Designing Illumination Lenses and Mirrors by the Numerical Solution of Monge-Ampère Equations

Kolja Brix, Yasemin Hafizogullari and Andreas Platen

Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany
DESIGNING ILLUMINATION LENSES AND MIRRORS BY THE NUMERICAL SOLUTION OF MONGE–AMPÈRE EQUATIONS

KOLJA BRIX§, YASEMIN HAFIZOGULLARI§, AND ANDREAS PLATEN§

Abstract. We consider the inverse refractor and the inverse reflector problem. The task is to design a free-form lens or a free-form mirror that, when illuminated by a point light source, produces a given illumination pattern on a target. Both problems can be modeled by strongly nonlinear second-order partial differential equations of Monge–Ampère type. In [Math. Models Methods Appl. Sci. 25 (2015), pp. 803–837, DOI: 10.1142/S0218202515500190] the authors have proposed a B-spline collocation method which has been applied to the inverse reflector problem. Now this approach is extended to the inverse refractor problem. We explain in depth the collocation method and how to handle boundary conditions and constraints. The paper concludes with numerical results of refracting and reflecting optical surfaces and their verification via ray tracing.

Key words. Inverse refractor problem, inverse reflector problem, elliptic Monge–Ampère equation, B-spline collocation method, Picard-type iteration

AMS subject classifications. 35J66, 35J96, 35Q60, 65N21, 65N35


1. Introduction. Both problems, the inverse refractor and the inverse reflector problem, from illumination optics can be formulated in the following framework: Let a point-shaped light source and a target area be given, e.g. a wall. Then we would like to construct an apparatus that projects a prescribed illumination pattern, e.g. an image or a logo, onto the target. Since we aim for maximizing the efficiency, we would like to construct our optical device in such a way that, neglecting losses, it redirects all light emitted by the light source to the target. We focus our attention to the design of such an optical system in the simple case that it either consists of a single free-form lens or of a single free-form mirror, see Figure 1.1 for an illustration of the former case. Our goal is now to compute the shape of the optically active surfaces, modeled as free-form surfaces, such that the desired light intensity distribution is generated on the target. Since these problems from illumination optics from the mathematical point of view conceptually fall into the class of inverse problems, they are also called inverse reflector problem and inverse refractor problem, respectively. In particular, since the size of the optical system is comparable to that of the projected image, we address the case of the near field problems.

There is a variety of technical applications of such optical systems, e.g. spotlights with prescribed illumination patterns used in street lamps or car headlamps, see e.g. [1] [6] [36].

The authors present in [5] a solution method for the inverse reflector problem via numerically solving a strongly nonlinear second-order partial differential equation (PDE) of Monge–Ampère type. Due to the high potential of this approach we now extend this method to the case of illumination lenses.

This paper is organized as follows: Since the reflector problem has been discussed in detail in [5] we mainly focus on the refractor problem. We start with the state of...
Figure 1.1: Setting of the refractor problem. The index of refraction of the lens material is $n_1$, while the surrounding has the refractive index $n_2$.

the art for its solution in Section 2. Then, we formulate the problem via a partial differential equation of Monge–Ampère type which we discuss in Section 3 for the construction of a refractor. For completeness we also give the Monge–Ampère formulation for the reflector problem in Section 4. Next, the numerical method is explained in Section 5. Since this type of optical design problem raises many difficulties in the solution process we discuss in Section 6 how these can be resolved. Finally, in Section 7 we look at numerical results for the inverse reflector and refractor problems and end this paper in Section 8 with our conclusions.

2. State of the art. In this section we discuss the methods available for the solution of the inverse design problems in nonimaging optics, see the monographies by Chaves [8] and by Winston, Miñano and Benítez [39] for an introduction to nonimaging optics and the paper by Patow and Pueyo [30] for a survey article on inverse surface design form the graphics community’s point of view. For a detailed survey of solution techniques for the inverse reflector problem, we refer the reader to [5, Section 2].

Focusing on the inverse refractor problem, in the paper by Wester an Bäuerle [37] there is a list of approaches, a discussion on practical problems, e.g. extended sources and Fresnel losses, and examples with LED lighting, e.g. a lens for automotive fog light and a lens producing a logo.

In the rest of this section, we first give a short overview of other solution techniques in Section 2.1 and then focus on methods based on PDEs in Section 2.2, which is also our problem formulation of choice. Finally, we discuss some advanced topics in Section 2.3 and draw our conclusions in Section 2.4.

2.1. Approaches for the solution of inverse problems in nonimaging optics not based on a PDE. We distinguish three different groups of techniques for the solution of inverse problems in nonimaging optics, which are not based on a PDE: there are methods resorting from optimization techniques, others built from Cartesian ovals and a third group of methods which are geometrical constructions.

