solution of the Hamilton–Jacobi equation, so, in particular, it holds for ζ. For the orbit we choose the curve that connects x with $T(x)$ and minimizes the action $Q[x, T(x)]$. We thus obtain the inequality

$$\zeta(T(x), Z_2) - \zeta(x, Z_1) \leqslant Q(x, T(x)).$$

Multiplying the last inequality by $I_1(x)$ and integrating with respect to x, we get

$$\int \zeta(T(x), Z_2)I_1(x)\mathrm{d}x - \int \zeta(x, Z_1)I_1(x)\mathrm{d}x \leqslant \int Q(x, T(x))I_1(x)\mathrm{d}x.$$

Since T transports I_1 into I_2, the first term on the left-hand side of the last inequality equals $\int \zeta(x, Z_2)I_2(x)\mathrm{d}x$. This completes the proof of the proposition and the theorem.

Theorem 1 says that the minimizing pair (I, ψ) for the functional W is a solution to the problem (**Op**). Theorem 3 states that the flow induced by ψ generates the optimal mapping $\bar{T}$. Therefore by solving the variational problem we obtain complete information on the ray mapping and hence the phase ψ of problem (**Op**). Notice that the functional $M(I_1, I_2; \bar{T})$ defines a metric measuring the distance between the intensities I_1 and I_2.

In the special case of the quadratic Monge problem, corresponding to the optical setup of a homogeneous medium, the action Q is minimized by the straight line (ray) connecting x and $T(x)$. Therefore the optimal map $\bar{T}$ in this case is given explicitly by

$$T_{Z_1}^{\ Z_2}(x) = x + \nabla_x\psi(x, Z_1)(Z_2 - Z_1). \tag{34}$$

Given two intensity distributions, and assuming that they are related by the paraxial Fresnel equations, the variational problem (**Mp**) provides us with a theoretical and practical tool to find a phase map that connects these intensities. The optimization formulation involves the action Q. It would be interesting to see how this action is related to the classical Fermat action. This analysis requires us to first understand the asymptotic regime in which the parabolic wave equation holds for uniform or nonuniform media.

To study this regime, we introduce a small positive parameter ε. We assume that the refraction index is of the form

$$n(x, z) = 1 + \varepsilon P(x, \varepsilon^{1/2}z), \tag{35}$$

where we assume without loss of generality that the background refraction index is 1. We then seek solutions of the eikonal equation $(\partial\phi/\partial z)^2 + |\nabla\phi|^2 = n^2$, where ϕ is the wave's phase, of the form

$$\phi(x, z) = z + \varepsilon^{1/2}\psi(x, \varepsilon^{1/2}z). \tag{36}$$

Substituting ϕ into the eikonal equation, we find that to leading order ψ satisfies Eq. (9).

The scaling above means that the variation in the refraction index is weak and slowly varying in the z direction. It also means that we deal with approximately paraxial waves. Consider now the Fermat variational principle

$$\min \int_x^y n(x)\mathrm{d}l, \tag{37}$$

where the minimization is over all orbits connecting x and y and $\mathrm{d}l$ is a length element of the orbit. The paraxial approximation amounts to $\mathrm{d}l = [1 + \frac{1}{2}(\mathrm{d}x/\mathrm{d}z)^2]\mathrm{d}z$. The scaling for ψ in Eq. (36) implies that the initial condition for the ray $x(z)$ must satisfy $\mathrm{d}x/\mathrm{d}z(z = Z_1) = O(\varepsilon^{1/2})$. Substituting the expansion for $\mathrm{d}l$ and the form (35) for n into the Fermat action [expression (37)], we obtain

$$\begin{aligned}\min \int_x^y n(x)\mathrm{d}l &= \min \int_{Z_1}^{Z_2}(1 + \varepsilon P)\left[1 + \frac{1}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}z}\right)^2\right]\mathrm{d}z \\ &\quad + o(\varepsilon) \\ &= (Z_2 - Z_1) + Q(x, y) + o(\varepsilon).\end{aligned} \tag{38}$$

Therefore the action Q is indeed an approximation of the Fermat action.

