RESEARCH

Open Access

CrossMark

Ray mappings and the weighted least action principle

Jacob Rubinstein[1][*][†] , Gershon Wolansky[1][†] and Yifat Weinberg[1]

*Correspondence:
koby@technion.ac.il
[1]Department of Mathematics,
Technion, Haifa, Israel
†Equal contributors

Abstract

Two basic problems in optics are presented. The solutions to both problems are formulated in terms of the associated ray mappings. An alternative formulation based on a weighted sum of the actions along the rays is derived. Existence of solutions is established via the Weighted Least Action Principle. Numerical methods for computing the ray mappings are discussed. Finally, we demonstrate the theoretical considerations by presenting complete solutions to a phase retrieval problem and to a specific beam shaping lens design.

Keywords: Ray mapping; Geometrical optics; Monge mass transport; Diffraction integral; Phase retrieval; Least action

1 Introduction

Ray mappings are fundamental objects in geometrical optics. Given an optical element, its Hamilton's point eikonal function includes the information on *all* possible ray mappings induced by it. However, in many applications one looks for specific ray mappings that satisfy certain constraints. In this paper we shall consider two such cases. In the first case we search for all ray mappings associated with a single monochromatic beam whose intensity is known at two planes. This question, known as the phase from intensity problem, arises in many applications ranging from astronomy [18, 24] to ophthalmology [12]. The second case concerns the shaping of a collimated beam with arbitrary incident and refracted intensity distributions. Beam shaping has many applications ranging from solar energy to chip manufacturing and medical instruments [6].

While these two cases involve two completely different problems in optics, we shall show that their *mathematical* formulation is essentially identical. We shall then introduce two characterizations of these problems, provide practical methods to compute the ray mappings, and from the mappings obtain the unknown phase in case one, and the beam shaping lens in case two.

2 Methods

In the next section we formulate the wave propagation problem via the Rayleigh-Sommerfeld diffraction integral. This formulation is natural for problems where the wave might pass through a caustic. The large wave number expansion of the diffraction integral is used to derive the ray mapping equations for the phase retrieval problem. The ray mapping condition for the beam shaping lens, that was derived in an earlier work by us, is

presented in Sect. 4. In Sect. 5 we show that both ray mapping problems can be solved by the Weighted Least Action Principle (WLAP). This variational method, which is a natural extension of the Fermat principle of least time [19, 20], was extensively studied in recent years both theoretically [25] and in a number of applications. In particular we characterize important properties of the critical points of the weighted least action functional, and associate them with the phase retrieval and beam shaping problems. Numerical methods for computing the ray mappings, or equivalently for finding the critical points of the Weighted Least Action functional, along with one example of phase retrieval and one example of lens design are provided in Sect. 6.

3 Phase from intensity

We denote $u(x, y, z)$ a monochromatic wave propagating in the positive z direction. We express u in the form

$$u(x, y, z) = A(x, y, z)e^{ik\phi(x,y,z)}. \tag{1}$$

The goal is to find the phase (or eikonal) ϕ from measurements of its intensities $I = A^2$ at two planes, say $z = 0$ and $z = h$:

$$I(x, y) := I(x, y, 0), \qquad I_p(x, y) := I(x, y, h). \tag{2}$$

The wave u is assumed to obey the Helmholtz equation $\Delta u + k^2 u = 0$, where k is the wave number, and we assume that the refraction index is $n = 1$.

In the geometrical optics limit the phase is equivalent to the rays, which are the normals to the wavefronts. A point (x, y) at the aperture Σ on the plane $z = 0$ is mapped into a point $T(x, y) = (x_p, y_p)$ at the imaged aperture Σ_p on the second plane $z = h$. To find the relation between the ray mapping and the phase $\phi(x, y, 0)$ we use the Rayleigh-Sommerfeld diffraction integral [3], [23]:

$$u(x_p, y_p, h) = \frac{-ik}{2\pi} \int_\Sigma A(x, y, 0) \frac{\exp(ik(\phi(x, y, 0) + d(x_p, y_p, x, y)))}{d(x_p, y_p, x, y)} \cos\theta \, dx \, dy. \tag{3}$$

