

Solving the Monge-Ampère Equations for the Inverse Reflector Problem

Kolja Brix[§], Yasemin Hafizogullari[§] and Andreas Platen[§]

Bericht Nr. 396

April 2014

Key words: inverse reflector problem, elliptic Monge-Ampère equation,
B-spline collocation method, Picard-type iteration

AMS Subject Classifications: 35J66, 35J96, 35Q60, 65N21, 65N35

**Institut für Geometrie und Praktische Mathematik
RWTH Aachen**

Templergraben 55, D-52056 Aachen, (Germany)

[§] Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55,
52056 Aachen, Germany, (brix@igpm.rwth-aachen.de, yasemin.hafizogullari@rwth-aachen.de,
andreas.platen@rwth-aachen.de), <http://www.igpm.rwth-aachen.de>

SOLVING THE MONGE-AMPÈRE EQUATIONS FOR THE INVERSE REFLECTOR PROBLEM

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

Abstract. The inverse reflector problem arises in geometrical nonimaging optics: Given a light source and a target, the question is how to design a reflecting free-form surface such that a desired light density distribution is generated on the target, e.g., a projected image on a screen. This optical problem can mathematically be understood as a problem of optimal transport and equivalently be expressed by a secondary boundary value problem of the Monge-Ampère equation, which consists of a highly nonlinear partial differential equation of second order and constraints. In our approach the Monge-Ampère equation is numerically solved using a collocation method based on tensor-product B-splines, in which nested iteration techniques are applied to ensure the convergence of the nonlinear solver and to speed up the calculation. In the numerical method special care has to be taken for the constraint: It enters the discrete problem formulation via a Picard-type iteration. Numerical results are presented as well for benchmark problems for the standard Monge-Ampère equation as for the inverse reflector problem for various images. The designed reflector surfaces are validated by a forward simulation using ray tracing.

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

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

1. Introduction. Suppose we have a light source with a given directional characteristic and we would like to generate a prescribed illumination pattern on a target, e.g., project a logo onto a wall. The classical approach using an aperture has the disadvantage that a part of the light hits the aperture and is lost for target illumination. This efficiency reduction also exists for slide projectors, where some light is absorbed by the film.

In order to use the entire luminous flux emitted by the light source to illuminate the target an optical system comprised of one or more free-form reflectors or free-form lenses can be used instead of an aperture. Then the loss of light can be neglected and the desired illumination pattern is encoded in the shape of the optical surfaces, that in general is unknown. In the following we will mainly focus on this kind of problem with only one mirror, i.e., we shall determine one desired reflecting free-form surface. This nonlinear inverse problem is called the *inverse reflector problem*, which is a well-known problem in nonimaging optics [14, 70].

Already 2000 years ago reflectors projecting images, called Chinese magic mirrors, have been hand-crafted of bronze in China and Japan, but the recipe has been lost and reconstructed several times over the ages; see [7] and [46]. Today such free-form optics are important in illumination applications. For example they are used in automotive industry for the construction of headlights that use the full light emitted by the lamp to illuminate the road but at the same time do not glare oncoming traffic; see, e.g., [71]. Some other applications include homogeneous illumination for machine-vision purposes or the realization of prescribed patterns in architecture illumination.

However, the solution of the inverse reflector problem is anything but trivial. There are many different schemes to determine a desired reflector like the method of supporting ellipsoids [40, 41] and trial and error approaches [2, 30, 47, 60]. The biggest

[§]Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Germany (brix@igpm.rwth-aachen.de, yasemin.hafizogullari@rwth-aachen.de, andreas.platen@rwth-aachen.de), <http://www.igpm.rwth-aachen.de>.

problem of these methods is the computational effort needed to accurately compute reflectors for complex desired illumination patterns on the target. In the beginning of the 21st century it turned out that solving a corresponding partial differential equation (PDE) is a high potential approach for this problem as already indicated in [59] in 2002. This equation is of *Monge-Ampère type*, which is a family of strongly nonlinear second order PDEs; see [36] for details on the theory of Monge-Ampère equations. A first equation of this family was presented by Gaspard Monge at the beginning of the 19th century in his work “Mémoire sur la théorie des déblais et des remblais” [50] and studied later again by André-Marie Ampère in 1820 [1]. Equations of this type often arise in the context of *optimal transport problems*, where the task is to find the optimal way to transport excavated material (French: déblai), e.g., sand, to piles (French: remblai) without losing any of the total mass. These problems often have an economic background: If one minimizes the cost of transport, which is measured by a quadratic cost functional, under the constraint of *total mass conservation*, one obtains a Monge-Ampère equation. [66] In this spirit, the inverse reflector problem deals with the transportation of light under the constraint that the total light flux emitted by the source is redistributed to the target surface.

The rest of this paper is arranged as follows: In Section 2 we review approaches to solve the inverse reflector problem. Section 3 focuses on the modeling of the inverse reflector problem via an equation of Monge-Ampère type. Our new approach for numerically solving equations of Monge-Ampère type is detailed in Section 4. In particular, we explain how to handle the boundary conditions arising in the inverse reflector problem. Numerical results for benchmark problems for the Monge-Ampère equations and for the inverse reflector problem are presented in Section 5. The paper closes with the conclusion and an outlook in Section 6.

2. State of the art of the solution of the inverse reflector problem.

There are plenty of existing approaches for solving the inverse reflector problem; for a detailed overview we refer the reader to [57]. Most of those methods can be grouped into three classes, namely *brute-force approaches*, *methods of supporting ellipsoids*, and *Monge-Ampère approaches*. We give an overview of these methods in Sections 2.1, 2.2, and 2.4, respectively. Other techniques which do not fit into these three classes are discussed in Section 2.3.

2.1. Brute-force approaches. At the beginning of the 21st century many trial and error methods have been developed. The idea of these iterative schemes is as follows: For an initial reflector in optical setup the resulting illumination pattern on the target area is computed using a ray tracing software. Then the illumination pattern is compared with the desired one, where a typical measure of the error is the deviation at previously selected test points in the Euclidean norm. Afterwards the reflector surface is slightly perturbed using ideas from optimization (e.g., simulated annealing) and the setup with the new reflector is simulated again. If the error has decreased, the new reflector is used as initial guess for the next iteration and the procedure is repeated; see, e.g., [2, 30, 47, 60].

The advantage of these methods is, that there are only very few restrictions for the optical setting. For example even extended light sources and mirrors with different reflectivities can be considered. However, the main drawback is the computing time required, because repeated simulations of the setup using costly ray tracing techniques are needed.

Another approach is presented by Weyrich, Peers, Matusik and Rusinkiewicz [69], where the mirror is assumed to be comprised of many small facets. Each facet is

directed to a prescribed position on the target area. This composite reflector in general has a discontinuous surface, which causes artifacts and is not favorable for production. Therefore, in order to smoothen the solution in a post-processing step, the facets are sorted according to their slopes and adjusted with respect to their heights. Nevertheless, the results have relatively low quality and still show many artifacts.

2.2. Methods of supporting ellipsoids. An ellipse in the plane in general has two foci with the property that light emitted at one focus and reflected at the interior of the ellipse is focused at the second focus. Ellipsoids of revolution in three dimensions have the same property. Kochengin and Olikier [40, 41] (see also [56]) therefore proposed the following method: For each point on the target which needs to be illuminated one defines an ellipsoid of revolution whose one focus is located at the light source and the other one at the target point. However, for each target point there are infinitely many ellipsoids of revolution with this property only differing in their diameters. Therefore in an iterative process the diameters of the ellipsoids are determined starting from some initial guess for each ellipsoid. In each iteration the reflector is defined as the convex hull of the intersection of the interiors of all ellipsoids. The result is a reflector whose surface consists of glued surface segments of ellipsoids of revolution. Since the initial reflector in general does not produce the desired illumination on the target in every iteration the diameters of each ellipsoid besides the first one are shrunk until convergence.

This method requires in each step a numerical integration over the emission solid angles of the source and a large number of optimization steps. Therefore the complexity of this scheme quickly grows with the number K of ellipsoids of revolution. Suppose we would like to determine a reflector whose reflection on the target is exact for each target point up to an accuracy $\gamma > 0$, then the number of iterations scales like $\mathcal{O}(\frac{K^4}{\gamma} \log \frac{K^2}{\gamma})$; see [42]. Therefore it is difficult to use this method for practical illumination patterns of higher resolution.

For the special case where the target is assumed to be infinitely far away from the light source Caffarelli, Kochengin, and Olikier [11] developed a variant, which uses paraboloids of revolution instead of ellipsoids.

2.3. Other methods. The simultaneous multiple surfaces (SMS) method developed by Miñano et al. [49] constructs rotational symmetric optical systems which couple a prescribed set of incoming wave fronts with prescribed conjugate wave fronts. This method was also extended to design optics in non-rotational symmetric cases; see, e.g., [6]. Optical systems which are computed using the SMS method can be found in [5, 48, 51]. This method always computes a pair of surfaces. Hence it cannot be applied to solve the inverse reflector problem with just one reflective surface.

Wang [67, 68] shows that the inverse reflector problem is in fact an optimal transportation problem. By taking into account also the dual reflector the problem can be reformulated as a linear optimization problem. Unfortunately, since the number of linear inequality constraints quickly grows with the number of pixels is too large, the complexity for the linear programming is very high and thus this method is not feasible for images of medium or higher resolution.

Another scheme has been developed by Feßler et al. [43, 72], which computes single refractive or reflective surfaces to produce a prescribed density distribution on a prescribed target. To the best of the authors' knowledge the method is not completely published.

2.4. Monge-Ampère approaches. The inverse reflector problem statement for a point light source can be considered as an optimal transportation problem leading to strongly nonlinear second order PDEs of Monge-Ampère type; see, e.g., [39, 61].

Brickell, Marder and Westcott [10] already in 1977 started to develop methods to solve the inverse reflector problem based on Monge-Ampère type equations. Later Engl and Neubauer [25, 52] investigated a conjugate gradient method with certain constraints to solve this problem via a Monge-Ampère type equation. Ries and Muschaweck [59] also developed a method based on this type of equations. However, this numerical method has only been published very fragmentarily to the best of the authors' knowledge.

A compromise between a trial and error method and the solution of a PDE is proposed by Fournier, Cassarly, and Rolland; see [13, 31]. While the case of extended light sources is considered, which poses many additional problems, the discussion is restricted to the special case of rotationally symmetric reflectors. In an iteration at first the reflector for a point light source is computed by solving an ordinary differential equation. Then the resulting surface is tested using a ray tracing software in a setup with an extended light source. If the result is not good enough, the differential equation for a point light source is solved again for an adjusted target illumination pattern, where the modification comes from the difference between the simulated and the desired target illumination. This procedure is repeated until a stopping criterion holds true.