Optimization approaches. There are methods for the design of optical surfaces, which are based on optimization techniques, see e.g. [34, 40]. Starting from an initial guess, the outline of the iterative optimization process for the determination of the optical surfaces is as follows: First, the current approximation of the optical surfaces is validated by ray tracing. In a second step, using an objective function, which is
often closely related to the Euclidean norm, the resulting irradiance distribution is compared to the desired one and a correction of the optical surfaces is determined. The process ends, when a suitable quality criterion is fulfilled, otherwise these two steps are repeated.

The advantage of this method is that it is very flexible. However, optimization procedures are very costly because of the repeated application of the ray tracing and it is unclear if the iterative methods converge at all.

*Cartesian ovals methods.* Cartesian ovals are curves of fourth order in the plane. They can be associated with two foci such that light emitted at one focus is collected at the other focus. Here the Cartesian oval coincides with the interface of two optical media with different refractive indices. Cartesian ovals can be extended to surfaces in 3d with the same property. By combining clippings of several of these surfaces in an iterative procedure a new segmented surface can be constructed that approximates the solution. This strategy has first been developed by Kochengin and Oliker [20, 21] for the construction of solutions for the inverse reflector problem. Later this has been extended to the inverse refractor problem [25] using Cartesian ovals, see also [15, Section 2] for some theoretical background, and for a collimated light beam instead of a point light source [26] using hyperboloids.

Although this technique has the advantage to permit the construction of continuous but non-differentiable surfaces [25], the number of clippings \( K \) required grows linearly in the number of pixels in the image. For example, using ellipsoids of revolution for the construction of a mirror with accuracy \( \gamma > 0 \), the complexity of the method scales like \( O(\frac{K^4}{\gamma} \log \frac{K^2}{\gamma}) \), see [22], such that it quickly becomes infeasible for higher resolutions.

*Geometric construction methods.* Reflective and refractive free-form surfaces can also be designed by geometric approaches. Probably the most famous of these techniques is the simultaneous multiple surfaces (SMS) method extending the ideas of Cartesian-oval methods, see e.g. [39, Chapter 8] and [3, 24] and the references therein. The main idea of the SMS method is the simultaneous construction of two optical surfaces, e.g. both surfaces of a lens, which permits to couple two prescribed incoming wave fronts, e.g. coming from two point light sources, with two prescribed outgoing wave fronts. While in its 2D version, the method is used to design rotationally symmetric optical surfaces, in a 3D variant it is also capable to construct free-form optical surfaces. However, the authors could not find any hint on the computational costs in the literature but conjecture that this scheme is expensive especially for complex target illumination patterns.

In several publications for the inverse refractor problem a PDE is derived, whose solution models the desired optical free-form surfaces, see e.g. [36, 15, 16, 27, 41, 32, 33, 11]. In these approaches usually the low wavelength limit is assumed to hold, i.e. the problems are formulated using the geometrical optics approximation.

Some examples for the inverse refractor problem with a more complex target illumination pattern are shown in [36, 41, 32, 33]. However, in all four articles the descriptions and discussions of the numerical methods are incomplete. To the best of the authors’ knowledge the solution method is not fully documented in the literature.

While we consider the case of a point light source, an interesting and closely related problem is shaping the irradiance distribution of a collimated light beam, see e.g. [26, 27] for the theory including some results on existence and uniqueness of solutions.
We refer the reader to the monography by Gutiérrez [14] for a general overview of Monge–Ampère-type equations.

Since we are looking for an optical surface which redirects light coming from a source onto a target, one can model this problem in terms of optimal transportation. Optimal transport. There are also methods which are based on a problem of optimal transport which leads to Monge–Ampère-type equations, see e.g. [1, 6, 28]. First the ray mapping, i.e. the mapping of the incoming light rays onto the points at the target, is computed via an optimal transport approach. At this point the optical surface is still unknown but in a next step it is constructed from the knowledge of the target coordinates for each incoming light ray. In 1998 Parkyn [29] already described a very similar procedure.

2.3. Advanced topics. In the current formulation of the problem only one single idealized point light source has been used. An extension to multiple point light sources is discussed by Lin [23] where the optical refractors are determined from those calculated for single point light sources by a weighted least-squares approach. More techniques for the case of extended light sources can be found in the papers by Bortz and Shatz [4] and Wester et al. [38].

In particular for the refractor problem, some energy is lost for the illumination of the target because of internal reflections in the lens material. A theoretical discussion of these Fresnel losses can be found in the publications by Gutiérrez [15] Section 5.13] and Gutiérrez and Mawi [17]. In [1, 6] the losses are minimized by free-form shaping of both refractive surfaces of the lens.

2.4. Conclusion. Our approach is motivated by the fact that even for the special case of a single point light source and the computation of just one surface of the lens we could not find any fully detailed method in the literature which can produce complex illumination patterns on the target area. From the authors’ point of view, the most promising approach is the one by solving a PDE of Monge–Ampère type.

3. The inverse refractor problem. This section is devoted to the formulation of the Monge–Ampère equation that models the near field refractor problem as given in the paper by Gutiérrez and Huang [16]. Since the full theory is a bit involved, we restrict ourselves to a summary of the most important aspects and refer the reader to [16, Appendix A] and the paper by Karakhanyan and Wang [19] for the details. Our notation also follows these sources.