Here $d = ((x_p - x)^2 + (y_p - y)^2 + h^2)^{1/2}$, and θ is the angle made by the line connecting (x, y) and (x_p, y_p) with the plane $z = 0$. The integral is approximated in the large k limit by the stationary phase method [10], [22]. The stationarity condition is $\nabla_x(\phi(x, y, 0) + d(x_p, y_p, x, y)) = 0$, or equivalently:

$$\nabla\phi = -\nabla_x d = \frac{(x_p - x, y_p - y)}{((x_p - x)^2 + (y_p - y)^2 + h^2)^{1/2}}, \tag{4}$$

where the notation $\nabla_x s(x_p, y_p, x, y)$ denotes the gradient of a function s with respect to the (x, y) variables. We assume that equation (4) implies a one to one correspondence between Σ and Σ_p, which is equivalent to assuming that no caustic surface intersects either planes $z = 0, h$.

Using equation (4), we can write down the classical stationary phase approximation of $u(x_p, y_p)$ and then use it to obtain the following relation between the intensities at the two

planes:

$$I_p\big(T(x,y)\big)\big|J\big(T(x,y)\big)\big| = I(x,y). \tag{5}$$

Here J is the Jacobian of the ray mapping $(x_p, y_p) = T(x,y)$, and $|J|$ denotes the absolute value of the determinant of J. We denote the set of mappings T satisfying condition (5) as $\mathcal{T}$. Therefore, the phase retrieval problem is equivalent to finding a ray mapping T : $(x,y) \to (x_p, y_p)$ that satisfies both equations (4) and (5).

4 Beam shaping

Consider an incident collimated beam, propagating in the z direction, whose cross sectional intensity is $I(x,y)$. The goal is to design a lens with two freeform surfaces, denoted f and g, that converts it into a new collimated beam with cross sectional intensity $I_p(x,y)$. Such problems are of great importance in the laser industry [6]. The case where I and I_p are both radially symmetric is well-known [15], and can be reduced to simple ODEs. However, we consider the most general case of arbitrary distributions I and I_p. In this case beam shaping can only be achieved by freeform surfaces [7], [21], [13], [14], [16], [4].

The schematic beams and optical element are depicted in Fig. 1. In Fig. 2 we present an example, where the target is to design a lens that converts an incoming collimated beam with six spots (for instance a beam created by a LED-based device) into a homogeneous collimated beam. Later, in Sect. 6, we shall compute the surfaces f and g for this example. A striking fact, proved in Sect. 5, is that both these surfaces are convex.

Let $f(x,y)$ and $g(x,y)$ be the front and back surfaces of the lens, respectively. The refraction index of air is 1, and we denote the lens refraction index n. Consider a ray that starts from a point (x,y) on a surface $z = 0$ on the left side of the lens. After two refractions, the ray is mapped into a point $(x_p, y_p) = T(x,y)$ on a plane $z = h$ on the right side of the lens. The ray mapping T is constrained by the energy conservation equation (5). In addition, the mapping T and the lens surfaces f and g are related by Snell's law of refraction and by the fact that both planes $z = 0, h$ are wavefronts, since the incident and refracted beams are collimated. This implies that the optical path length between all points (x,y) and their
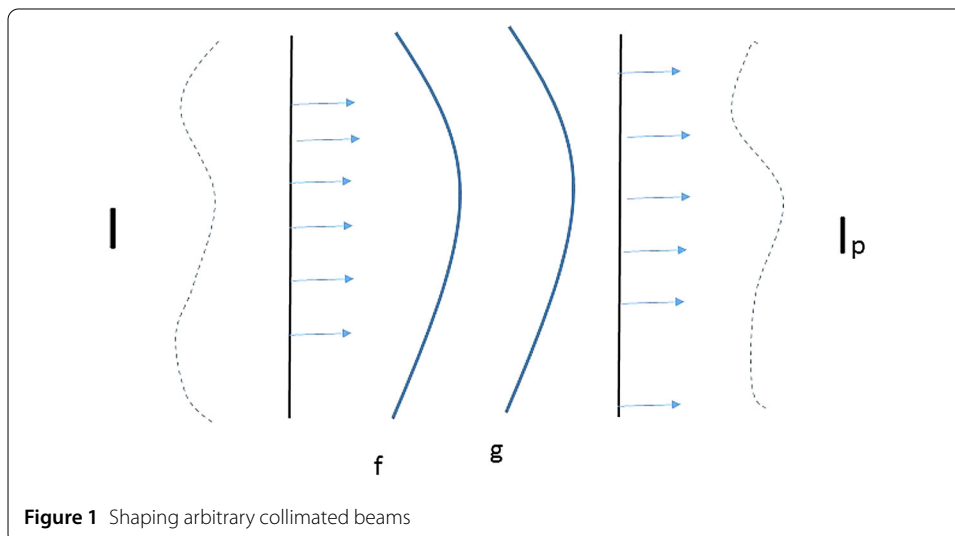