The mathematical analysis is valid for the optical problem (**Op**) regardless of its origin. It is particularly interesting to note that Eqs. (8) and (9) form the semiclassical limit of the Schrödinger equation. The function P then has the interpretation of the potential of the physical system, and the z coordinate represents time. Therefore the variational principle (**Mp**) means that if we are given the absolute value of the wave function everywhere in space at two different times Z_1 and Z_2 we can find the phase of the wave function at all times $t \in (Z_1, Z_2)$.

Some of the results presented in this section, and, in particular, the connection between problems (**Mp**) and (**Wp**) were also derived (by different arguments) for the special case of the quadratic Monge problem in Refs. 7–9. Our proofs are formal in the sense that we tacitly assume that all the functions are sufficiently smooth. A complete rigorous analysis of existence, uniqueness, and regularity of the solutions to problems (**Mp**) and (**Wp**) is delicate and lies beyond the scope of this paper. We refer the reader to Refs. 7 and 8. For the sake of completeness, though, we list a number of basic results that can be obtained for our action Q by the tools of these references:

1. There exists a minimizer $I = I(x, z)$ of (**Wp**) that satisfies the end conditions $I(x, Z_1) = I_1(x)$, $I(x, Z_2) = I_2(x)$. This minimizer may be nonunique.
2. If P is continuously differentiable, then there exists a maximizer ψ of E that is a Lipschitz function and satisfies the equation $(\partial\psi/\partial z) + \frac{1}{2}|\nabla_x\psi|^2 = P$ almost everywhere.
3. A lot more is known in the special but important case of homogeneous media where $P \equiv 0$. For example, if the intensities I_1 and I_2 are continuous (or even just L_1) functions, then the minimizer of the Monge problem is unique. Furthermore, a wealth of regularity results are known in this case.[7]

4. VARIATIONAL PROBLEM II: GENERAL WAVES

We showed that the data embedded in the intensity distributions given on two planes suffice to find the phase of a wave in the paraxial regime. This raises the natural question of whether the result can be extended to general waves. In this section we provide a positive answer to this question. One of the basic differences in the general case is that we should not talk any more of transporting intensities but rather of transporting radiance functions. The main ideas are to use the Fermat action itself (in a suitable form) as the action in the optimization problem and to phrase the intensity equation as an equation for the radiance.

We start by formulating the optical setup. Consider a solution to the Helmohltz equation

$$\Delta u + k^2 n^2(x, z) u = 0 \tag{39}$$

of the form $u = A(x, z)\exp[ik\phi(x, z)]$. Expanding as usual in large k, we obtain[6] the eikonal equation $(\partial\phi/\partial z)^2 + |\nabla\phi|^2 = n^2$, for the phase ϕ, and the transport equation

$$\partial_z\left(I\frac{\partial\phi}{\partial z}\right) + \nabla\cdot(I\nabla\phi) = 0, \tag{40}$$

for the intensity $I = A^2$. Here, as before, x denotes a point in the plane orthogonal to the z direction, and ∇ is the two-dimensional gradient. Since we are not limited now to paraxial rays, we have some freedom in choosing the z axis. When considering general waves, it is more appropriate to analyze the radiance and not the intensity, since the radiance is the conserved quantity. We therefore define the radiance $\rho(x, z) = I(\partial\phi/\partial z) = I(n^2 - |\nabla\phi|^2)^{1/2}$ and write the transport equation (40) with respect to it. We consider the case in which the radiance is given on two parallel planes and thus choose the z axis to be orthogonal to the planes.

We are now ready to formulate the optical problem (**O**): Find the phase function $\phi(x, z)$ and the radiance function $\rho(x, z)$ such that ρ and ϕ satisfy

$$\frac{\partial\rho}{\partial z} + \nabla\cdot\left[\rho\frac{\nabla\phi}{(n^2 - |\nabla\phi|^2)^{1/2}}\right] = 0, \qquad Z_1 < z < Z_2, \tag{41}$$

$$\left(\frac{\partial\phi}{\partial z}\right)^2 + |\nabla\phi|^2 = n^2, \qquad Z_1 < z < Z_2, \tag{42}$$

subject to $\rho(z = Z_1, x) = \rho_1(x)$, $\rho(z = Z_2, x) = \rho_2(x)$, where ρ_1 and ρ_2 are two given radiance distributions. The question we pose is whether problem (**O**) is solvable and, in particular, whether we can associate it with a variational principle. The analysis in this section is often similar to that in Section 3. Therefore we only outline the proofs and highlight the differences.