We now proceed as follows: At first, we fix the geometric setting and the implicit definition of the refracting and the target surfaces in Section 3.1. Then we apply Snell’s law of refraction in Section 3.2 and follow the path of the light ray in Section 3.3. Finally, in Section 3.4 we obtain the desired equation of Monge–Ampère type.

3.1. The Geometric Setting. Since a lens has two surfaces we need to design both of them. For simplicity we choose a spheric inner surface, i.e. the surface which faces the light source is a subset of a sphere with center at the position of the light source. Thus there is no refraction of the incoming light at this interface, the inner surface is optically inactive.

It remains to compute the shape of the outer surface facing the target area. To that end let us define the quotient \( \kappa = \frac{n_2}{n_1} \) of the refractive indices of the lens material \( n_1 \) and the environment \( n_2 \). We assume that the light source illuminates a non-empty subset \( U \) of the northern hemisphere of the unit sphere \( S^2 \subset \mathbb{R}^3 \). The third component of an incoming light ray with direction \( \mathbf{x} = (x_1, x_2, x_3)^T \in U \) is then given as \( x_3 = \sqrt{1 - x_1^2 - x_2^2} \). Thus we define \( \mathbf{x'} := (x_1, x_2)^T \) and parametrize
our outer lens surface by the distance function $\rho := \rho(x')$, i.e. the surface is given as $\Gamma := \{\rho(x')x : x \in U\}$.

The target $\Sigma$ is defined as a subset of a hypersurface implicitly given by the zero level set of a continuously differentiable function $\psi$ via

$$
\Sigma \subset \{ z \in \mathbb{R}^3 : \psi(z) = 0 \}.
$$

Note that for the numerical solution procedure in the Newton-type method we require that $\psi$ is twice continuously differentiable. While in general much more complicated situations are supported [16], for simplicity we restrict ourselves to the case where the target $\Sigma$ is on a shifted $x$-$y$-plane such that $\psi(z) := z_3 - \gamma$ for a shift $\gamma > 0$.

To model the luminous intensity of the source we define the density function $f : U \to \mathbb{R}^+$, where $\mathbb{R}^+ := \{ x \in \mathbb{R} : x > 0 \}$. The corresponding density function for the desired illumination pattern on the target $\Sigma$ is denoted by $g : \Sigma \to \mathbb{R}^+$. Since we want to redirect all incoming light onto the target the density functions need to fulfill the energy conservation condition

$$
\int_U f \, dS = \int_\Sigma g \, dS.
$$

Note that for simplicity we neglect the loss of reflected light intensity. For a more complicated derivation of a Monge–Ampère-type equation for the refractor problem taking losses into account see [17].

3.2. Snell’s law of refraction. According to Snell’s law of refraction in vectorial notation (see e.g. [13, Chapter 4.4] or [8, Chapter 12]), the direction of the light ray after refraction at the point $\rho x$ is $y = \frac{1}{\kappa}(x - \Phi(x \cdot \nu)\nu) \in S^2$, where $\Phi(s) := s - \sqrt{\kappa^2 + s^2 - 1}$ and $\nu$ is the outer unit normal on $\Gamma$ defined as a function on $U$.

As detailed in [10, (2.15)], for the outer normal unit vector at $x \in U$ we find

$$
\nu = \frac{-\hat{\nabla}\rho(x') + x(\rho(x') + \nabla\rho(x') \cdot x')}{\sqrt{\rho^2(x') + |\nabla\rho(x')|^2 - (\nabla\rho(x') \cdot x')^2}},
$$

where $\nabla f$ denotes the gradient of a function $f$ and $\hat{\nabla}\rho(x') := (\nabla\rho(x'), 0) \in \mathbb{R}^3$. To ease notation, we define the utility function $G$ which represents the denominator in (3.3), i.e.

$$
G(x', u, p) := \sqrt{u^2 + |p|^2 - (p \cdot x')^2}.
$$

3.3. Following the light ray. Next, we consider the line which contains the light ray after refraction, defined by the point $\rho x$ and the direction vector $y$. We now turn to finding the point $z = (z_1, z_2, z_3)^T$ where the refracted light ray hits the target $\Sigma$. In order to determine the third component $z_3$ of $z$, we first define the utility point $w = (w_1, w_2, 0)^T$ as the intersection point of this line with the plane $\{ x \in \mathbb{R}^3 : x_3 = 0 \}$ which is given as

$$
w = \rho(x') x + d_0 \nu
$$

for a $d_0 \in \mathbb{R}$. For a proof of the existence of $w$ see [10, Appendix A.2]. Using (3.3), we confirm that

$$
w = F(x', \rho(x'), \nabla\rho(x')) \hat{\nabla}(\rho^2),
$$
where the utility function $F$ is given by

$$
F(x', u, p) := \frac{1}{2} \Phi\left(\frac{u}{G(x', u, p)}\right) - G(x', u, p) + (u + p \cdot x')\Phi\left(\frac{u}{G(x', u, p)}\right).
$$

After refraction at the point $\rho x$ the light ray hits the target $\Sigma$ at point $z = (z_1, z_2, z_3)^T$ given as

$$
z(x') = \rho(x')x + d_1 y = \rho(x')x + t(w - \rho(x')x).
$$

From the third component we know that $t = \frac{\rho x_3 - z_3}{\rho x_3}$.