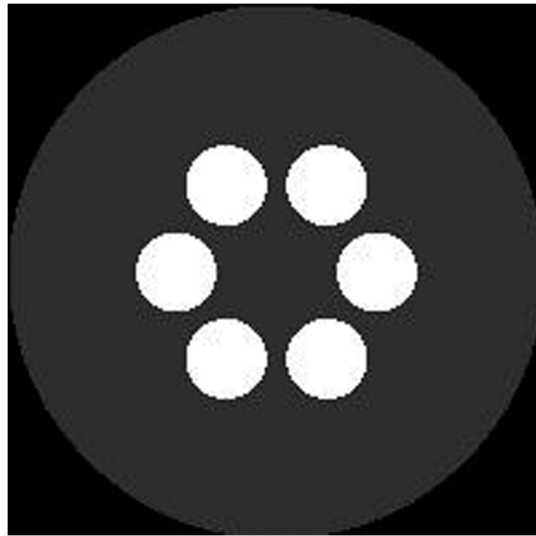
Figure 1 Shaping arbitrary collimated beams

Figure 2 An incoming beam consisting of six spots arranged over a circle

images $T(x,y)$ is constant. Combining these facts, and after some algebra, the following equation is obtained [21] that relates f and T.

$$\nabla f = \frac{-n((x_p, y_p) - (x, y))}{\sqrt{\chi - bn|(x_p, y_p) - (x, y)|^2}},$$

$$(6)$$

where χ and b are two constants that depend upon n, the distance h between the planes, and an integration factor [21]. A similar equation holds for the second surface g.

5 The weighted least action principle

We presented in Sects. 3 and 4 two different optical problems that are fully determined by specific ray mappings. Surprisingly, both problems lead to a similar mathematical formulation. Both ray mappings are required to satisfy the energy conservation condition (5). In addition, in both cases the ray mapping T must satisfy a 'symmetry' condition such as (4) or (6). Three questions arise. The first is whether such ray mappings T exist at all. If the answer to this question is positive, the second question is how many solutions exist, and what are their properties. The third question is how to *compute* the ray mappings.

It is very hard to answer these questions directly from equations (4)–(5). Together they constitute a very complicated PDE of the notorious Monge–Ampère type [17]. Instead, we shall answer all three questions using an equivalent formulation of the problem. For this purpose, we associate each ray connecting (x, y) and $(x_p, y_p) = T(x, y)$ with an *action* $C(T, (x, y))$, and then define the *weighted* total action to be

$$M(T) = \int_\Sigma C(T, (x, y)) I(x, y) \, dx \, dy,$$

$$(7)$$

where Σ denotes the wave's aperture, or the beam's cross sectional domain. The weighted action M is restricted to the set $\mathcal{T}(I, I_p)$ of mappings that satisfy condition (5). The functional $M(T)$ was introduced (for the special case $C(x_p, y_p, x, y) = |(x_p, y_p) - (x, y)|$) by Monge.

It has been extensively studied in recent years both theoretically and in the context of several applications [25]. In the optical set up, the special case $C(x_p, y_p, x, y) = |(x_p, y_p) - (x, y)|^2$, naturally denoted the L_2 Monge problem, is related to paraxial wave propagation [22], and to the Schrödinger equation [20]. For the applications in the present paper we consider cost functions C of the form

$$C(x_p, y_p, x, y) = C\big((x_p, y_p) - (x, y)\big), \tag{8}$$

with $C(z)$ being a smooth convex function.

The stationary points of $M(T)$ are characterized by the following condition:

Theorem 1 *Assume I, I_p are positive continuous functions, and the domains Σ, Σ_p are compact. A ray mapping $(x_p, y_p) = \bar{T}(x, y)$ is a critical point of M in the class $\mathcal{T}(I, I_p)$ if and only if T satisfies the following relation*

$$\nabla_x C(x_p, y_p, x, y) = \nabla \zeta(x, y), \tag{9}$$

for some 'potential' function $\zeta(x, y)$.