3. Mathematical formulation of the inverse reflector problem. In Section 2 we saw that there are many different approaches to solve the inverse reflector problem. As proposed in the approaches discussed in Section 2.4 we follow the strategy of first modeling the inverse reflector problem using a PDE of Monge-Ampère type and solving this equation in a second step. We now therefore turn to a formulation of the problem in mathematical terms. There exist different approaches to deduce an equation of Monge-Ampère type for this kind of problem [10, 25, 39, 61]. We choose the formulation from Karakhanyan and Wang [39, Proposition 2.2], which is based on an energy conservation equation. The advantage of this approach is that for the special case in which the light source is located in-plane with the target, we obtain a relatively simple Monge-Ampère equation which will be called a Monge-Ampère equation of standard type. In the following we use the notation and results given by Karakhanyan and Wang [39].

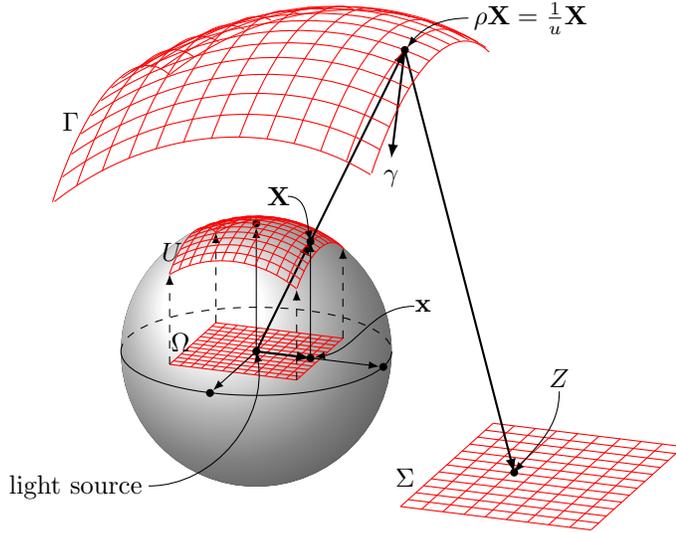
In Subsection 3.1 we first define the problem statement. Then a corresponding equation of Monge-Ampère type is set up in Subsection 3.2 and a result and some remarks on the existence and uniqueness of the solution of this problem are discussed in Subsection 3.3.

3.1. Problem statement. Let us now fix the mathematical description of the problem.

PROBLEM IR. (see, e.g., the introduction in [39])

Let a point light source be given, which emits light in all directions given by the set $U \subset S_2 := \{\mathbf{X} \in \mathbb{R}^3 : \|\mathbf{X}\|_2 = 1\}$. The luminous intensity of the source is modeled by the density function $f : U \rightarrow \mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$. Furthermore we have a target area Σ given in implicit form as $\Sigma := \{\mathbf{Z} \in \mathbb{R}^3 : \psi(\mathbf{Z}) = 0\}$ for an appropriate function ψ . Let $g : \Sigma \rightarrow \mathbb{R}^+$ be a prescribed density function on the target area Σ .

Find a reflector Γ which redistributes the entire light emitted from the light source, such that the given target illumination defined by g is generated; see Figure 3.1.

Figure 3.1: Sketch of the point light source, reflector Γ , and target Σ .

3.2. Monge-Ampère equation. We now aim at deriving the Monge-Ampère type equation corresponding to Problem IR. Therefore we first parameterize the desired reflector by $\Gamma := \{\rho(\mathbf{X})\mathbf{X} : \mathbf{X} \in U\}$, where $\rho : U \rightarrow \mathbb{R}^+$ is an appropriate distance function. As already indicated in Figure 3.1 we assume that U is located in the northern hemisphere of S_2 , such that each $\mathbf{X} \in U$ can be expressed by $\mathbf{X} = (\mathbf{x}, \omega(\mathbf{x}))^T$ where $\mathbf{x} \in \mathbb{R}^2$ and $\omega := \omega(\mathbf{x}) := \sqrt{1 - \|\mathbf{x}\|_2^2}$. Therefore we can use the domain $\Omega := \{\mathbf{x} \in \mathbb{R}^2 : (\mathbf{x}, \omega(\mathbf{x}))^T \in U\}$ in \mathbb{R}^2 to parameterize the set $U \subset \mathbb{R}^3$.

Under these assumptions Karakhanyan and Wang [39, Proposition 2.2] give a PDE for Problem IR for the desired function ρ . However, substituting ρ by $\frac{1}{u}$ results in an equation with slightly simpler expressions (see [39, Remark 2.1]), which leads to the following result.

THEOREM 3.1. (see [39, Remark 2.1 and Proposition 2.2])

Let the density functions f and g be given, satisfying the energy conservation

$$(3.1) \quad \int_U f \, dS = \int_\Sigma g \, dS.$$

Moreover, let the shape of Σ be defined by the function ψ as in Problem IR. Define

$$\begin{aligned} \omega &:= \sqrt{1 - \|\mathbf{x}\|_2^2}, & t &:= 1 - u \frac{z_3}{\omega}, \\ \tilde{a} &:= \|Du\|_2^2 - (u - Du^T \mathbf{x})^2, & \hat{D}u &:= (Du, 0)^T, \\ \tilde{b} &:= \|Du\|_2^2 + u^2 - (Du^T \mathbf{x})^2, & \mathcal{N} &:= I + \frac{\mathbf{x}\mathbf{x}^T}{\omega^2}, \end{aligned}$$

and

$$\mathbf{X} = (\mathbf{x}, \omega), \quad \mathbf{Z}_0 := \frac{2}{\tilde{a}} \hat{D}u, \quad \mathbf{Z} = \frac{1}{u} \mathbf{X} + t \left(\mathbf{Z}_0 - \frac{1}{u} \mathbf{X} \right),$$

where $\mathbf{X} = (x_1, x_2, x_3)^T \in U$ and $\mathbf{Z} = (z_1, z_2, z_3)^T \in \Sigma$. We assume that $t > 0$, i.e., $\frac{z}{u} > z_3$, and $\nabla\psi^T(\mathbf{Z} - \frac{1}{u}\mathbf{X}) > 0$. Then the inverse reflector problem states:

Find a function $u : \Omega \rightarrow \mathbb{R}$, such that

$$(3.2) \quad \det \left(D^2u + \frac{\tilde{a}z_3}{2tx_3} \mathcal{N} \right) = - \frac{(u\mathbf{Z}_0 - \mathbf{X}) \cdot \nabla\psi}{t^2 \|\nabla\psi\|_2 \omega} \cdot \frac{\tilde{a}^3}{4\tilde{b}} \cdot \frac{f(\mathbf{X})}{\omega g(\mathbf{Z})} \quad \text{in } \Omega \text{ and}$$

$$(3.3) \quad T : U \rightarrow \Sigma, \quad \mathbf{X} \mapsto \mathbf{Z} \quad \text{is surjective.}$$

Let us note some comments on the above theorem.

REMARK 1.

- a) The assumption $\frac{z}{u} > z_3$ means, that the third component of the vector that corresponds to the position where a light ray hits the reflector is larger than the third component of the point where this light ray hits the target Σ after reflection. In other words, the light ray must be directed downwards after the reflection.
- b) The second condition $\nabla\psi^T(\mathbf{Z} - \frac{1}{u}\mathbf{X}) > 0$ is needed for a technical reason and just fixes the direction of the normal on Σ . If this constraint is not fulfilled one can simply replace $\psi(\mathbf{Z})$ by $-\psi(\mathbf{Z})$ for all $\mathbf{Z} \in \mathbb{R}^3$.
- c) If one sets $z_3 \equiv 0$, the equation (3.2) simplifies to

$$(3.4) \quad \det(D^2u) = - \frac{(u\mathbf{Z}_0 - \mathbf{X}) \cdot \nabla\psi}{\|\nabla\psi\|_2 \omega} \cdot \frac{\tilde{a}^3}{4\tilde{b}} \cdot \frac{f(\mathbf{X})}{\omega g(\mathbf{Z}_0)},$$

which is a Monge-Ampère equation of standard type, i.e., the left hand side is only the determinant of the Hessian of u ; see [39, Remark 2.1].

3.3. Existence and uniqueness. The existence of solutions of Problem IR is ensured by the following result.

THEOREM 3.2. (see [39, Theorem A])

Suppose we have two functions f and g which fulfill the energy conservation condition (3.1) as in Theorem 3.1. Let moreover \mathbf{p} be an element of the light cone \mathcal{C}_U of the light source, i.e.,

$$\mathbf{p} \in \mathcal{C}_U := \left\{ \mathbf{p} \in \mathbb{R}^3 : \frac{\mathbf{p}}{\|\mathbf{p}\|_2} \in U \right\},$$

and one of the conditions

a) $\|\mathbf{p}\|_2 > 2 \sup_{\mathbf{q} \in \Sigma} \|\mathbf{q}\|_2$ or

b) $\Sigma \subset \mathcal{C}_V := \{t\mathbf{X} : t > 0, \mathbf{X} \in V\}$ for a region $V \subset S^2$ with $\bar{U} \cap \bar{V} = \emptyset$

is satisfied. Then there exists a reflecting surface, which is the solution of the inverse reflector problem IR and contains the point \mathbf{p} .

We first notice that the solution is not unique. If we have a solution for one $\mathbf{p} \in \mathcal{C}_U$, we know that there exist other solutions for $c\mathbf{p}$ for each $c > 1$.

Even if we fix $\mathbf{p} \in \mathcal{C}_U$, there are at least two solutions that contain \mathbf{p} . A solution is called R -convex if it fulfills an *ellipticity constraint*, i.e., the matrix $D^2u + \frac{\tilde{a}z_3}{2tx_3} \mathcal{N}$ in (3.2) is required to be positive definite. Moreover, a solution is defined to be R -concave, if this matrix is negative definite. In [39, Section 7] one can find a sketch of a proof that for a fixed $\mathbf{p} \in \mathcal{C}_U$ there are exactly one R -convex and one R -concave solution. Therefore we need at least to fix the size of the reflector by a point \mathbf{p} and search for either a R -convex or a R -concave solutions to obtain uniqueness. This is a necessary condition to ensure well-posedness of the problem.

4. Solving the Monge-Ampère equations. We now discuss methods to numerically solve equations of Monge-Ampère type, which is particularly difficult due to the strong nonlinearity of this type of equations.