We introduce the short notation $x \otimes y := xy^T$. Let us define $F := F(x', \rho, \nabla \rho)$ and denote its partial derivatives by $D_{x'} F$, $D_{\rho} F$ and $D_p F$, respectively.

A lengthy computation using standard calculus and some tensor identities of Sherman–Morrison type yields

$$
Dw' = 2\rho F_M D^2 \rho + B
$$

where $M := I + \frac{1}{2} \nabla \rho \otimes D_p F$ and $B := 2F\nabla \rho \otimes \nabla (\rho^2) \otimes D_x F + D_u F \nabla (\rho^2) \otimes \nabla \rho$.

Note that

$$
M^{-1} = I - \frac{\nabla \rho \otimes D_p F}{F + \nabla \rho \cdot D_p F}.
$$

In a bit more involved computation along the same lines we compute

$$
Dz' = 2t \rho F_M (1 - \beta(w' - \rho x') \cdot \tilde{\nabla} \psi)(D^2 \rho + A)
$$

with $\beta := (\nabla \psi \cdot (w - \rho x))^{-1}$, $\tilde{\nabla} \psi := (\psi_{p_1}, \psi_{p_2})$ and $A = A(x', \rho, \nabla \rho)$, where

$$
A := \frac{1}{2t \rho F} M^{-1}(t B + (1 - t)C) \quad \text{and}
$$

$$
C := D(\rho x') + \frac{1}{\rho x_3}(w' - \rho x') \otimes \nabla (\rho x_3).
$$

### 3.4. Monge–Ampère equation.

The energy conservation (3.2) clearly also holds if we replace $U$ with any arbitrary subset $\bar{U} \subset U$ and $\Sigma$ with $\Sigma := T(\bar{U}) \subset \Sigma$, where $T : U \to \Sigma$, $x \mapsto z(x')$. By coordinate transformation this yields the identity $\det(Dz) = f/(g \sqrt{1 - |x'|^2})$. Finally, we can derive the Monge–Ampère equation for the refractor problem

$$
\det(D\rho + A) = \frac{f(x)}{g(z(x'))H} \quad \text{for} \quad x' \in \Omega
$$

where $\Omega := \{(x_1, x_2)^T \in \mathbb{R}^2 : (x_1, x_2, x_3)^T \in U\}$ and $H = H(x', \rho, \nabla \rho)$ is computed by

$$
H := (1 - |x'|^2)|\nabla \psi|(2t)^2 \rho^3 (-\beta)^2 F(\rho \cdot D_p F),
$$

see [16, Appendix A].
Existence and uniqueness of solutions. In general, for boundary value problems with Monge–Ampère equations proving well-posedness, i.e. existence and uniqueness of the solution and continuous dependency on the parameters, is a hard problem, e.g. see [12, Section 1.4] for an example of a discretized Monge–Ampère equation obtained by finite differences on a grid of $4 \times 4$ cells which has 16 different solutions.

Some theoretical results for the existence of a solution for the refractor problem under some appropriate conditions can be found in [16] in Theorem 5.8 for $\kappa < 1$ and Theorem 6.9 for $\kappa > 1$. Additionally there are results on the uniqueness of the solution if just finitely many single points on the target are illuminated, see [16, Theorem 5.7] for $\kappa < 1$ and [16, Theorem 5.8] for $\kappa > 1$.

For proving existence and uniqueness of a solution one typically requires the equation of Monge–Ampère type to be elliptic. A necessary condition is that the right-hand side of (3.7) is positive. For this reason we demand that $\beta < 0$ or, equivalently, $\nabla \psi \cdot (w - \rho x) < 0$. If this term is positive we can simply replace $\psi$ by $-\psi$.

4. The inverse reflector problem. The inverse reflector problem can be modeled as a Monge–Ampère-type equation very similarly to the case of the inverse refractor problem in Section 3 see [19].