The stationarity condition (9) can be proved (e.g. [22]) by computing the first variation of M under the constraint (5). The next question is how many solutions exist.

Theorem 2 *Under the assumptions of Theorem 1 there exist a minimizer and a maximizer for the functional $M(T)$.*

The existence of a minimizer of M, was proved by Brenier [5]; see also [25]. The existence of a maximizer follows from the fact that the conditions on I, I_p and Σ, Σ_p imply weak star compactness for the optimization problem; see [26]. Furthermore, we conjecture that there always exist also critical points that are neither minimizers nor maximizers.

As will be shown later, it is important to characterize in some detail the stationary solutions. In the L_2 Monge problem, it can be shown that both the minimizer and the maximizer are gradients of some function ψ, where ψ is convex for the minimizer, and concave for the maximizer. For a more general convex cost function C, we now prove:

Theorem 3 *Assume I and I_p are positive continuous functions, the two domains Σ, Σ_p are bounded, and $C(z)$ is a smooth convex function. Let $\zeta(x)$ be the potential in equation (9) associated with the maximizer of M. Then the maximizer $(x_p, y_p) = \bar{T}(x, y)$ of the functional M is of the form*

$$\bar{T}(x, y) = (x, y) - \nabla C^*\big(\nabla \zeta(x, y)\big), \tag{10}$$

where

$$C^*(p_1, p_2) = \max_{x, y}(p_1, p_2) \cdot (x, y) - C(x, y), \tag{11}$$

and ζ is a convex function.

Proof Define the class $\mathcal{A}$ to be:

$$\mathcal{A} = \alpha(x,y), \beta(x_p,y_p); \qquad \alpha(x,y) + \beta(x_p,y_p) \geq C\big((x_p,y_p) - (x,y)\big). \tag{12}$$

We also introduce the class of joint intensity functions Π, such that the marginal intensities of any function $\pi \in \Pi$ are I and I_p respectively, i.e.

$$\int \pi(x,y,x_p,y_p)\,dx\,dy = I_p(x_p,y_p), \qquad \int \pi(x,y,x_p,y_p)\,dx_p\,dy_p = I(x,y). \tag{13}$$

Clearly, for every $\pi \in \Pi$

$$\int \big(\alpha(x,y) + \beta(x_p,y_p)\big)\pi\,dx\,dy\,dx_p\,dy_p$$

$$= \int \alpha(x,y)I(x,y)\,dx\,dy + \int \beta(x_p,y_p)I_p(x_p,y_p)\,dx_p\,dy_p$$

$$\geq \int C\big((x_p,y_p) - (x,y)\big)\pi\,dx\,dy\,dx_p\,dy_p. \tag{14}$$

Since the inequality in equation (14) holds for all π and all α, β we can apply a min-max argument and conclude

$$\inf_{\alpha,\beta\in\mathcal{A}} \int \alpha(x,y)I\,dx\,dy + \int \beta(x_p,y_p)I_p\,dx_p\,dy_p$$

$$\geq \sup_{\pi\in\Pi} \int C\big((x_p,y_p) - (x,y)\big)\pi\,dx\,dy\,dx_p\,dy_p. \tag{15}$$

In particular, for the special set of π of the form $\pi = I(x,y)\delta((x_p,y_p) - T(x,y))$, where $T(x,y)$ is a mapping in $\mathcal{T}$, we have

$$\inf_{\alpha,\beta\in\mathcal{A}} \int \alpha(x,y)I\,dx\,dy + \int \beta(x_p,y_p)I_p\,dx_p\,dy_p$$

$$\geq \sup_{T\in\mathcal{T}} \int C\big(T(x,y) - (x,y)\big)I(x,y)\,dx\,dy. \tag{16}$$

Recalling the potential function $\zeta(x,y)$ we define

$$\eta(x_p,y_p) = \max_{x,y} C\big((x_p,y_p) - (x,y)\big) - \zeta(x,y). \tag{17}$$

Moreover, solving the optimization problem in equation (17) provides a ray mapping

$$(x_p,y_p) = \bar{T}(x,y) = (x,y) - \nabla_x C^*(\nabla_x \zeta). \tag{18}$$

Similarly, ζ can be expressed in terms of η:

$$\zeta(x,y) = \sup_{x_p,y_p} C\big((x_p,y_p) - (x,y)\big) - \eta(x_p,y_p). \tag{19}$$