After an overview of numerical solvers for Monge-Ampère type equations in Subsection 4.1 we discuss our approach in Subsection 4.2. In order to improve the convergence properties and to speed up the solution, a multilevel technique will be introduced in Subsection 4.3. Since the existence and uniqueness of a solution is guaranteed, if the Monge-Ampère type equation fulfills an ellipticity condition (see, e.g., [64, Theorem 1.1 and Remarks (i)]), we ensure that this condition holds true by adding a convexity constraint to the equation as detailed in Subsection 4.4. Since the boundary condition (3.3) cannot be considered directly, the section closes with the presentation of a technique to realize this type of boundary conditions in Subsection 4.5.

4.1. State of the art. There are several other numerical methods for the solution of Monge-Ampère type equations known, but most of them have clear limitations. Usually the algorithms are only designed to handle boundary value problems for the *Monge-Ampère equation of standard type*

$$(4.1) \quad \det(D^2u(\mathbf{x})) = f(\mathbf{x})$$

for any $\mathbf{x} \in \Omega \subset \mathbb{R}^2$ with Dirichlet boundary conditions $u(\mathbf{x}) = g(\mathbf{x})$ for $\mathbf{x} \in \partial\Omega$. In particular, in this case the left-hand side determinant may only depend on the Hessian of the solution, but no perturbations in the determinant like $\det(D^2u + A) = f$ for a matrix A are permitted. Numerical methods for this Monge-Ampère equation of standard type can for example be found in [3, 4, 8, 18, 19, 27, 55].

In order to solve the inverse reflector problem we search for a solution of the Monge-Ampère equation given in Theorem 3.1. Since equation (3.2) is strongly nonlinear with rather cumbersome terms it is very difficult to analyze this equation particularly with regard to weak formulations. To the best knowledge of the authors, there is no closed theory on weak formulations in the classical sense for equations of Monge-Ampère type available. Hence, Feng and Neilan [27, 28, 29] introduce a new type of weak solution called the moment solution and investigate the following ansatz. They embed strongly nonlinear PDEs into linear PDEs of higher order and study the limit of vanishing highest order term. For example, the Monge-Ampère equation (4.1) is embedded into an quasilinear elliptic PDE of fourth-order with highest order term $\varepsilon\Delta^2u$ with $\varepsilon > 0$, where the limit $\varepsilon \rightarrow 0$ is studied. Such a scheme is called vanishing moment method; for details we refer to [27, 28, 29].

Another interesting approach has recently been published by Brenner et al. [8] and by Brenner and Neilan [9]. They propose a finite element method that leads to a sophisticated consistent discretization and show that the corresponding discrete linearized problem is stable. The main idea is to employ standard continuous Lagrange finite elements instead of finite elements with higher smoothness and to use penalty terms, like those applied in Discontinuous Galerkin methods, to demand for regularity of the solution across interfaces.

In view of practical applications like the inverse reflector problem, it is necessary that a numerical solver efficiently treats Monge-Ampère equations of standard type as well as perturbed equations with Neumann boundary conditions. At least one of the two methods given by Benamou, Froese, and Oberman [4] supports Neumann boundary conditions but is only suited to solve Monge-Ampère equations of standard type. In a subsequent work Froese [32] presents a method for Monge-Ampère type equations arising in optimal transportation problems. She uses a Neumann boundary

condition and the right hand side of the Monge-Ampère equation of standard type (4.1) is allowed to also depend on the gradient of u . The method by Brenner et al. [8] also permits this dependency on the right hand side.

A detailed overview for numerical methods for fully nonlinear second order PDEs including methods for Monge-Ampère type equations can be found in the review article [26] by Feng, Glowinski, and Neilan.

4.2. Spline collocation method. We now explore a different simple but very flexible approach for the solution of the Monge-Ampère type equations: The *collocation method* can directly be applied to the strong formulation such as (3.2).

The solution is approximated in a finite-dimensional trial space, in our case we choose the space of spline functions because of its good approximation properties. The spline space is spanned by B-spline functions which form an advantageous basis due to its flexible manageability. Moreover this basis is well-known to be numerically very stable and the functions are of minimal support, which favors sparsity in the collocation matrices.

The rest of this subsection is arranged as follows: First we formulate the collocation method in Subsection 4.2.1. The trial space and a modified B-spline basis that is suited for our particular choice of the collocation points are set up in one spatial dimension in Subsections 4.2.2 and 4.2.3. Finally, the trial space is extended to the two-dimensional case in Subsection 4.2.4 via a tensor construction.

4.2.1. Collocation. Let us now formulate the collocation method for a general second order PDE and therefore introduce some notation. Let $\Omega := (a, b) \times (c, d) \subset \mathbb{R}^2$ be a rectangular domain and let the boundary value problem be given as

$$(4.2) \quad F(x, y, u(x, y), Du, D^2u) = 0, \quad \text{for } (x, y)^T \in \Omega,$$

$$(4.3) \quad G(x, y, u(x, y), Du) = 0, \quad \text{for } (x, y)^T \in \partial\Omega.$$

We approximate the exact solution in a finite-dimensional trial subspace, say of dimension $n \in \mathbb{N}$, spanned by some basis functions $B_1, \dots, B_n \in C^2(\Omega)$. Then the approximate solution is written as $\hat{u}(x, y) := \sum_{i=1}^n c_i B_i(x, y)$, where $(x, y)^T \in \bar{\Omega}$ and c_1, \dots, c_n are the desired coefficients.

Of course we cannot expect such a discrete solution to fulfill the PDE exactly in the whole domain Ω . The idea of the collocation method is that the PDE should be fulfilled pointwise at certain *collocation points*. Therefore we choose n appropriate pairwise different collocation points, i.e., we select two finite and nonempty subsets $\hat{\Omega} \subset \Omega$ and $\partial\hat{\Omega} \subset \partial\Omega$ with cardinality $|\hat{\Omega} \cup \partial\hat{\Omega}| = n$. Our problem (4.2), (4.3) is then required to be fulfilled exactly at these points, i.e., we end up with a discrete problem which is the nonlinear system of equations

$$(4.4) \quad F(\tau, \mu, \hat{u}(\tau, \mu), D\hat{u}, D^2\hat{u}) = 0, \quad \text{for } (\tau, \mu)^T \in \hat{\Omega},$$

$$(4.5) \quad G(\tau, \mu, \hat{u}(\tau, \mu), D\hat{u}) = 0, \quad \text{for } (\tau, \mu)^T \in \partial\hat{\Omega}.$$

Now we can use some *Newton-type method* to solve (4.4). For reasons of better stability we favor the application of the *double-dogleg method* (see, e.g., [24]), which is a trust region algorithm of quasi-Newton type. We use the variant proposed by Dennis and Mei [23], where we invoke an algorithm by Nielsen [53] for the choice of the trust region radius after each iteration step. Adequate stopping criteria for the iteration process can be found in [45].

Next we will discuss the choice of appropriate basis function and collocation points.

4.2.2. B-splines. Let us briefly recall the definition of *B-splines*, which are the fundament of our basis functions, using the notation of [15].

DEFINITION 4.1. *Let $[a, b] \subset \mathbb{R}$ be a given interval that serves as the domain of the basis functions. For B-splines of order $n \in \mathbb{N}$ we define for a fixed $N \in \mathbb{N}$ a strictly increasing knot sequence $T = \{t_i\}_{i=1}^{N+n} \subset \mathbb{R}$ with*

$$(4.6) \quad t_1 < \dots < t_n := a < \dots < t_{N+1} := b < \dots < t_{N+n}.$$

For $i = 1, \dots, N$ the i -th B-spline $N_{i,n}$ of order n is then defined by the Cox-de Boor recursion formula

$$N_{i,1}(x) := \chi_{[t_i, t_{i+1})}(x) := \begin{cases} 1, & \text{if } x \in [t_i, t_{i+1}), \\ 0, & \text{otherwise,} \end{cases} \quad \text{if } n = 1 \text{ and}$$

$$N_{i,n}(x) := \frac{x - t_i}{t_{i+n-1} - t_i} N_{i,n-1}(x) + \frac{t_{i+n} - x}{t_{i+n} - t_{i+1}} N_{i+1,n-1}(x) \quad \text{if } n \geq 2.$$

REMARK 2. *The condition of a strictly increasing knot sequence (4.6) can be relaxed using de l'Hôpital's rule and multiple knots can be permitted; see, e.g., [17, Chapter IX] or [58, Section 2.2].*

In the following we fix the outer knots in (4.6) at the boundary, i.e., we set $t_1 := \dots := t_{n-1} := a$ and $t_{N+2} := \dots := t_{N+n} := b$. Multiple knots for the interior knots t_n, \dots, t_{N+1} result in less smooth B-splines. Since we want to solve a PDE of second order, we need basis functions which are twice differential. To have minimal computational efforts while fulfilling this constraint we choose the lowest possible order, which is $n = 4$, i.e., *cubic splines*, and avoid multiple knots inside (a, b) .

REMARK 3. *Since the left outer knots all coincide with the left boundary a , $N_{1,n}$ is the only B-spline with a non-zero function value at the left boundary point a . Moreover, the first derivative of $N_{j,n}$ vanishes at a for all $j > 2$ and the second derivative of $N_{j,n}$ vanishes at a for all $j > 3$. By symmetry, the same holds for the B-splines at the right boundary point b .*

4.2.3. Collocation points and modification of the B-spline basis. In order to uniquely define a spline from our spline space it suffices to set N linear independent conditions, e.g., to prescribe the function values at N appropriate different nodes. The knots themselves are possibly a good choice for these nodes. However, we only have $N - 2$ knots in $[a, b]$, that is there are two open degrees of freedom. There are different ways to handle open degrees of freedom, e.g., by setting a not-a-knot condition [17, Chapter IV]. Another possibility is to modify the trial space by lowering the dimension of the spline space, which we will discuss next.

For our purpose it is crucial to keep the approximation properties of the trial space. Therefore it is necessary that each (Taylor-)polynomial of degree $\leq n - 1$ can still be reproduced and consequently at least n basis functions must be supported in each subinterval. With regard to the boundary conditions and the clearness of the construction of the modified basis, we build new basis functions at the boundary of the interval $[a, b]$. Without loss of generality we can restrict ourselves to the left boundary case, the right boundary case is handled analogously by symmetry.