Using the same definitions and notation as in Section 4 and introducing the substitution $u := \frac{1}{w}$, we first define

$$t := 1 - \frac{z_3}{x_3}, \quad \tilde{a} := |\nabla u|^2 - (u - \nabla u \cdot x)^2,$$

$$N := I + \frac{x \otimes x}{x_3^2}, \quad \tilde{b} := |\nabla u|^2 + u^2 - (\nabla u \cdot x)^2,$$

$$w := \frac{2}{\tilde{a}} \nabla u, \quad \text{and} \quad z := \frac{1}{\tilde{a}} x + t \left( w - \frac{1}{w} x \right).$$

We assume that $t > 0$, i.e., $\frac{x_3}{\tilde{a}} > z_3$, and $\nabla \psi \cdot (w - \frac{1}{u} x) > 0$. Then the Monge–Ampère equation for the inverse reflector problem reads

$$\det \left( D^2 u + \frac{\tilde{a} z_3}{2t x_3} N \right) = -\frac{(uw - x) \cdot \nabla \psi}{t^2 |\nabla \psi| x_3^2} \cdot \frac{\tilde{a}^3}{4b} \cdot f(x) \cdot g(z),$$


5. Numerical solution of partial differential equations of Monge–Ampère type. The numerical solution of strongly nonlinear second-order PDEs, including those of Monge–Ampère type, is a highly active topic in current mathematical research. There are many different approaches available on the market, see the review paper by Feng, Glowinski and Neilan [12] and also [5] for an overview.

However, most methods are not well-suited for all equations of Monge–Ampère type such that it remains unclear if a particular method can be successfully applied to our problems. In [5] the authors propose to use a spline collocation method which turns out to provide an efficient solution strategy for Monge–Ampère equations arising in the inverse reflector problem.

In Section 5.1 we explain the idea of a collocation method, which reduces the problem to finding an approximation of the solution within a finite dimensional space. Then we discuss the choice of appropriate basis functions in Section 5.2.
5.1. Collocation method. As discretization tool for the Monge–Ampère equations arising in the reflector and refractor problem, we propose a collocation method, see e.g. Bellomo et al. [2] for examples of collocation methods applied to nonlinear problems. Let the PDE $F(x, u, \nabla u, D^2 u) = 0$ in $\Omega$ and constraints $G(x, u, \nabla u) = 0$ on $\partial \Omega$ be given. In this setting we approximate $u$ in a finite-dimensional trial subspace of $C^2(\Omega)$, i.e. for some finite set $I$ and basis functions $(B_i)_{i \in I} \subset C^2(\Omega)$ we choose the ansatz $\hat{u} = \sum_{i \in I} c_i B_i$. Next, we only require that the PDE holds true on a collocation set $\hat{\Omega} \subset \Omega$ which contains only finitely many points. So our approximation $\hat{u}$ of the solution of our PDE satisfies

\[
F(\tau, \hat{u}(\tau), \nabla \hat{u}, D^2 \hat{u}) = 0, \quad \tau \in \hat{\Omega},
\]

\[
G(\tau, \hat{u}(\tau), \nabla \hat{u}) = 0, \quad \tau \in \partial \hat{\Omega}.
\]

This discrete nonlinear system of equations is solved by a quasi-Newton method, which uses trust-region techniques for ensuring global convergence of the method, see Chapter 4.2.1 in [5] and the references cited therein for the details and the proofs.

5.2. Splines and collocation points. We choose to apply a space of spline functions as ansatz space because of their advantageous properties, see e.g. [10, 31, 35] for details on the theory of splines.

For a given interval $[a, b]$ we fix an equidistant knot sequence $T = \{t_i\}_{i=1}^{n+N}$ with $n$-fold knots at the interval end points $a = t_i$ for $1 \leq i \leq n$ and $b = t_i$ for $N + 1 \leq i \leq n$. Moreover, we require that the knot sequence is strictly increasing inside the interval $(a, b)$, i.e. $t_i < t_{i+1}$ for $n \leq i \leq N$.

Then, an appropriate basis for our spline space is given by the B-spline functions $N_{i,n}$ of order $n$ which can be defined via the recursion formula

\[
N_{i,1}(t) = \chi_{[t_i, t_{i+1}]}(t), \quad N_{i,n}(t) = (N_{i,n-1} \ast N_{i,1})(t),
\]

where $\chi_{[t_i, t_{i+1}]}$ is the characteristic function of the interval $[t_i, t_{i+1}] \subset \mathbb{R}$ and the convolution of two functions is defined as $(f \ast g)(x) := \int_\mathbb{R} f(s)g(x - s) \, ds$. Since we require that the ansatz functions are twice differentiable, we choose cubic splines, i.e. $n = 4$.

In two dimensions the ansatz functions on a rectangular domain are obtained via a tensor ansatz and then are used as the $B_i$ in the previous subsection. The collocation points are chosen to coincide with the sequence of equidistant knots. Since this leads to an underdetermined system of equations we use a not-a-knot condition at the very but last knot at each interval end, i.e. we require that the spline function is three times continuously differentiable at this knot. In other words, the restriction of the spline to the union of the two subintervals closest to each interval end is a cubic polynomial and the knot could be removed without changing the spline function. This is a much simpler approach than the one used in the previous work [5 Section 4.2.3] but provides approximately the same accuracy.

6. Numerical solution of equations of Monge–Ampère type for optical applications. Next, we consider the particular difficulties that we have to overcome to efficiently solve the equations of Monge–Ampère type that arise in the reflector and refractor problems.