We now select $\bar{\pi} = I(x,y)\delta((x_p,y_p) - \bar{T}(x,y))$ and $\alpha(x,y) = \zeta(x,y)$, $\beta(x_p,y_p) = \eta(x_p,y_p)$. Using the definition of ζ, η and $\bar{T}$ we can calculate the integrals in the inequality (15), and

obtain that for this choice of $(\bar{T}, \alpha, \beta)$ both sides equal $M(\bar{T})$. Therefore $\bar{T}$ is the mapping that maximizes $M(T)$ in the class $\mathcal{T}$. Finally, since $\zeta(x, y)$ is obtained by maximizing a convex function C, it must be convex, which completes the proof. $\qquad\square$

It is now possible to identify both equations (4) and (6) to be of the form of equation (9). Therefore the phase from intensity problem can be solved by associating the unknown phase $\phi(x, y, 0)$ with the critical points of M for the choice $C_1(x_p, y_p, x, y) = d(x_p, y_p, x, y)$. Similarly the beam shaping lens can be solved by identifying the front lens surface f with the critical points of M for the choice $C_2(x_p, y_p, x, y) = -\sqrt{\chi - bn|(x_p, y_p) - (x, y)|^2}$. It remains to consider the computation of the critical points.

We point out that in general the selection of the appropriate cost function C follows from the nature of the wave equation, or more precisely from its Hamiltonian nature. A number of canonical wave equations, their Hamiltonian, and the associated C can be found for instance in [20].

6 Results

The problem of computing minimizers of M received quite a bit of attention. For example we refer to [1, 2, 8] where PDE-based approach is suggested. An alternative combinatorial approach is to sample the two intensities I, I_p faithfully by two sets of points (x^j, y^j) for I and (x_p^j, y_p^j) for I_p, where $j = 1, 2, \ldots, N$. The problem of minimizing M is the same as computing the permutation π that minimizes

$$\sum_{j=1}^{N} C\left(x_p^{\pi(j)}, y_p^{\pi(j)}; x^j, y^j\right).$$

The same method can be used to find the maximizer of $M(T)$. While searching over all permutations has of course huge complexity, much more efficient algorithms, such as the "Hungarian algorithm" [9] were proposed with complexity of $O(N^3)$. An even more effective multiscale algorithm was proposed recently by Merigot [11], although it is limited to quadratic cost functions. We recently developed a very fast $O(N \log N)$ multiscale algorithm that works for all convex C, at the price of some moderate assumptions on the potential function ζ. However, in the simulations in the present paper we use the Hungarian algorithm.

We proceed to present a few examples. We first demonstrate the WLAP as a method to retrieve an arbitrary phase. We select the initial aperture to be the unit disc, and the initial phase is $\phi(x, y) = 0.1(\frac{x^2}{2} + y^2 + 3x^4 + y^3)$ to mimic defocus, coma and spherical aberrations. We assumed uniform intensity on the first screen, and computed the exact intensity at the second screen. The sampling of both intensities (200 sampling points) are depicted in Figs. 3(a) and 3(b), respectively.

The ray mapping was computed by minimizing the Monge functional C_1 defined in Sect. 5. The calculated ray mapping, and the exact ray mapping are depicted in Figs. 4 and 5, respectively.

The expected convergence rate of the algorithm can be bounded by the size of an average 'cell'. In the case of a unit disc aperture, the bound is thus $\sqrt{\pi / N}$ where N is the number of sampling point. In practice the convergence is better, as can seen in Table 1.

In the second example we consider the beam shaping problem of Sect. 4. The intensities were sampled with 400 points (right drawing of Fig. 6. In this case we computed the
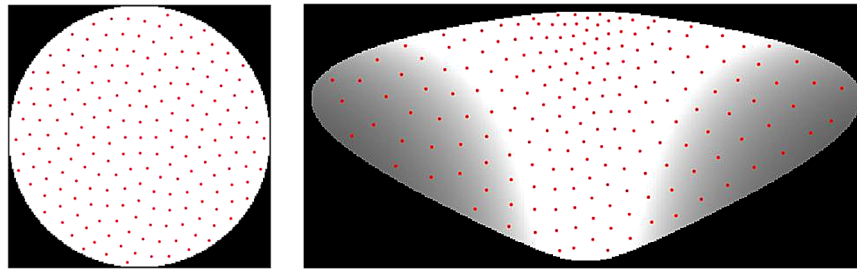
(a) The sampling of the wave's
intensity on the first screen