In order to modify as few basis functions as possible we choose four new basis functions from the span of the first five B-splines, i.e., we write

$$B_{i,4} := \begin{cases} \sum_{j=1}^5 a_{i,j} N_{j,4} & \text{for } i = 1, \dots, 4, \\ N_{i+1,4} & \text{for } i = 5, \dots, N-6, \\ \sum_{j=N-4}^N a_{i,j} N_{j,4} & \text{for } i = N-5, \dots, N-2 \end{cases}$$

for some coefficients $a_{i,j} \in \mathbb{R}$, which still have to be determined. Note that this approach of gluing B-splines at the boundary is similar to the procedure used in the construction of WEB-splines [38] where some B-splines are glued to improve the numerical stability of the basis, i.e., to lower its condition number.

From *Marsden's identity* we can derive

$$(4.7) \quad x^m = \sum_{j=1}^N \frac{(-1)^{n-1-m} \psi_{j,n}^{(n-1-m)}(0)}{(n-1) \cdot \dots \cdot (m+1)} N_{j,n}(x) \quad \text{with} \quad \psi_{j,n}(y) := \prod_{l=1}^{n-1} (t_{j+l} - y)$$

for $m = 0, \dots, n-1$; see, e.g., [15, (2.3.3)].

Furthermore the coefficient matrix $A := [a_{i,j}]_{i=1, \dots, 4, j=1, \dots, 5}$ has to be chosen in such a way, that for each $m \in \{0, 1, 2, 3\}$ there exist coefficients $\{c_{i,m}\}_{i=1}^{N-2}$ so that

$$(4.8) \quad \begin{aligned} x^m &= \sum_{i=1}^N c_{i,m} B_{i,4}(x) \\ &= \sum_{j=1}^5 \left(\sum_{i=1}^4 c_{i,m} a_{i,j} \right) N_{j,4} + \sum_{i=6}^{N-5} c_{i,m} N_{i+1,4} + \sum_{j=N-4}^N \left(\sum_{i=N-5}^{N-2} c_{i,m} a_{i,j} \right) N_{j,4} \end{aligned}$$

holds. Setting $C := [c_{i,m}]_{1 \leq i \leq 4, 0 \leq m \leq 3}$ and equating the coefficients of $N_{j,4}$ for $j = 1, \dots, 5$ in (4.7) and (4.8) yields the system of linear equations

$$(4.9) \quad A^T C = B$$

where $B := [b_{j,m}]_{1 \leq j \leq 5, 0 \leq m \leq 3}$ with $b_{j,m} := \frac{(-1)^{n-1-m} \psi_{j,n}^{(n-1-m)}(0)}{(n-1) \cdot \dots \cdot (m+1)}$. Now we have to choose the matrix A , such that (4.9) has a solution matrix C .

The trivial solution is $A := B^T$ and $C := I$, where $I \in \mathbb{R}^{4 \times 4}$ is the identity matrix. In other words the new basis elements then are exactly the monomials 1 , x , x^2 , and x^3 at the first subinterval. The disadvantage of this solution is, that the new basis functions are not shift invariant, i.e., they depend on the location of the knots in the knot sequence T . Therefore we will search for a more favorable solution of shift invariant functions.

For this purpose let us choose the ansatz $(BP)P^{-1} = B$ for an invertible matrix $P \in \mathbb{R}^{4 \times 4}$ and set $A := (BP)^T$ and $C := P^{-1}$. Thus the idea is to define the matrix A^T using appropriate column operations acting on B , such that A^T has a preferably simple form.

Of course the choice of P is not unique, such that we can prescribe additional conditions, e.g., in the spirit of the properties of the B-spline functions observed in Remark 3. For Dirichlet boundary conditions it is advantageous, if only one basis function is nonzero at the boundary. The same holds true for Neumann boundary conditions when the first derivative of the basis functions is considered. Moreover, for natural spline interpolation (see, e.g., [17, Chapter IV]) it is convenient if only one

basis function has a non-vanishing second derivative at the boundary. The following choice provides us with a basis which has all these handy properties.

Suppose that we have an equidistant knot sequence with multiple knots at the boundary, i.e., $T = \{t_i\}_{i=1}^{N+4} \subset [a, b] \subset \mathbb{R}$ with

$$(4.10) \quad t_i := \begin{cases} a, & \text{if } 1 \leq i \leq 4, \\ a + \frac{b-a}{N-3}(i-4), & \text{if } 5 \leq i \leq N, \\ b, & \text{if } N+1 \leq i \leq N+4. \end{cases}$$

After some elementary computations for this particular case, we can determine an invertible matrix P such that

$$(4.11) \quad A = (BP)^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{4} & 0 & 0 \\ 1 & \frac{3}{4} & \frac{1}{2} & 0 \\ \frac{3}{4} & \frac{15}{16} & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}^T.$$

Note that the matrix A is independent of the knots, such that our new basis functions inherit the shift invariance from the B-splines. A plot of the resulting four new basis functions is given in Figure 4.1.

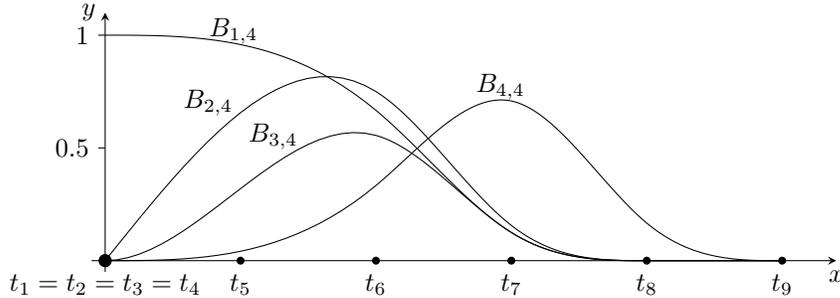


Figure 4.1: Modified basis functions given by A in (4.11).

REMARK 4. Due to the conditions imposed and the properties pointed out in Remark 3, the matrix A^T in (4.11) has a lower triangular structure.

4.2.4. Tensor-product B-spline basis. Since we aim at solving equations of Monge-Ampère type on a two-dimensional rectangle $R = [a, b] \times [c, d] \subset \mathbb{R}^2$, we next define the spline space on R using the usual tensor product construction. Let T_1 be the knot sequence as in (4.10) for the interval $[a, b]$ and T_2 the analog knot sequence for $[c, d]$. As in Section 4.2.2 for both T_1 and T_2 we define B-splines of order n_1 and n_2 , respectively. We then define *tensor-product B-splines* by

$$N_{i,j;n_1,n_2}(x, y) := N_{i,n_1}(x) N_{j,n_2}(y)$$

for $i = 1, \dots, N_1$, $j = 1, \dots, N_2$ and $(x, y) \in [a, b] \times [c, d]$.

Using the same arguments as in Subsection 4.2.2 we set $n_1 = n_2 = 4$ and restrict ourselves to knot sequences that are equidistant in the inner of the intervals. As

ansatz functions in our collocation method in Subsection 4.2.1 we use the modified tensor-product basis functions

$$B_{i,j;4}(x, y) := B_{i,4}(x) B_{j,4}(y)$$

for $i = 1, \dots, N_1 - 2$ and $j = 1, \dots, N_2 - 2$ and, as in [17, Chapter XIII], the collocation points

$$\begin{aligned} \hat{\Omega} &:= \{(\tau, \mu)^T \in \Omega : \tau \in T_1, \mu \in T_2\} \subset \Omega, \\ \partial\hat{\Omega} &:= \{(\tau, \mu)^T \in \partial\Omega : \tau \in T_1, \mu \in T_2\} \subset \partial\Omega, \end{aligned}$$

which are the knots of the B-splines.

4.3. Nested iteration. Splines are particularly suitable for multilevel strategies, because of the following property: Suppose we have a knot sequence $\hat{T} := \{\hat{t}_i\}_{i=1}^{M+n} \subset \mathbb{R}$ for the interval $[a, b] \subset \mathbb{R}$ as in (4.10) and a second knot sequence $T := \{t_i\}_{i=1}^{N+n}$ for $[a, b] \subset \mathbb{R}$ that is obtained from \hat{T} by inserting new knots. Then the corresponding spline spaces $S_n(\hat{T})$ and $S_n(T)$ of order $n \in \mathbb{N}$ are nested, i.e., $S_n(\hat{T})$ is a subspace of $S_n(T)$.

Since we use a Newton-type method for solving the discrete nonlinear problem (4.4), (4.5) we need an initial guess that is preferably close to the solution. Otherwise solving the problem could be very time consuming or even infeasible, because the Newton-type method might not converge. We therefore follow an approach based on *nested iteration*: Our computation starts on a very coarse grid. After calculating the solution for this coarse problem, we refine the mesh by knot insertion, e.g., we halve the mesh size in each coordinate direction. Since the spline spaces of the coarse and refined knot sequences are nested, the coarse solution is also contained in the spline space corresponding to the finer mesh and can be used as the initial guess. We continue with this process until we reach the desired grid resolution.

To be more precise, assume we have $N = N_1 = N_2$, i.e., the number of knots is the same in each coordinate direction. Let us now solve the discrete nonlinear problem on a grid with N^* knots in each direction. We start with the coarsest reasonable grid possible in our situation such that the boundary basis functions do not overlap, which is of size $N^0 := 11$, and after each iteration we halve the mesh size. Then after nested iteration $k \in \mathbb{N}$ there are $N^k := 2N^{k-1} - 1$ knots in the grid in each coordinate direction. If $N^{k+1} \geq N^*$ we interpolate the solution obtained from the k -th iteration to the fine grid with $N^* \times N^*$ knots using spline interpolation and solve the final nonlinear problem.

4.4. Convexity constraint for Monge-Ampère equations. In order to show the existence and the uniqueness of solutions of Monge-Ampère type equations, it is often required that the equation is elliptic with respect to the solution; see, e.g., [64, Theorem 1.1 and Remarks (i)] or [63].

A nonlinear PDE $F[u] = F(\cdot, u, Du, D^2u) = 0$ is said to be *elliptic*, if the matrix

$$(4.12) \quad \left[\frac{\partial F}{\partial r_{ij}}(\gamma) \right]_{i,j} := \begin{bmatrix} \frac{\partial F}{\partial r_{11}}(\gamma) & \dots & \frac{\partial F}{\partial r_{1n}}(\gamma) \\ \vdots & & \vdots \\ \frac{\partial F}{\partial r_{n1}}(\gamma) & \dots & \frac{\partial F}{\partial r_{nn}}(\gamma) \end{bmatrix}$$

is positive definite for all $\gamma = (x, z, p, r) \in V \subset \Omega \times \mathbb{R} \times \mathbb{R}^n \times \tilde{\mathbb{R}}^{n \times n}$, where $\tilde{\mathbb{R}}^{n \times n}$ denotes the space of symmetric $n \times n$ matrices; see, e.g., [35, Chapter 17].