6.1. Boundary conditions. The boundary conditions for both, the inverse reflector and refractor problems, are realized via a Picard-type iteration as similarly proposed by Froese [13 Section 3.4].
We assume that the light rays hitting the boundary of the optical surface also hit the boundary of the target, i.e. \( z(\partial \Omega) = \partial \Sigma \), see [5, Section 4.5] for the details. This assumption is related to the edge ray principle, see e.g. [39, Appendix B]. Note that \( z \) also depends on the solution \( \rho \) and its derivative \( \nabla \rho \). In order to have a boundary condition which is easier to handle, we do not fix the target coordinate on the boundary but only its normal component. Since we do not know the normal component of the mapping \( z \) for the exact solution we proceed as follows: For solving the nonlinear system of equations from our collocation method we use a Newton-type method producing iterations \( \rho_k \), \( k = 1, 2, \ldots, n_{\text{max}} \), starting with an initial guess \( \rho^0 \).

We denote the corresponding mappings by \( z^k := z(x', \rho_k, \nabla \rho_k) \). In the \( k \)th iteration we require that the outer normal of the mapping \( z^k \) of the current iteration and of the orthogonal projection of the mapping \( z^{k-1} \) of the last iteration onto the boundary coincide, i.e.

\[
\left( z^k - \arg \min_{\tilde{z} \in \partial \Sigma} |\tilde{z} - z^{k-1}|^2 \right) \cdot \nu(x') = 0 \quad \text{for } x' \in \partial \Omega,
\]

see [5, Section 4.5] (cf. also [13, Section 3.3]). The left-hand side is then used as function \( G \) in Section 5.1.

Since the last iteration is involved in the boundary condition the function \( G \) changes in each iteration so that we solve different problems in successive steps. In order to ensure the existence of a solution of the subproblems we follow the approach by Froese [13, Section 3.4] and add a parameter \( c \) in front of the right-hand side of the Monge–Ampère equation (3.7), i.e. we replace \( f \) by \( cf \) where \( c \) is an additional unknown in our equation. An additional constraint to compensate this new degree of freedom is discussed in Section 6.3.

6.2. Ellipticity constraint. For proofs of results for existence and uniqueness of a solution we require the equation of Monge–Ampère type to be elliptic. In order to ensure ellipticity we manipulate the determinant in the same way as explained in [5, Section 4.4] (cf. also [13, Section 4.3]): Let \( W = [W_{i,j}]_{1 \leq i,j \leq 2} \in \mathbb{R}^{2 \times 2} \) be a matrix. For a penalty parameter \( \lambda > 0 \) we define the modified determinant

\[
\det^+ \lambda W := \max\{0, W_{1,1}\} \max\{0, W_{2,2}\} - W_{1,2}^2
- \lambda \left[ \left( \min\{0, W_{1,1}\} \right)^2 + \left( \min\{0, W_{2,2}\} \right)^2 \right]
\]

which we use instead of the determinant in the left-hand side of the Monge–Ampère equation (3.7). For an elliptic solution of this equation the left-hand side is exactly the same for the determinant and the modified determinant, see [5, Lemma 4.2]. Furthermore, each non-elliptic solution of (3.7) is not a solution of this equation, when the determinant is replaced by the modified determinant.

6.3. Choice of the “size” of the refractor. Up to now the refractor is at most uniquely determined up to its size. Therefore we define our initial guess \( u_0 \) of the problem appropriately and search for a solution \( u \) of same size requiring that \( \int_{\Omega} u \, ds = \int_{\Omega} u_0 \, ds \) holds true, see also [5, Section 4.5].

Note that this condition is taken account of by the additional unknown \( c \) introduced in Section 6.1.

6.4. Total internal reflection. In case that \( \kappa < 1 \) it is possible that a ray of light exceeds the critical angle and total internal reflection occurs, such that this light ray does not reach the target. Of course we know that this is not true for the
solution, because we require that all light rays hit the target. However during the iteration process of our nonlinear solver this phenomenon can appear. If this is the case the argument of the square root in the definition of $\Phi$ in Section 3.2 is negative at this position.

To overcome this instability we replace $\Phi(s)$ by its stabilized counterpart $\tilde{\Phi}(s) := s - \sqrt{\max\{0, \kappa^2 + s^2 - 1\}}$. Then the situation of total internal reflection is treated like the case when the light ray hits the surface exactly at the critical angle. The refracted light ray most likely also misses the target and therefore this intermediate step cannot satisfy the Monge–Ampère equation \((3.7)\) such that further iterations are performed.

If total internal reflection doesn’t occur, which is the case we intend to have for our solution, we have $\Phi(s) = \tilde{\Phi}(s)$ and therefore obtain an equivalent problem.

6.5. Nested iteration. The convergence of Newton-type methods sensitively depends on the choice of an initial guess that is close enough to the solution. We apply a nested iteration strategy in order to largely increase the stability of the solver but also in order to accelerate the solution procedure. We start with a coarse grid for the spline surface and a blurred version of the image for the illumination pattern.