Our second variational problem, denoted by problem (**M**), is the following:

Find a map $\bar{T}$ such that $\bar{T}_{\#}\rho_1 = \rho_2$ and

$$\int \mathcal{Q}(x, \bar{T}(x))\rho_1(x)\mathrm{d}x \leqslant \int \mathcal{Q}(x, T(x))\rho_1(x)\mathrm{d}x, \qquad \forall T_{\#}\rho_1 = \rho_2, \tag{43}$$

where the action Q is given by

$$\mathcal{Q}(x, y) = \min\int_x^y n\,\mathrm{d}l = \min\int_{Z_1}^{Z_2} n(x, z)\left(1 + \frac{1}{2}\left|\frac{\mathrm{d}x}{\mathrm{d}z}\right|^2\right)^{1/2}\mathrm{d}z \tag{44}$$

and where the minimization is among all orbits $x(z)$ such that $x(Z_1) = x$, $x(Z_2) = y$.

We now show that the optimal mapping $\bar{T}$ is the ray mapping associated with (**O**). A key point in the analysis in Section 3 was the introduction of the corresponding Lagrangian (17). We thus proceed to define an appropriate Lagrangian and argue that its minimization is equivalent to the optical problem (**O**):

Theorem 5. Let $\rho = \rho(x, z) \geqslant 0$ and $v = v(x, z) \in R^2$ be solutions of the following optimization problem:

$$\inf_{\rho,v}\mathcal{W}(I_1, I_2; P) := \inf_{\rho,v}\int_{Z_1}^{Z_2}\int n\rho\sqrt{1 + v^2}\,\mathrm{d}x\mathrm{d}z, \tag{45}$$

subject to the constraints

$$\frac{\partial\rho}{\partial z} + \nabla\cdot(\rho v) = 0, \qquad Z_1 \leqslant z \leqslant Z_2,$$

$$\rho(x, Z_i) = \rho_i(x), \qquad i = 1, 2. \tag{46}$$

Then

$$v = \frac{\nabla\phi}{(n^2 - |\nabla\phi|^2)^{1/2}}, \tag{47}$$

where ϕ solves Eq. (42).

Proof. We first characterize v. For this purpose we introduce a Lagrange multiplier φ, and, fixing ρ, look for

$$\inf_v \int\int \left[\rho n\sqrt{1 + v^2} + \varphi\left(\frac{\partial\rho}{\partial z} + \nabla\cdot(\rho v)\right)\right]\mathrm{d}x\mathrm{d}z. \tag{48}$$

Integrating by parts the term multiplying φ, we obtain

$$\inf_v \int\int \rho\left(n\sqrt{1 + v^2} - \frac{\partial\varphi}{\partial z} - v\cdot\nabla\varphi\right)\mathrm{d}x\mathrm{d}z + \int\varphi\rho_2\mathrm{d}x - \int\varphi\rho_1\mathrm{d}x. \tag{49}$$

Using the inequality

$$n\sqrt{1 + v^2} - \sqrt{n^2 - p^2} \geqslant v\cdot p, \tag{50}$$

which holds for any vectors v and p, we obtain that the functional is minimized at

$$v = \frac{\nabla\varphi}{(n^2 - |\nabla\varphi|^2)^{1/2}}. \tag{51}$$

Notice that, for this choice of v, the functional $\mathcal{W}$ takes the form

$$\mathcal{W} = \iint \rho\left[(n^2 - |\nabla\varphi|^2)^{1/2} - \frac{\partial\varphi}{\partial z}\right]\mathrm{d}x\mathrm{d}z + \int \varphi\rho_2\mathrm{d}x - \int \varphi\rho_1\mathrm{d}x. \tag{52}$$

The characterization of the Lagrange multiplier φ as the solution of the eikonal equation (42) is similar to the proof of Theorem 1, so we do not spell out the details.