(b) The sampling of the wave's intensity on the second screen

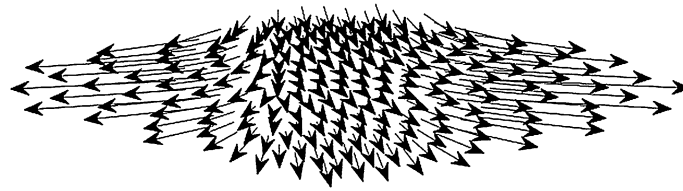**Figure 3** The sampling of the wave's intensity

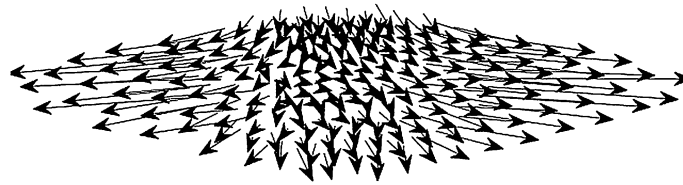**Figure 4** The theoretical ray mapping

Figure 5 The ray mapping obtained by minimizing the Monge functional

Table 1 Convergence of the ray mapping algorithm

Number of sampling points	Average mapping error	Theoretical error bound
200	0.038	0.125
400	0.021	0.088
800	0.017	0.062
1000	0.015	0.056

maximizer of M, for the cost function C_2 defined in Sect. 5, since this solution guarantees that both surfaces f and g are convex, which is a very important consideration in the lens manufacturing. The ray mapping is depicted in the left drawing of Fig. 6. Finally, we depict the lens surfaces in Fig. 7.

Indeed, as proved in Theorem 3 in Sect. 5, both surfaces are convex, in contrast to the highly nonconvex intensity distribution of the incident beam.

7 Discussion and conclusions

The notion of ray mapping was presented in the context of two canonical optical problems. Although the problems are physically completely different, it was shown that they both
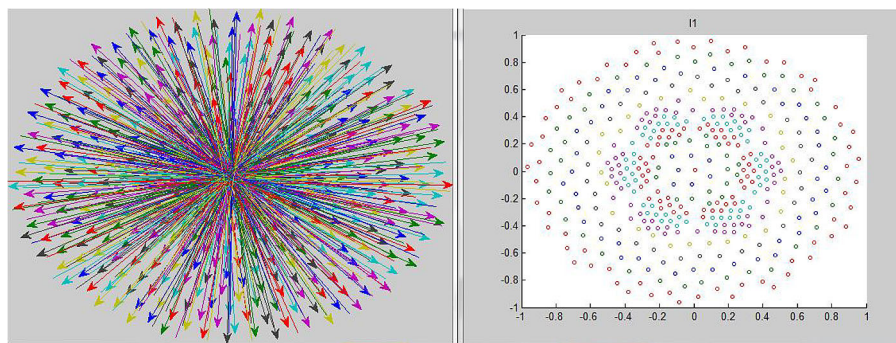
Figure 6 A discrete sampling of the intensity of the incoming beam (right) and the ray mapping (left)
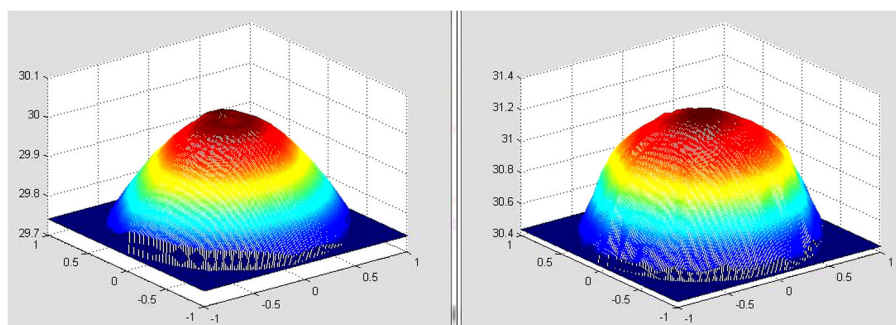

Figure 7 The surfaces (*f* on the left and *g* on the right) of the beam shaping lens

can be solved within the same mathematical framework of the Monge optimal transport paradigm. We showed that the weighted least action functional has at least two critical points, one a minimizer and one a maximizer. We conjecture that there exist also other (finitely many) critical points.