The blurring process is necessary because a coarse grid cannot produce a very detailed image on the target area. For this reason we convolve the image, which is given as a raster graphic in our case, with a discrete version of the standard mollifier function $\varphi(x) := \exp(-1/(1-|x|^2))$ if $|x| < 1$ and zero otherwise, namely with $\varphi_n(i,j) := \varphi(2i/n,2j/n)/\sum_{r,s} \varphi(r/n,s/n)$ for $n \in \mathbb{N}$ and indices $i,j \in \mathbb{Z}$ for the pixel coordinates.

If our grid has $N \times N$ nodes we alternately increase the resolution $N$ of the grid and decrease the strength $n$ of blurring, i.e. we solve the problem for different pairs of $(N,n)$, see also \cite{5} Sections 4.3 and 5.2.

6.6. Initial guess. For the refractor problem we simply use the surface of a sphere with center at the position of the light source and a prescribed radius as initial guess.

For the reflector problem we start with a reflective surface producing a homogeneous illumination pattern on the target. We obtain this reflector by first using the method of supporting ellipsoids \cite{20,21} and our collocation technique afterwards, see also \cite{5} Section 5.2.3.

6.7. Minimal gray value. The density function $g$ corresponds to the target illumination on $\Sigma$ and is given by 8 bit digital grayscale images (integer gray values in the range from 0 to 255). Since we divide by $g$ in right-hand side of the Monge–Ampère equation \((3.7)\), the function $g$ should be bounded away from zero. To guarantee this lower bound we adjust the image and use the modified function

\[(6.1)\]

$$\tilde{g}(Z) := g(Z) + \max\{0, L - \min_{Z' \in \Sigma} g(Z')\}$$

with $L \in \mathbb{N}$, see also \cite{5} (5.9). Numerical experiments show that the value $L = 20$ leads to good results. In order to satisfy the energy conservation condition \((3.2)\) the function $\tilde{g}$ needs to be scaled accordingly.

7. Simulation results. In this section we discuss some numerical simulation results obtained by the collocation method for the inverse reflector and refractor problems.
7.1. Lambertian radiator and target illumination. For both optical problems and all of our simulations we use the domain \( U = \{ \mathbf{x} \in S^2 : \mathbf{x}' \in (-\frac{3}{10}, \frac{3}{10})^2 \} \) and a light source with a Lambertian-type emission characteristics. Its emitted luminous intensity \( I(\theta) \) is rotationally symmetric, shows a fast decay and is proportional to \( \cos\left(\frac{2\theta}{3}\right) \), where \( \theta \in [0, \frac{3\pi}{20}] \) is the angle between the \( z \)-axis and the direction of observation. Figure 7.1 shows the emission density function \( f \) depending on our two-dimensional parameter \( \mathbf{x}' \) and on \( \theta \). We choose a light source with this characteristic because the maximum possible angular direction for our rectangular domain is about \( \theta_{\text{max}} = 25^\circ \) and we therefore have a very low intensity at the edges of \( \Omega \) to make the setting more difficult.

As desired target illumination patterns we chose four images with a variety of characteristics, i.e. many different patterns and features, see first row in Figure 7.4. The first three test images are taken from [9], while the fourth test image is our institute’s logo.

7.2. Geometrical setting and verification. Figure 7.2 shows our geometrical setting for the inverse reflector problem where the resulting reflectors are approximately of the size as in this figure. Here we have \( \Sigma = [4, 12] \times [-4, 4] \times \{20\} \).

For the refractor problem the dimensions including the size of the optical surfaces are chosen very similarly to the case of the reflector problem to have a comparable
The calculated reflector or lens is verified using the ray tracing software POV-Ray [7].

7.3. Choice of the parameters. In the nested iteration we successively solve the nonlinear systems of equations for the following pairs \((N,n)\) of grid resolutions: \((16,163), (31,163), (31,55), (61,55), (61,19), (121,19), (121,7), (241,7), (241,3), \) and \((481,3)\), see Section 6.5 for the details. The Newton-type method ends after at most 200 iterations.

The regularization parameter in the modified determinant as defined in Section 6.2 is set to \(\lambda = 10^3\) which turns out to be an appropriate choice for all examples.

7.4. Results. The results of the numerical simulations are depicted in Figure 7.4. In the first row the original test images are shown. The first three of them are chosen to examine different characteristics within the images, like thin straight lines and lettering as in the image “Boat”, see Figure 7.4 (a). Different patterns of high and low contrast are present in the image “Goldhill”, see Figure 7.4 (b), in particular at the windows and roofs of the houses and the surrounding landscape in the background. The image “Mandrill” in Figure 7.4 (c) shows the face of a monkey with a lot of fine details like the whiskers. The fourth and most challenging of our test pictures is the logo of our institute in Figure 7.4 (d) because it shows the highest possible contrast and contains jumps in the gray value from black to white. The iteration counts and timings for the numerical experiments are given in Tables 7.1 and 7.2.