By the same method of proof as for Proposition 2, one can show that the flow $T^z_{Z_1}(x) = \bar{x}(z)$, defined by

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}z} = v = \frac{\nabla\phi}{(n^2 - |\nabla\phi|^2)^{1/2}}, \qquad \bar{x}(Z_1) = x, \tag{53}$$

transports ρ_1. We thus proceed to show that this flow induces the optimal flow $\bar{T}$. The only difference from the proof of Theorem 3 is in the step (28) where we integrated the eikonal equation. We need to integrate an arbitrary solution ψ of the eikonal equation (42) along an arbitrary orbit. We obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}\psi(\xi(z), z) &= \nabla\psi\cdot\frac{\mathrm{d}\xi}{\mathrm{d}z} + \frac{\partial\psi}{\partial z} = \nabla\psi\cdot\frac{\mathrm{d}\xi}{\mathrm{d}z} \\ &\quad + (n^2 - |\nabla\psi|^2)^{1/2} \\ &\leqslant n\left[1 + \left(\frac{\mathrm{d}\xi}{\mathrm{d}z}\right)^2\right]^{1/2},\end{aligned} \tag{54}$$

where we used inequality (50).

We can therefore state the following:

Theorem 6.

(1) The optimal mapping $\bar{T}$ for problem (**M**) induces a ray mapping for optical problem (**O**).

(2)

$$\begin{aligned}\inf_{\rho,v}\int_{Z_1}^{Z_2}\int \rho n\sqrt{1+v^2}\mathrm{d}z\mathrm{d}x &= \inf_{T}\int \mathcal{Q}(x, T(x))\rho_1\mathrm{d}x \\ &= \sup_{\phi}\int \phi\rho_2\mathrm{d}x - \int\phi\rho_1\mathrm{d}x,\end{aligned} \tag{55}$$

where the first integral is minimized under constraints (46), the second integral is minimized among all maps T that transport ρ_1 to ρ_2, and the last integral is maximized among all functions ϕ that satisfy $(\partial\phi/\partial z)^2 + |\nabla\phi|^2 \leqslant n^2$.

5. DISCUSSION

The theory developed in the previous sections provides us with a useful tool for determining the phase from intensity measurements. The optical problem (**Op**) or (**O**) is hard to solve, since we are given information on the intensity or radiance at two separate planes, constrained by certain differential equations connecting them. The variational problem (**Mp**) or (**M**), on the other hand, provides a direct variational characterization of the solution.

One may interpret our result as a way to give a physical meaning to the notion of a ray. Rays are mathematical rather than physical entities. One cannot measure rays directly. The variational principle we derived provides a means for measuring rays. We actually measure intensities, and then, through the mathematical procedure of finding the optimal mapping, we identify the individual rays.

Another interesting conclusion relates to the transport of the radiance ρ [Eq. (41)]. One may consider two adjacent planes and use the measured ρ on both of them to estimate ρ_z. Equation (41) can then be treated as an equation for the unknown phase ϕ. Therefore Eq. (41) can be called the transport-of-radiance equation. This equation defines a nonlinear curvature sensor that replaces the TIE (8) in the nonparaxial case.

The solution of the variational problems (**Mp**) and (**M**) is nonsingular in the sense that we now explain. Consider first the paraxial homogeneous case [problem (**Mp**) with constant refraction index]. The solution of the transport equations (8) and (9) may develop caustics. From the ray-mapping perspective, a caustic is a manifold in space in which rays intersect. The solution to our variational principle (**Mpq**) cannot capture such singularities; namely, the optimal ray mapping $\bar{T}$ does not have intersecting rays. To see why, look at a discrete approximation of the problem by finitely many rays, each carrying with it an equal amount of intensity. Consider two rays, $x_i(z)$, $i = 1, 2$, connecting $x_i(Z_1) = a_i$ and $x_i(Z_2) = b_i$. If the rays intersect at some point $\tau \in (Z_1, Z_2)$, we can swap the orbits between τ and Z_2. The new orbits also transport I_1 to I_2, and the integral $\int Q(x, \bar{T}(x))I_1(x)$ is unchanged by the swapping. In general, however, the newly generated orbits do not minimize this integral among all orbits connecting a_1 to b_2 and a_2 to b_1. Thus the map $\bar{T}$ cannot be optimal.