In case of a general Monge-Ampère equation

$$(4.13) \quad \det(D^2u + A(\cdot, u, Du)) = f(\cdot, u, Du) \quad \text{in } \Omega$$

with appropriate boundary conditions we have the following result.

LEMMA 4.2. *Let $\Omega \subset \mathbb{R}^n$ be a domain, $u \in C^2(\Omega)$, and let A be a symmetric $n \times n$ matrix. It is necessary and sufficient to ensure ellipticity for (4.13) that*

$$(4.14) \quad D^2u + A(\cdot, u, Du) \text{ is positive definite.}$$

Consequently the right-hand side f has to be positive to allow elliptic solutions.

Proof. For $A \equiv 0$ this result has been proven, e.g., in [33, Lemma 1]. Otherwise, the result is also well-known (see, e.g., [63]), but the authors could not find a proof in the literature. We therefore extend the proof given by Froese and Oberman [33, Lemma 1] to this more general result.

Let \mathcal{W} be the cofactor matrix of the symmetric matrix $\mathcal{W} := D^2u + A$; see, e.g., [62, Section 4.3] for a definition. As a consequence of Cramer's rule we have $\mathcal{W}^{-1} \det(\mathcal{W}) = \tilde{\mathcal{W}}^T$. It follows that $\tilde{\mathcal{W}}$ is positive definite if and only if \mathcal{W} is positive definite. Therefore we only have to prove that

$$(4.15) \quad D_r \det(\mathcal{W}) := \left[\frac{\partial \det(\mathcal{W})}{\partial (D^2u)_{i,j}} \right]_{i,j} = \tilde{\mathcal{W}}$$

is true, which is an alternative expression of (4.12) for the Monge-Ampère equation (4.13).

Expanding the determinant along the i -th row using Laplace's formula yields

$$\det(\mathcal{W}) = \sum_{j=0}^n \tilde{\mathcal{W}}_{i,j} \mathcal{W}_{i,j} = \sum_{j=0}^n \tilde{\mathcal{W}}_{i,j} ((D^2u)_{i,j} + A_{i,j}).$$

By definition the cofactor $\tilde{\mathcal{W}}_{i,j}$ is independent of $\mathcal{W}_{i,j}$ and therefore of $(D^2u)_{i,j}$ as well. In addition, the matrix A is independent of D^2u , such that we have

$$\frac{\partial \det(\mathcal{W})}{\partial (D^2u)_{i,j}} = \tilde{\mathcal{W}}_{i,j}.$$

We therefore proved that (4.12) for the Monge-Ampère equation (4.13) is positive definite if and only if $D^2u + A$ is positive definite. \square

REMARK 5. *In case that $A \equiv 0$, the Hessian matrix D^2u must be positive definite to ensure ellipticity for (4.13), which means that u has to be strictly convex. Therefore condition (4.14) can be viewed as some kind of convexity condition.*

As proposed in [33, 34] we take (4.14) into account as an additional constraint. Note that the positive definiteness of a symmetric matrix from $\mathbb{R}^{2 \times 2}$ is equivalent to a positive determinant and positive diagonal entries. The positivity of the determinant is guaranteed if $f > 0$. To forbid solutions with non-positive diagonal entries in the matrix (4.14) we define the modified determinant

$$\det^+(\mathcal{W}) := \max\{0, \mathcal{W}_{1,1}\} \max\{0, \mathcal{W}_{2,2}\} - \mathcal{W}_{1,2}^2,$$

which is non-positive if at least one diagonal entry is non-positive and otherwise equals $\det(\mathcal{W})$. This idea is similar to the modified determinant in [33, 34] and also in [32, Section 4.3].

To avoid problems, such as non-uniqueness of a solution in case of a singularity, where $D^2u + A$ has an eigenvalue equal to zero, we introduce a parameter $\lambda \geq 0$ and subtract a penalty term

$$(4.16) \quad \det_{\lambda}^+ \mathcal{W} := \det^+ \mathcal{W} - \lambda [(\min\{0, \mathcal{W}_{1,1}\})^2 + (\min\{0, \mathcal{W}_{2,2}\})^2]$$

to ensure that the modified determinant has a negative value instead of just a non-positive value. This was done similarly by Froese [32, Section 4.3]. Now we replace the determinant in (4.13), consider the equation

$$(4.17) \quad \det_{\lambda}^+ (D^2u + A(\cdot, u, Du)) = f(\cdot, u, Du)$$

instead, and obtain the following result.

LEMMA 4.3. *Let $\Omega \subset \mathbb{R}^2$ be a domain, $u \in C^2(\Omega)$, and $f > 0$. If u is a classical solution of the modified Monge-Ampère equation (4.17), then u also solves the original equation (4.13) and simultaneously fulfills the ellipticity constraint (4.14) and vice versa.*

Proof. Let $u \in C^2(\Omega)$ be a classical solution of (4.17) and $\mathcal{W} := D^2u + A(\cdot, u, Du)$. Since $\det_{\lambda}^+ \mathcal{W} = f > 0$, the diagonal entries of \mathcal{W} are positive because otherwise we would have $\det_{\lambda}^+ \mathcal{W} \leq 0 < f$. Therefore $f = \det_{\lambda}^+ \mathcal{W} = \det \mathcal{W}$ holds true. It immediately follows, that u also solves (4.13) and (4.14) simultaneously.

Conversely, if u solves (4.13) and (4.14), the diagonal entries of \mathcal{W} are positive and we have $f = \det \mathcal{W} = \det_{\lambda}^+ \mathcal{W}$. \square

4.5. Boundary conditions for the inverse reflector problem. Let us now come back to the solution of the inverse reflector problem. The equation of Monge-Ampère type (3.2) in Theorem 3.1 already is of the desired form (4.2) to be treated by the collocation method. But this is not true for the boundary condition (3.3), because it is a constraint for the desired mapping T on the whole domain U and not only for T restricted to the boundary of U , i.e., for $T|_{\partial U}$. Thus condition (3.3) is not of the general form (4.3), which makes it difficult to handle. To overcome this problem we first assume that the boundary of U is supposed to be mapped by T onto the boundary of Σ and the interior of U to the interior of Σ . To the best of the authors' knowledge this assumption has not yet been proven to hold for the inverse reflector problem. But it is worthwhile noting that when the mirror surface is interpreted as an extended light source this assumption corresponds to the *edge ray principle* from nonimaging optics [70, Chapter 3]. In brief, it states that when the rays emitted at the boundary of the light source, the edge rays, are mapped to the boundary of the target it is ensured that all other rays emitted by the light source are also mapped to the target, i.e. energy conservation holds.

However, Froese [32] discussed a simpler but related problem where a similar assumption to (3.3) holds true and she proposed to replace this assumption by a simpler boundary condition. Therefore we follow her strategy proposed in [32, Section 3.3] to render our boundary condition manageable.

The idea is to first replace the constraint (3.3), i.e., $T(U) = \Sigma$, by $T(\partial U) = \partial \Sigma$. Since we work on Ω , which is isomorphic to U , we write $T(\partial \Omega) = \partial \Sigma$. We then only need to prescribe the normal component of the mapping $T|_{\partial \Omega}$, such that we obtain the boundary condition

$$(4.18) \quad T(\cdot, u, Du)^T \nu(\cdot) = \phi(\cdot) \quad \text{in } \partial \Omega,$$

where ν is the outer normal vector of $\partial\Omega$ and the normal component ϕ of T is an a priori unknown function.

Unfortunately, we now have a circular dependency problem. If we knew ϕ , we would have the boundary condition in the desired form (4.2). Of course if the solution u^* of the problem and therefore the correct mapping T^* is known, then $\phi(\mathbf{x}) = T^*(\mathbf{x}, u^*, Du^*)^T \nu(\mathbf{x})$. But the function ϕ is unknown unless the solution of Problem IR is solved, where the boundary condition is needed to identify the solution.

In order to disrupt this circular dependency problem, a *Picard-type iteration* has been proposed in [32] for a similar but different problem: We iterate and start with an initial guess ϕ^0 . For $k = 1, 2, \dots$ we determine ϕ^k by first solving the Monge-Ampère equation (3.2) with boundary condition (4.18) using ϕ^{k-1} instead of ϕ . The solution u^{k-1} defines a reflector mapping $T^{k-1} := T(\cdot, u^{k-1}, Du^{k-1})$, which not necessarily maps $\partial\Omega$ onto $\partial\Sigma$ but onto $\partial\Sigma^{k-1}$ for the image Σ^{k-1} of the mapping T^{k-1} , which in general differs from Σ . In order to correct this we apply the orthogonal projection of $\partial\Sigma^{k-1}$ onto $\partial\Sigma$ using the standard scalar product and define

$$(4.19) \quad \phi^k(\mathbf{x}) := \left[\arg \min_{\mathbf{z} \in \partial\Sigma} \|\mathbf{z} - T^{k-1}(\mathbf{x}, u^{k-1}, Du^{k-1})\|_2^2 \right] \nu(\mathbf{x}) \quad \text{for } \mathbf{x} \in \partial\Omega.$$

The boundary condition then reads

$$\left[T(\mathbf{x}, u^k, Du^k) - \arg \min_{\mathbf{z} \in \partial\Sigma} \|\mathbf{z} - T(\mathbf{x}, u^{k-1}, Du^{k-1})\|_2^2 \right]^T \nu(\cdot) = 0 \quad \text{for } \mathbf{x} \in \partial\Omega,$$

where u^{k-1} is known and u^k is the desired interim solution in step k .

Existence. For the solution of the inverse reflector problem we need to ensure the conservation of energy (3.1). In order to obtain the correct ϕ , we solve the reflector problem for ϕ^k and therefore for a different target Σ^k for $k = 1, 2, \dots$ in the Picard-type iteration. The energy conservation then holds for Σ^k . Note that the prescribed density function g on the target Σ of Problem IR can be continued with zero outside of Σ . However, we need to compensate for the difference in energy and ensure the energy conservation condition to hold on Σ by scaling the density function f with an appropriate constant $c > 0$ defined by

$$c := \frac{\int_{\Sigma^k} g \, dS}{\int_{\Sigma} g \, dS};$$

see [32, Section 3.4]. Since we do not know Σ^k , also c is unknown. Thus, we introduce c as a new degree of freedom in our subproblems for the different right hand sides ϕ^k of the boundary condition (4.18) for $k = 1, 2, \dots$ and replace f in the Monge-Ampère equation (3.2) by cf which guarantees the existence of a solution.