The characterization of the optical problems as minimizers or maximizers of the functional $M(T)$ provides a very useful tool for computing the solutions of these problems. The minimizer (or maximizer) can be found by a relatively simple combinatorial optimization tools. We thus provided an example demonstrating the application of the theory to phase retrieval and an example of beam shaping. An important aspect of the beam shaping problem is that by *maximizing* the functional $M(T)$, we obtain a lens with two convex freeform surfaces. The convexity was proved in a very general context under certain convexity assumption on the action C. This assumption holds for the Helmholtz equation or Fresnel equation for monochromatic waves, for the Schrödinger equation (in the semi-classical limit) and in other cases.

Abbreviations
WLAP, Weighted Least Action Principle.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Angenent S, Tannenbaum A, Haker S. Minimizing flows for the Monge–Kantorovich problem. SIAM J Math Anal. 2003;35:61–97.
2. Benamou JD, Brenier Y. Minimizing flows for the Monge–Kantorovich problem. Numer Math. 2000;84:375–93.
3. Born M, Wolf E. Principles of optics. Cambridge: Cambridge University Press; 1999.
4. Bosel C, Gross H. Single freeform surface design for prescribed input wavefront and traget irradiance. J Opt Soc Am A. 2017;34:1490–9.
5. Brenier Y. Polar factorization and monotone rearrangement of vector-valued functions. Commun Pure Appl Math. 1991;64:375–417.
6. Dickey FM. Laser beam shaping. Opt Photonics News. 2003;14:31–5.
7. Glimm T, Oliker VI. Optical design of two-reflector systems, the Monge–Kantorovich mass transfer problem and Fermat's principle. Indiana Univ Math J. 2004;53:1255–87.
8. Haber E, Rehman T, Tannenbaum A. An efficient numerical method for the solution of the L_2 optimal mass transfer problem. SIAM J Sci Comput. 2010;32:197–211.
9. Kuhn HW. The Hungarian method for the assignment problem. Nav Res Logist Q. 1955;2:83–97.
10. Mendel L, Wolf E. Optical coherence and quantum optics. Cambridge: Cambridge University Press; 1995.
11. Merigot Q. Multiscale approach to optimal transport. Comput Graph. 2011;30:1548–92.
12. Nam J, Rubinstein J, Thibos L. Wavelength adjustment using an eye model from aberrometry data. J Opt Soc Am A. 2010;27:1561–74.
13. Oliker V. On design of freeform refractive beam shapers, sensitivity to figur eerror and convexity of lenses. J Opt Soc Am A. 2008;25:3067–76.
14. Oliker V. Designing freeform lenses for intensity and phase control of coherent light with help from geometry and mass transport. Arch Ration Mech Anal. 2011;201:1013–45.
15. Oliker V, Rubinstein J, Wolansky G. Ray mappings and illumination control. J Photonics Energy. 2013;3:035599.
16. Oliker V, Rubinstein J, Wolansky G. Supporting quadric method in optical design: precise illumination control of a collimated light with a single convex freeform lens. Adv Appl Math. 2015;62:160–83.
17. Prins CR, Ten Thije Boonkkamp JHM, Van Roosmalen J, Ijerman WL, Tukker TM. A Monge–Ampere-solver for free-form reflector design. SIAM J Sci Comput. 2014;36:640–60.
18. Roddier F. Curvature sensing and compensation: a new concept in adaptive optics. Appl Opt. 1988;27:1223–5.
19. Rubinstein J, Wolansky G. A variational principle in optics. J Opt Soc Am A. 2004;21:2164–72.
20. Rubinstein J, Wolansky G. A weighted least action principle for dispersive waves. Ann Phys. 2005;316:271–84.
21. Rubinstein J, Wolansky G. Intensity control with a free-form lens. J Opt Soc Am A. 2007;24:463–9.
22. Rubinstein J, Wolansky G. Optimal transport and geometrical optics. J Opt Soc Am A. 2017;34:1817–23.
23. Smith DG. Field guide to physical optics. Bellingham: SPIE; 2013.
24. Teague MR. Deterministic phase retrieval: a Green's function solution. J Opt Soc Am. 1983;73:1434–41.
25. Villani C. Topics in optimal transportation. Providence: Am. Math. Soc.; 2003.
26. Wells BB. Weak compactness of measures. Proc Am Math Soc. 1969;20:124–30.