Since caustic solutions are, of course, feasible for appropriately selected I_1 and I_2, it follows that there are solutions for Eqs. (8) and (9) that are not captured by the variational problem (**Mpq**). In fact, the entire formulation must be modified, since the optical problem (**Op**) is not well defined in the presence of caustics. A generalized formulation of the paraxial equations (8) and (9) can be written down in terms of the Wigner transform of the wave function. The Wigner transform describes the wave in phase space. The geometrical-optics equations for the wave function are replaced by an appropriate Liouville equation for its Wigner transform. Indeed, in Ref. 10 we show that certain critical points (other than the global minimum) of the functional $\int Q(x, T(x))I_1(x)\mathrm{d}x$ (under the constraint $T_\# I_1 = I_2$) provide ray mappings associated with solutions of the Liouville equation for the Wigner transform that are associated with caustics for the solutions of Eqs. (8) and (9).

Another type of singularity in wave problems is the vortex or phase singularity.[11] Vortices are associated with zeros of the intensity function. It is known that the intensity distributions alone cannot determine the degree (circulation) of the phase around the zeros of I, and therefore the solution to the optical problem (**Op**) may not be unique. In such cases, the variational principle (**Mpq**) captures the solution for which the phase has degree zero around all zeros of I. This is a consequence of a deep result by Brenier[12] stating that in the (**Mpq**) problem the optimal mapping $\bar{T}$ is a gradient of convex function Ψ. Recalling Eq. (34), we conclude $\psi = 1/(Z_2 - Z_1)(\Psi - |x|^2)$. Since Ψ is convex, it must be a single-valued function, and the last formula implies that ψ is also single valued. Therefore the degree of ψ cannot be nonzero. Our discussion of singular solutions was restricted to the paraxial regime. Upgrading it to more general wave problems is very challenging, and we are now pursuing this question.

It remains to look for solution methods for the variational problem. In the remainder of this section we shall briefly discuss this problem. In a sequel paper we shall elaborate on the subject and report on numerical simulations. To simplify the presentation, and since the problem of phase reconstruction is commonly considered for homogeneous media in the paraxial limit, we shall concentrate now on the quadratic Monge variational problem (**Mpq**). Similar methods and considerations also apply to problems (**Mp**) and (**M**).

The first thing to notice is that our derivation in Section 2 was formulated for intensity functions defined over the entire plane. The theory is valid, of course, also for the special case where I_1 and I_2 are supported on finite domains, say, D_1 and D_2, respectively. The question is how to define these domains in the optical situation that is of interest here. A natural option is to choose the aperture's image. It is not obvious, however, how to define this image. One way of doing it is to select an initial intensity threshold and use it to determine the location of the image boundary. In addition, we have to handle the question of how to deal numerically with cases in which I vanishes or nearly vanishes at the boundary.

A small number of numerical algorithms have been proposed for the Monge mass transportation problem. We shall briefly describe here two such methods. The first method converts the Monge problem into a linear programming problem. This is done by associating problem (**Mp**) [or (**M**)] with the Kantorovich minimization problem (**K**).

Kantorovich[7] considered the problem of minimizing the functional

$$K(m) = \int |x - y|^2 m(x, y)\mathrm{d}x\mathrm{d}y. \tag{56}$$

The minimization is over all densities m such that their marginal density is given by I_1 and I_2, respectively, i.e.,

$$\int m(x, y)\mathrm{d}x = I_2(y), \qquad \int m(x, y)\mathrm{d}y = I_1(x). \tag{57}$$

Recall that both x and y denote points in the plane. It is known[7] that the problem of minimizing K under Eqs. (57) has a unique solution. Moreover, the minimizer m is supported exactly on the graph of the optimal mapping $\bar{T}$ defined above, namely, $\bar{m}(x, y) = \delta[x - \bar{T}(x)]$, where $\bar{m}$ is the optimal Kantorovich density. The Kantorovich problem provides a key tool in the theoretical analysis of the Monge problem. We use it here as a potential numerical tool.