Uniqueness. However, we cannot expect that there is only one R -convex solution, i.e., a solution of the Monge-Ampère equation (3.2) in the elliptic case, for each inverse reflector problem. In fact there are infinitely many solutions; see Subsection 3.3. One possible choice of a condition to ensure uniqueness is to fix the size of the reflector. The reflector is parameterized by the distance function u , which controls the size of its shape. Hence, similar to [32, Section 3.4], we fix a parameter $\mathcal{G} > 0$ and add the constraint

$$(4.20) \quad \int_{\Omega} u(\mathbf{x}) \, d\mathbf{x} = \mathcal{G}.$$

Resulting problem. Collecting all the conditions, we obtain the subproblems

$$\begin{aligned} \det_{\lambda}^{+} (D^2 u^k + A(\mathbf{x}, u^k, Du^k)) &= cb(\mathbf{x}, u^k, Du^k) && \text{for } \mathbf{x} \in \Omega, \\ T^k(\mathbf{x})^T \nu(\mathbf{x}) &= \phi^{k-1}(\mathbf{x}) && \text{for } \mathbf{x} \in \partial\Omega, \\ \int_{\Omega} u(\mathbf{x}) \, d\mathbf{x} &= \mathcal{G} \end{aligned}$$

for $k = 1, 2, \dots, k_{\max}$, where u^k and c are the unknowns, $\mathcal{G} > 0$ and $\lambda \geq 0$ are fixed parameters, b equals the right hand side of the Monge-Ampère equation (3.2), and ϕ^{k-1} is the orthogonal projection as defined in (4.19).

REMARK 6. *There is not plenty of existence and uniqueness theory available for equations of Monge-Ampère type, in particular not for the general case where the determinant does not only depend on the Hessian of u . The most adequate theorems for our situation found by the authors are formulated in [44, Theorem 1.1] and [64, Theorem 1.1] with Neumann boundary conditions. But unfortunately this cannot be applied because of the missing regularity of the boundary $\partial\Omega$ and the fact that our boundary conditions are nonlinear. However, it is worth noting that both theorems say that under some conditions there exists a unique solution of the elliptic Monge-Ampère type equation, so the ellipticity constraint (4.14) is important.*

5. Numerical simulations. In this section we present some test cases which we use to verify our numerical solver for the inverse reflector problem. First we consider in Subsection 5.1 five benchmark test cases for the Monge-Ampère equation of standard type (4.1). Since these have also been discussed by Froese and Oberman [33, 34], we can compare the convergence behavior of their and our methods. Afterwards we discuss results for the inverse reflector problem in Subsection 5.2.

General implementation remarks. All the computations have been carried out on a standard personal computer equipped with an AMD Phenom II X4 955 processor running at 3.2 GHz. In order to verify the numerical solutions of the inverse reflector problem we perform a forward simulation of the reflector using the ray tracing software POV-Ray [12].

Note that in each iteration of the nonlinear solver, i.e., in each Newton iteration, in our collocation method we solve a sparse system of linear equations. For this purpose we use the unsymmetric multifrontal sparse LU factorization package (UMFPACK) [16].

5.1. Test cases for the Monge-Ampère equation of standard type. We first define in Subsection 5.1.1 five test cases. The results are given in Subsection 5.1.2.

5.1.1. Five test cases. Let us define $\Omega := (0, 1) \times (0, 1) \subset \mathbb{R}^2$, $\mathbf{x} := (x, y)^T \in \Omega$, and $\mathbf{x}_0 := (\frac{1}{2}, \frac{1}{2})^T \in \Omega$. We consider five examples for the Monge-Ampère equation of standard type with Dirichlet boundary conditions, i.e.,

$$(5.1) \quad \det(D^2 u(\mathbf{x})) = f(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega \quad \text{and} \quad u(\mathbf{x}) = g(\mathbf{x}) \quad \text{for } \mathbf{x} \in \partial\Omega.$$

In the following examples the exact convex solution is known and the boundary function g is given as the restriction of the exact solution to the boundary $\partial\Omega$.

In the *first example* [4, 20, 33, 34] the solution is in $C^2(\Omega)$ and radially symmetric. The exact solution and the right hand side of the Monge-Ampère equation (5.1) are given by

$$(5.2) \quad u(\mathbf{x}) = \exp\left(\frac{\|\mathbf{x}\|_2^2}{2}\right) \quad \text{and} \quad f(\mathbf{x}) = (1 + \|\mathbf{x}\|_2^2) \exp\left(\frac{\|\mathbf{x}\|_2^2}{2}\right).$$

The *second example*, also taken from [33, 34], has a solution u in $C^1(\Omega)$ and is defined by

$$(5.3) \quad u(\mathbf{x}) = \frac{1}{2} (\max\{0, \|\mathbf{x} - \mathbf{x}_0\|_2 - 0.2\})^2 \quad \text{and} \quad f(\mathbf{x}) = \max\left\{0, 1 - \frac{0.2}{\|\mathbf{x} - \mathbf{x}_0\|_2}\right\}.$$

A solution which is in $C^2(\Omega)$ but whose gradient has a singularity near $(1, 1)^T \in \partial\Omega$ has also been discussed in [4, 19, 20, 33, 34]. This *third example* is given by the exact solution

$$(5.4) \quad u(\mathbf{x}) = -\sqrt{2 - \|\mathbf{x}\|_2^2} \quad \text{with right-hand side} \quad f(\mathbf{x}) = 2(2 - \|\mathbf{x}\|_2^2)^{-2}.$$

The solution u is also in $W^{1,p}(\Omega)$ for any $p \in [1, 4)$; see also [20].

In the *fourth example*, also taken from [4, 33, 34], the solution is only Lipschitz continuous, i.e., the solution u is in $C^{0,1}(\Omega)$, and is defined by

$$(5.5) \quad u(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_0\|_2 \quad \text{with right-hand side} \quad f = \pi\delta_{\mathbf{x}_0},$$

where f is defined by the Dirac delta distribution. Note that u is an Aleksandrov solution; see [33, 34] for the details. In [4, 33, 34] the distribution is approximated by a piecewise constant function. On a ball of radius $h/2$, where h is the spatial resolution of the grid, the approximation f_h takes a value such that integral over the ball is conserved. This leads to

$$f_h(\mathbf{x}) := \begin{cases} 4/h^2, & \text{for } \|\mathbf{x} - \mathbf{x}_0\|_2 \leq h/2, \\ 0, & \text{otherwise.} \end{cases}$$

For the *fifth and last example* the exact solution is unknown such that we can not compare the results with the exact solution. Nevertheless, Dean and Glowinski [18, 21, 22] and also Feng and Neilan [28] discussed this test case with right-hand side

$$(5.6) \quad f := 1$$

and Dirichlet homogeneous boundary condition. Feng and Neilan [28] remark that there exist a unique convex viscosity solution but no classical one.

REMARK 7. *Froese and Oberman [34] claim the solution of (5.5) to be $u(\mathbf{x}) = \sqrt{\|\mathbf{x} - \mathbf{x}_0\|_2}$ which is neither a convex function nor a solution to this right-hand side, but it seems that they used the correct version of (5.5) in their numerical experiments. This is probably a mistake in writing.*

REMARK 8. *The existence of a solution is guaranteed for the first four test cases due to their definitions. For the uniqueness we refer to the general result [35, Corollary 17.2] for classical solutions, which states for our case in (5.1) as follows:*

Let $\Omega \subset \mathbb{R}^2$ be a domain. Suppose $u, v \in C^0(\bar{\Omega}) \cap C^2(\Omega)$ are strictly convex and we have $\det(D^2u) = \det(D^2v)$ in Ω and $u = v$ on $\partial\Omega$. We then have $u \equiv v$ in Ω as well.

5.1.2. Results for the five test cases. We now apply our solver to the five test cases introduced in Subsection 5.1.1, that all impose Dirichlet boundary conditions. We still have to fix $\lambda \geq 0$ for the penalty term of our modified determinant \det_λ^\dagger in (4.17), which we use to ensure the ellipticity of the Monge-Ampère equation (5.1). Since this additional term vanishes for the exact solution we should use a large value. Preliminary tests reveal that $\lambda = 10^3$ is a good choice for all subsequent numerical

experiments. Our collocation grid is defined as an equidistant grid of $N \times N$ points for different values of $N \in \mathbb{N}$. Note that the grid points coincide with the knots of the B-splines.

For the first four test cases Benamou, Froese, and Oberman [4, 33, 34] suggest to use as initial guess the solution of the Poisson boundary value problem

$$(5.7) \quad \Delta u = \sqrt{2f}$$

with the same Dirichlet boundary conditions and for the same right hand side function f as for the corresponding Monge-Ampère equation. In [33, 34] this Poisson equation is solved in a preprocessing step and after that the result is convexified by the method of Oberman [54] to ensure a convex initial guess. Here we also use same initial guess for all five test cases, but, however, it turned out that our method works well even without convexifying the solution of (5.7). We therefore omit this step.

The spline collocation method has been used to solve (5.7). In order to validate the numerical result, we compare the maximum absolute error of the numerical solution u at the N^2 collocation points with the exact solution u^* . The absolute errors are given in Table 5.1, while the corresponding computing times are denoted in Table 5.2. Note that the computing time measurements indicate the overall time for the computation including the computation of the initial guess and the nested iteration scheme. In Figure 5.1 the dependency of the maximum error and the computing time on the number of unknowns are shown in a plot with logarithmic scale on both axes. We observe that the complexity of the solution method is proportional to N^3 .

Table 5.1: Maximum error $\|u - u^*\|_\infty$ for the first four test cases of Subsection 5.1.1.