First, we notice that for a given original image the output images obtained by forward simulation for the reflector and refractor problem look very similar. In comparison to the original images the output images are slightly blurred and have a little less contrast but visually they only differ locally at very few locations. Major deviations can be observed in the background of the institute’s logo which is not completely black after the forward simulation of the mirror and the lens. This is because of the minimal gray value needed to avoid the division by zero, see Section 6.7.

We see that all of these characteristics of the first three test images are well preserved by our method. The computing time for the refractor is approximately twice as long as for the reflector but still acceptable with about 4 minutes.

For the fourth test image we had to adjust the parameters in the nested iteration process to handle the sharp edges and work with a finer grid, see Table 7.2. We also raised the minimal gray value in Section 6.7 from 20 to 30 obtaining a proportion...
between black and white of 1 : 9.5. These parameters lead to results showing also a very sharp logo for both the inverse refractor and reflector problems. Note that nevertheless in two stages of the nested iteration for the refractor problem the quasi-Newton method was stopped because the maximal number of iterations was reached without meeting the required tolerances, see Table 7.2. This happened only for two intermediate steps of the nested iteration process while we observe convergence in the last iteration, which shows us that this does not affect the overall method. In the case of the refractor problem the gray line below the letters is irregularly illuminated and slightly too bright. Nevertheless, the shape of this line is reproduced very precisely.

The optically active surface of the lens for the projection of the institute’s logo is displayed in Figure 7.5. Note that the characters used in the logo can be recognized on the surface. We observe that they cover about the half of the lens’ surface while this is not the case in the original image. Of course this is what we expect because we want to redirect a maximal amount of incoming light onto these letters.
Table 7.1: Number of iterations of the Newton-type method for each of the ten nested iterations and overall computing time in seconds for the standard test images in Figure 7.4 (a) "Boat", Figure 7.4 (b) "Goldhill", and Figure 7.4 (c) "Mandrill".

<table>
<thead>
<tr>
<th>Iterations ((N, n))</th>
<th>Refractor</th>
<th>Reflector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boat</td>
<td>Goldhill</td>
</tr>
<tr>
<td>(16, 163)</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>(31, 163)</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>(31, 55)</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td>(61, 55)</td>
<td>43</td>
<td>15</td>
</tr>
<tr>
<td>(61, 19)</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>(121, 19)</td>
<td>41</td>
<td>32</td>
</tr>
<tr>
<td>(121, 7)</td>
<td>47</td>
<td>41</td>
</tr>
<tr>
<td>(241, 7)</td>
<td>46</td>
<td>38</td>
</tr>
<tr>
<td>(241, 3)</td>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td>(481, 3)</td>
<td>45</td>
<td>36</td>
</tr>
</tbody>
</table>

Time / s 227 210 234 90 90 133

Table 7.2: Number of iterations of the Newton-type method for each of the ten nested iterations and overall computing time in seconds for the institute’s logo in Figure 7.4 (d).

<table>
<thead>
<tr>
<th>Iterations ((N, n))</th>
<th>Refractor Institute’s Logo</th>
<th>Reflector Institute’s Logo</th>
</tr>
</thead>
<tbody>
<tr>
<td>(21, 100)</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>(41, 100)</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>(41, 100)</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>(81, 100)</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>(81, 73)</td>
<td>54</td>
<td>19</td>
</tr>
<tr>
<td>(161, 73)</td>
<td>35</td>
<td>13</td>
</tr>
<tr>
<td>(161, 25)</td>
<td>200</td>
<td>90</td>
</tr>
<tr>
<td>(321, 25)</td>
<td>66</td>
<td>20</td>
</tr>
<tr>
<td>(321, 9)</td>
<td>200</td>
<td>155</td>
</tr>
<tr>
<td>(641, 9)</td>
<td>59</td>
<td>22</td>
</tr>
</tbody>
</table>

Time / s 1390 928

Figure 7.5: Outer refractor surface for projecting our institute’s logo.
8. Summary and outlook. For the efficient and stable solution of the inverse reflector and refractor problems we propose a numerical B-spline collocation method which is applied to the formulation of the inverse optical problems as partial differential equations of Monge–Ampère type and appropriate boundary conditions. Several challenges for the construction of a stable numerical solution method have been met, e.g. we detailed how to enforce ellipticity constraints to ensure uniqueness of the solution and how to handle the involved boundary conditions. A nested iteration approach simultaneously considerably improves the convergence behavior and speeds up the numerical procedure.

For the inverse refractor problem our algorithm provides a reliable and fast method to compute one of the two surfaces of the lens under the assumption of a point-shaped light source. Shaping the second surface of the lens, e.g. to minimize Fresnel losses, and exploring possible solution strategies for the problem for extended real light sources are topics of upcoming research.

Acknowledgments. The authors are deeply indebted to Professor Dr. Wolfgang Dahmen for many fruitful and inspiring discussions on the topic of solving equations of Monge–Ampère type. We thank Elisa Friebel, Silke Glas, and Gudula Kämmer for proofreading the manuscript.

REFERENCES