To solve the Kantorovich problem, we need to discretize the densities I_1, I_2. That is, we approximate the intensities I_1, I_2 with discrete distributions:

$$I_1 \approx \sum_1^N m_i^{(1)}\delta_{x_i}; \qquad I_2 \approx \sum_1^N m_i^{(2)}\delta_{y_i}, \tag{58}$$

where $\{\delta_{x_i}, \delta_{y_j}\}$ are unit point masses and $\Sigma\, m_i^{(1)} = \Sigma\, m_i^{(2)} = 1$. The corresponding Kantorovich problems (57) takes the form

$$\min_M \sum_{i=1}^N \sum_{j=1}^N M_{i,j}|x_i - y_j|^2, \tag{59}$$

where the minimum is taken on all nonnegative, $N \times N$ matrices M, which satisfy $\Sigma_i M_{i,j} = m_j^{(2)}$, $\Sigma_j M_{i,j} = m_i^{(1)}$.

It turns out, however, that the naive discretizations [(58) and (59)] leads to a minimizer that is not supported on a graph. Therefore, instead of expressions (58), we select the points $\{x_i\}$, $\{y_i\}$ according to empirical distributions, namely,

$$I_1 \approx \frac{1}{N}\sum_i^N \delta_{x_i}, \qquad I_2 \approx \frac{1}{N}\sum_i^N \delta_{y_i}. \tag{60}$$

We then consider the minimization problem (59) subject to the condition $\Sigma_i M_{i,j} = \Sigma_j M_{j,i} = 1/N$. For this formulation there exists a unique minimizer that is a permutation matrix $M_{i,j} = (1/N)\delta_{i,j(i)}$. This permutation defines a discrete mapping

$$\Pi: \quad \{1,...N\} \to \{1,...N\},$$

which is the solution of the discrete optimal mapping via

$$T(x_i) = y_{\Pi(i)}.$$

This algorithm requires a selection of points [expressions (60)] that carry identical mass with respect to the intensities I_1 and I_2. This can be done in a variety of ways. For instance, Ref. 13 proposes a simple suitable sampling method for the case in which D_1 and D_2 are rectangles.

The main drawback of the linear programming formulation, based on the Kantorovich problem, is that it requires us to work with functions of four variables. This implies a need for large computer memory. We therefore consider an alternative numerical approach based on a continuous flow. The idea is that we need not use a flow that is faithful to the optical equations (8) and (9). It is only required that the flow will start from I_1 and will end in I_2. A natural candidate for such a flow is the steepest-descent method. Here we proceed along the gradient of the Monge problem while always preserving the constraint that the mapping transports I_1 into I_2.

A steepest-descent flow for the problem (**Mpq**) was computed by Angenent *et al.*[13] To emphasize that this flow of transformations is not directly related to the optical problem (only the terminal point of the flow, i.e., the optimal mapping $\bar{T}$ is), we introduce the flow as

$$U = U(x, t): R^2 \times [0, \infty) \to R^2$$

and the variable t to denote time for the flow. The flow starts with an initial mapping $U_0(x) = U(x, 0)$ that transports I_1 to I_2. For example, the sampling, [expressions (60)] with trivial permutation can be used to generate an initial mapping.

In the algorithm of Angenent *et al.*, one decomposes U at each time t into the orthogonal sum of a gradient and a divergence-free vector field:

$$U(x, t) = \nabla p(x, t) + v(x, t), \qquad \nabla \cdot v = 0. \quad (61)$$

When dealing with bounded domains, we must supplement the requirement that v is divergence free with the boundary behavior of v. The natural boundary condition is that the vector field v has no normal component at the boundary. Together, the requirements on v imply that p must solve the following Poisson–Neumann problem:

$$\Delta p = \nabla \cdot U, \qquad x \in D_1, \qquad \frac{\partial p}{\partial \nu} = U \cdot \nu, \qquad x \in \partial D_1, \quad (62)$$

where ν is the outward normal to D_1. Equivalently, the potential p can be characterized as the function that minimizes the functional

$$F(p) = \int |U - \nabla p|^2 \mathrm{d}x. \quad (63)$$

The evolution of U is driven by

$$\frac{\partial U}{\partial t} + \frac{1}{I_1} v \cdot \nabla U = 0. \quad (64)$$

We refer to the evolution problems (61), (62), and (64) as the Angenent–Haker–Tannenbaum (AHT) flow.