N	C^2 example (5.2)	C^1 example (5.3)	example with blow up (5.4)	$C^{0,1}$ example (5.5)
31	$9.60 \cdot 10^{-5}$	$1.18 \cdot 10^{-4}$	$3.76 \cdot 10^{-3}$	$1.25 \cdot 10^{-2}$
45	$4.53 \cdot 10^{-5}$	$7.90 \cdot 10^{-5}$	$3.21 \cdot 10^{-3}$	$1.10 \cdot 10^{-2}$
63	$2.29 \cdot 10^{-5}$	$4.40 \cdot 10^{-5}$	$2.75 \cdot 10^{-3}$	$9.00 \cdot 10^{-3}$
89	$1.14 \cdot 10^{-5}$	$2.86 \cdot 10^{-5}$	$2.34 \cdot 10^{-3}$	$8.34 \cdot 10^{-3}$
127	$5.58 \cdot 10^{-6}$	$2.37 \cdot 10^{-5}$	$1.97 \cdot 10^{-3}$	$8.50 \cdot 10^{-3}$
181	$2.74 \cdot 10^{-6}$	$1.58 \cdot 10^{-5}$	$1.66 \cdot 10^{-3}$	$8.48 \cdot 10^{-3}$
255	$1.37 \cdot 10^{-6}$	$1.01 \cdot 10^{-5}$	$1.41 \cdot 10^{-3}$	$8.73 \cdot 10^{-3}$
361	$6.84 \cdot 10^{-7}$	$7.27 \cdot 10^{-6}$	$1.18 \cdot 10^{-3}$	$8.68 \cdot 10^{-3}$

Table 5.2: Computing time in seconds for the five test cases of Subsection 5.1.1 (the wall-clock time on the otherwise idle computer has been measured).

N	C^2 example (5.2)	C^1 example (5.3)	example with blow up (5.4)	$C^{0,1}$ example (5.5)	Viscosity solution (5.6)
31	0.1	0.3	0.1	4.8	0.1
45	0.1	0.6	0.2	13.3	0.2
63	0.2	1.2	0.4	18.9	0.3
89	0.6	3.2	0.9	58.6	0.9
127	0.9	4.1	1.7	97.6	1.7
181	2.7	12.8	4.9	345.2	5.1
255	5.1	24.9	9.7	639.5	9.9
361	18.4	79.8	31.6	2193.7	32.9

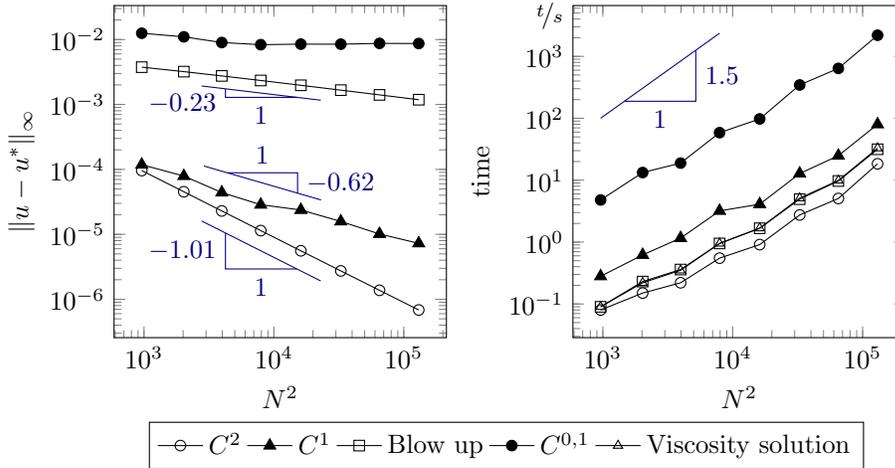


Figure 5.1: Plot of the maximum errors and computing times for the test cases of Subsection 5.1.1 (see also Table 5.1 and Table 5.2).

For a better comparison the grid sizes given by N were chosen to match the choices of Froese and Oberman [34]. In that paper three of their methods [4, 33, 34] are compared for the first four test cases, where it turned out that the standard finite difference method [4] performed best for the first example. Using regression analysis for our method we observe that the curve in the double-logarithmic plot has a slope of -1.01 which corresponds to a quadratic convergence rate. This rate agrees with that achieved by the finite difference scheme. Moreover, the differences in the maximum errors is less than a factor of $3/2$.

For the little less smooth solution in the second example the standard finite difference method proposed in [4] still leads to smaller errors than the two methods in [33, 34]. Comparing the absolute errors our method improves the results of the standard finite difference method by a factor of 2 to 3 and we observe a slightly superlinear convergence rate.

The third example is a big challenge for the methods because of the blow up of the gradient of the solution at the point $(1, 1)^T \in \partial\Omega$. Here the monotone scheme [33] and the hybrid scheme [34] perform best. Our method shows a convergence that is approximately proportional to the square root of the mesh size. It works more precise than the standard finite difference method by a factor of about 4 but is not as accurate as the other two schemes whose maximal errors are between $2 \cdot 10^{-3}$ and $4 \cdot 10^{-5}$. Due to the fact that the schemes [33, 34] are constructed to converge also to viscosity solutions these methods are suited for less smooth solutions.

The solution of the fourth example does not have a continuous first derivative in Ω such that it is very difficult to handle even for standard spline interpolation. Here all three methods given in [4, 33, 34] and also our method do not converge, the error does not drop below 10^{-3} . Interestingly, we observe that all three methods, the two methods of [33, 34], and our method, stagnate for N larger than 89. In fact, we do not even expect that our collocation method converges, because it requires the solution to be twice differentiable.

In Figure 5.2 we visualize the result for the fifth test case. In fact, the solution is convex. Figures 5.2(b) and 5.2(c) show cross section of the solution along the x -axis and along the diagonal, respectively. These can be compared with those of Dean and Glowinski [18, 21, 22] and Feng and Neilan [28]. We observe that both the curvature as well as the minimal values of the functions agree with those in the literature.

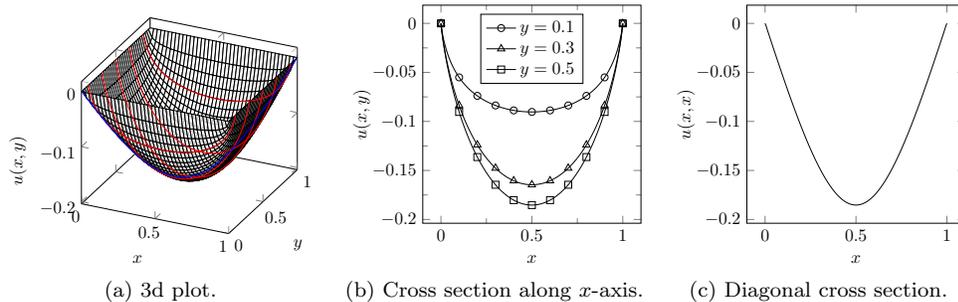


Figure 5.2: Solution of the fifth example computed for $N = 181$.

5.2. Test cases for the inverse reflector problem. First, in Subsection 5.2.1 we give some additional details that are crucial for numerically solving the inverse reflector problem. Then in Subsection 5.2.2 we define the geometric setting of our test case for the inverse reflector problem. Afterwards we describe in Subsection 5.2.3 how we obtain a good initial guess and in Subsection 5.2.4 we present the results for some examples.

5.2.1. Solution procedure. We now briefly discuss the procedure of numerically solving the inverse reflector problem. To this end we focus on three issues that particularly need to be handled to successfully solve the inverse reflector problem.

Boundary condition. In Subsection 3.2 we saw the mathematical formulation of the reflector problem for the near field and in Subsection 4.5 a relaxation to a sequence of subproblems. For each subproblem we have to solve an equation of Monge-Ampère type with an adjusted boundary function ϕ . Moreover, we use an iterative nonlinear solver for the solution of each subproblem. To avoid solving the Monge-Ampère equation many times, we intertwine the iterations and update the boundary function ϕ immediately after each correction step in the iterative nonlinear solver instead of not updating ϕ until the nonlinear solver terminates.

Since we work on a rectangular target set Σ we face the problem that the outer normal vector of $\partial\Sigma$ is not defined in a corner. Here we use the normalized sum vector of the two outer normal vectors of both adjacent edges, i.e., the outer normal vectors at the corners of a rectangle $(a, b) \times (c, d)$ are given by the vectors $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2})^T$.

Dark areas on the target. Let the density function g for our target illumination on Σ be given by 8 bit digital grayscale images, i.e. the gray values of the image are integer values in the range $0, \dots, 255$. Since g is in the denominator on the right hand side of the Monge-Ampère equation (3.2), we require that g is bounded away from zero. This lower bound should be as small as possible. To ensure this constraint we adjust brightness and contrast of the input image g and consider the image

$$(5.8) \quad \tilde{g}(\mathbf{Z}) := g(\mathbf{Z}) + \max\{0, 20 - \min_{\mathbf{Z}' \in \Sigma} g(\mathbf{Z}')\}$$

instead, where the value of 20 leads to good results. Afterwards this density function needs to be normalized such that the energy conservation (3.1) holds true.

However, if we try to realize black values in the target image by letting some pixel values go to zero, the right-hand side of the Monge-Ampère equation (3.2) tends to infinity. The left-hand side is more or less the determinant of the Hessian of u ; see, e.g., the special case in equation (3.4), which also needs to go to infinity at some points because of the surjectivity constraint (3.3). It follows that the curvature of the reflector must be infinitely large at these points which leads to a kink on its surface.

Nested iteration. On the one hand our nonlinear solver profits from a good choice of the initial guess. But on the other hand, if we want to produce a very complex image on the target, we need to define our ansatz functions on a very fine grid. Thus we have many degrees of freedom which makes it difficult to obtain a good initial guess.

An efficient way to address this problem is to apply the multilevel technique of Subsection 4.3. We therefore start the computation on a very coarse grid of dimension 21×21 and solve the inverse reflector problem. Our target density function g will be given by an image of size 512×512 pixels. Thus we cannot expect to be able to solve the reflector problem accurately on such a coarse grid. We therefore also coarsen the image. For this reason we define the *standard mollifier function* $\varphi(\mathbf{x}) := \exp(-1/(1-\|\mathbf{x}\|_2^2))$ if $\|\mathbf{x}\|_2 < 1$ and zero otherwise. A discrete approximation of the mollifier function with a support of size $n \times n$ pixels is given by

$$(5.9) \quad \varphi_n(i, j) := \frac{\varphi\left(2\frac{i}{n}, 2\frac{j}{n}\right)}{\sum_{r, s \in \mathbb{Z}} \varphi\left(2\frac{r}{n}, 2\frac{s}{n}\right)}$$

for $i, j \in \mathbb{Z}$. We now convolve the image with φ_n for different $n \in \mathbb{N}$ and solve the problem for these modified images on grids of size $N \times N$ for appropriate $N \in \mathbb{N}$, i.e., we solve the problem many times for different pairs (N, n) to improve the solution; see Figure 5.3. We use the following pairs in the given order: $(21, 55)$, $(41, 55)$, $(41, 19)$, $(81, 19)$, $(81, 7)$, $(161, 7)$, $(161, 3)$, $(321, 3)$.