The energy $\int |U(x) - x|^2 \mathrm{d}x$ monotonically decreases along the AHT flow. If no singularities develop, the flow is expected to converge to the unique minimizer $\bar{T}$ of the Monge problem. The discretization of the flow (64) and the projections (62) must be done with great care. Recall that the AHT flow reduces the energy, but it is constrained to lie in the manifold of maps U that transport I_1 to I_2. If we are not careful in designing our numerical scheme, the constraint will not be exactly enforced, and then the flow will carry the mapping into the identity mapping, which corresponds to zero energy. Therefore we must use a numerical scheme that maintains the constraint also on the discrete level. We found that this can be achieved by treating the gradient and divergence operators as conjugate operators; numerically, we write a forward-difference scheme for the gradient and a backward-difference scheme for the divergence.

It is interesting to note the similarity between the AHT flow and the associated projection (61) and the Chorin projection method in fluid mechanics. Chorin[14] pointed out that the Navier–Stokes equations for incompressible fluids contain no equation for the pressure. Rather, the pressure is determined from the constraint that the flow must be incompressible. He thus devised a numerical approach in which one advances the fluid velocity numerically while imposing the constraint at each time iteration through a projection such as Eq. (61).

The AHT evolution equation could be problematic when I_1 is small. As we pointed out earlier, this might indeed happen in realistic optical situations. One possibility to overcome this difficulty is to replace the AHT flow with an alternative gradient flow. In the new formulation we propose, we replace the projection (61) with the orthogonal projection

$$U(x, t) = \nabla q(x, t) + w(x, t), \qquad \nabla \cdot (I_1 w) = 0. \quad (65)$$

We then replace the gradient flow (64) with the flow

$$\frac{\partial U}{\partial t} + w \cdot \nabla U = 0. \quad (66)$$

It can be shown that with the AHT flow the energy decreases according to

$$\frac{\mathrm{d}}{\mathrm{d}t} M(I_1, I_2) = -\int |v|^2 \mathrm{d}x \quad (67)$$

while along the flow (66) the energy satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} M(I_1, I_2) = -\int I_1 |w|^2 \mathrm{d}x. \quad (68)$$

Indeed, along the flow (66), points where I_1 is small will not contribute much to the decrease of the energy toward its minimal value; these areas, however, are not significant, since they carry little energy.

REFERENCES

1. M. R. Teague, “Deterministic phase retieval: a Green’s function solution,” J. Opt. Soc. Am. **73**, 1434–1441 (1983).
2. F. Roddier, “Curvature sensing and compensation: a new concept in adaptive optics,” Appl. Opt. **27**, 1223–1225 (1988).
3. T. E. Gureyev and K. A. Nugent, “Phase retrieval with the transport-of-intensity equation. II. Orthogonal series solution for nonuniform illumination,” J. Opt. Soc. Am. A **13**, 1670–1682 (1996).
4. C. M. Lee and J. Rubinstein (manuscript in preparation; available from the author at the address on the title page).
5. M. A. van Dam and R. G. Lane, “Wave-front sensing from defocused images by use of wave-front slopes,” Appl. Opt. **41**, 5497–5502 (2002).
6. J. B. Keller and R. M. Lewis, “Asymptotic methods for partial differential equations: the reduced wave equation and Maxwell’s equations,” in *Surveys in Applied Mathematics*, J. B. Keller and G. C. Papanicolaou, eds. (Plenum, New York, 1993), Vol. 1, pp. 1–82.
7. C. Villani, *Topics in Optimal Transportation* (American Mathematical Society, Providence, R.I., 2003).
8. G. Wolansky, “Optimal transportation in the presence of a prescribed pressure field,” preprint, available from the author at the address on the title page.
9. J. D. Benamou and Y. Brenier, “A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem,” Numer. Math. **84**, 375–393 (2000).
10. J. Rubinstein and G. Wolansky, “A weighted least action principle for dispersive waves,” preprint.
11. J. F. Nye, *Natural Focusing and Fine Structure of Light* (Institute of Physics, Bristol, UK, 1999).
12. Y. Brenier, “Polar factorization and monotone rearrangement of vector-valued functions,” Commun. Pure Appl. Math. **64**, 375–417 (1991).
13. S. Angenent, S. Haker, and A. Tannenbaum, “Minimizing flows for the Monge–Kantorovich problem,” SIAM J. Math. Anal. **35**, 61–97 (2003).
14. A. J. Chorin, “Numerical solution of the Navier–Stokes equations,” Math. Comput. **22**, 742–762 (1968).