5.2.2. Optical and geometric setting. For the illumination of the mirror we choose an isotropic light source, i.e., $f \equiv 1$. Our geometrical setting is given by the target surface, defined by $\Sigma := \{\mathbf{Z} = (z_1, z_2, z_3)^T \in \mathbb{R}^3 : z_1 \in (-1.5, 1.5), z_2 \in (1, 4), z_3 = -5\}$ and by the position of the reflector, which is given in the dimensioned drawing in Figure 5.4(a). In the mathematical model, the size of the reflector is controlled by an appropriate constant \mathcal{G} in (4.20) which in the following examples is $\mathcal{G} = 0.417674$. For the modified determinant (4.16) we again choose the penalty constant $\lambda = 10^3$, which leads to good results for all of our examples.

5.2.3. Initialization. The choice of the initial guess is crucial for the convergence of the Newton-type scheme in the collocation method. Therefore we need an initial guess that is close enough to the solution. Since there are already other methods available to solve the inverse reflector problem, we can resort to one of them for the initialization. Due to the nested iteration approach we only need to calculate an initial guess for a strongly blurred input density distribution on a very coarse initial grid. For that reason we choose the method of supporting ellipsoids [40, 41] for this task, which is a viable choice, because of the low resolution the complexity of the method is not too high; see also Section 2.2.

In principle, we need to generate a new initial guess when the desired density function g changes. However, numerical evidence shows that this is not necessary

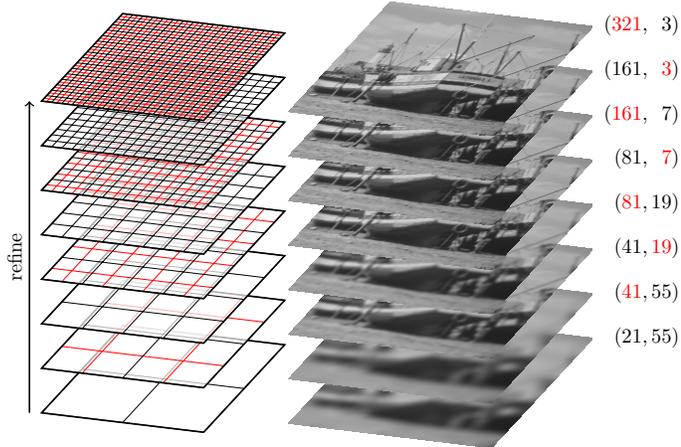


Figure 5.3: Nested iterations to improve the initial guess from bottom to top with given pairs (N, n) , where $N \times N$ is the resolution of the grid and n the parameter for the mollifier φ_n in (5.9).

and that we can prepare an universal initial guess that depends on the optical and geometric setup but no longer on g . This is probably possible, because the collocation method starts on a very coarse grid using a strongly blurred and thus an “almost” constant version of g .

We therefore invoke the method of supporting ellipsoids and compute a reflector surface that produces a constant density function g on the target. The solution specifies a surface that consists of segments of ellipsoids of revolution. Next we approximate the solution in our B-spline ansatz space corresponding to the coarse initial grid using spline interpolation and we solve the inverse reflector problem again on the same grid with our spline collocation method. The output of the forward simulation of the resulting reflector is shown in Figure 5.4(b). As desired the reflector produces a homogeneous illumination pattern on the target. In the following we use this reflector as the initial guess for all calculations in the same geometrical and optical setting but for different target illuminations g .

5.2.4. Results for the inverse reflector problem. We now calculate the reflector surfaces for three test cases, where the desired target illuminations g are given by three common grayscale test images from the USC-SIPI Image Database [65]. In a post-processing step we run the forward simulation by ray tracing to compute the actual illumination pattern produced on the target by the designed reflector surface. Figure 5.5 shows the simulation results. Each of the three output images is very close to the corresponding original image and, although the images are slightly blurred, even complex details can easily be identified.

Note that the differences in the illumination between dark areas in the original pictures and the simulations are resulting from the fact that we have to lift dark gray values up; see (5.8). Therefore it is impossible to produce real black areas on the target.

The three test images pose different challenges for our reflector design algorithm. In the first test image *Boat*, see Figure 5.5(a), the problem is to meet the straight

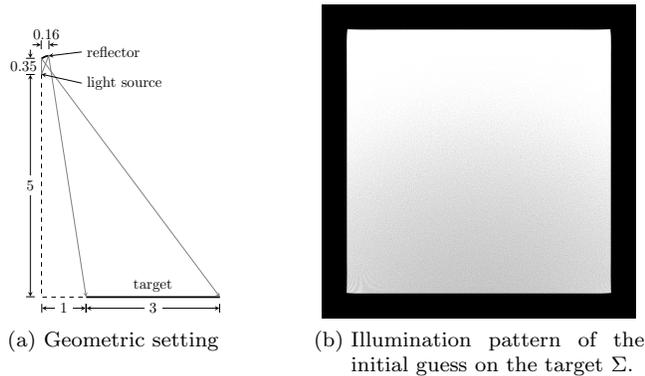


Figure 5.4: Setting of the problem and initial guess.

lines of the mast, the person standing next to the boat and the lettering on the stern of the vessel. The simulation result looks very good, straight lines are depicted almost perfectly in all directions and the name of the vessel is still readable but only barely. In our selection the second image *Goldhill*, see Figure 5.5(b), represents a different type of pictures. It is rich of different patterns, e.g. the patterns of the roofing tiles and the windows in the foreground as well as the patterns of the trees and bushes in the background, such that we can test how well different patterns are reproduced. Apart from the slight blurring effect the different patterns are well depicted and can be distinguished easily. The challenge of the image *Mandrill* is to depict the hair of the beard of the monkey. We can see in Figure 5.5(c) that our algorithm also passes this test.

As our final numerical example we compute the surface of a mirror that projects our institute's logo on the screen. Note that this type of cartoon-like images with high contrast and sharp edges is most challenging for our algorithm. Nevertheless, our algorithm achieves a very good reproduction of the logo; see Figure 5.6(a) for the desired intensity pattern and Figure 5.6(b) for the forward simulation result by ray tracing. In Figure 5.6(c) we show the position and the coarse shape of the mirror surface, while the fine structure that contains the information of the image is visualized in Figure 5.6(d) after a highpass filtering process. Note that lighter areas on the screen correspond to large areas on the mirror surface.

6. Conclusion and outlook. We have presented a new B-spline collocation method for the numerical solution of Monge-Ampère type equations, which are strongly nonlinear partial differential equations. Some extensions and manipulations of the equations and boundary conditions have been explained in detail that render it possible to apply the collocation method to the solution of the inverse reflector problem formulated as a Monge-Ampère type equation.

In comparison with existing schemes developed in [4, 33, 34] our B-spline collocation method produces results that are similar to the finite difference scheme in [4], which is the most accurate method for smooth solutions proposed in these publications.

The largest obstructions encountered for the numerical solution of the inverse reflector problem are how to handle the boundary conditions, how to ensure the



Figure 5.5: Simulation results for three test images. First row: desired distribution (original image); second row: distribution after forward simulation by ray tracing).

uniqueness of the solution, and how to achieve convergence in the numerical scheme. In fact, we explain how these issues can be resolved and that the B-spline collocation method is well-suited even for this strongly nonlinear problem.

In future work the authors plan to extend the numerical method to support other optical devices, in particular lenses. The problem is then, for example, to determine the two surfaces of a lens such that all light passing through this lens is redirected onto the target and produces a prescribed illumination pattern. This is a strongly related problem and can also be modeled by a Monge-Ampère type equation; see, e.g., [37].

In modern lighting applications sources can often no longer be assumed to be point sources, e.g., in compact optical systems using light emitting diodes (LEDs). Therefore the challenging question arises how to handle extended light sources in the inverse reflector problem.

Acknowledgment. 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.

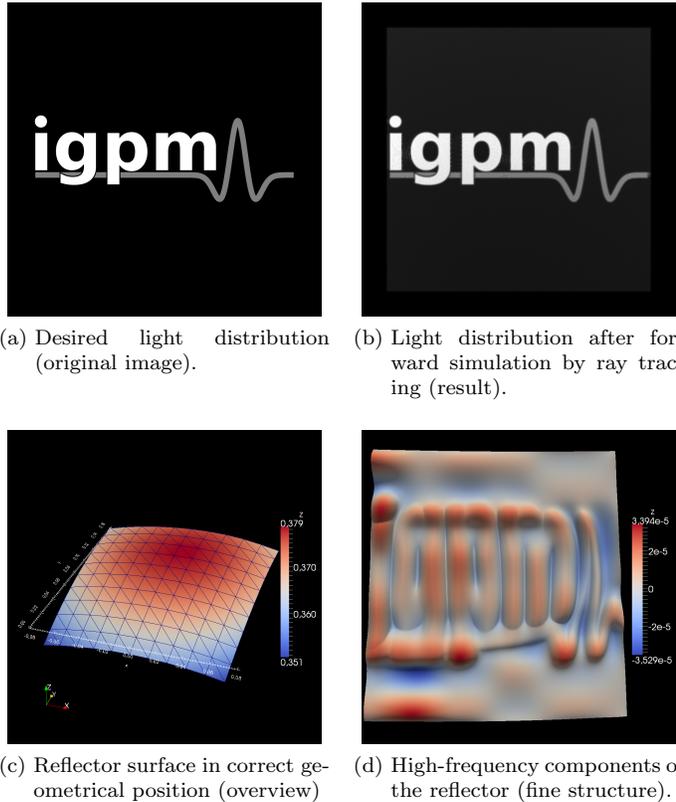


Figure 5.6: Simulation results for our institute's logo.

REFERENCES

- [1] A.-M. AMPÈRE, *Mémoire concernant l'application de la théorie exposée dans le XVIIe cahier du journal de l'école polytechnique, à l'intégration des équations aux différentielles partielles du premier et du second ordre.*, J. École. R. Polytech., 11 (1820), pp. 1–188. <http://gallica.bnf.fr/ark:/12148/bpt6k4336744/f2>.
- [2] O. ANSON, J. F. SERON, AND D. GUTIERREZ, *NURBS-based inverse reflector design*, in Proceedings of Congreso Español de Informática Gráfica (CEIG) 2008, Barcelona, Spain, 2008, pp. 65–74. DOI: 10.2312/LocalChapterEvents/CEIG/CEIG08/065-074.
- [3] G. AWANOU, *Spline element method for the Monge-Ampère equation*. [arXiv:1012.1775](https://arxiv.org/abs/1012.1775) [math.NA], 2010.
- [4] J.-D. BENAMOU, B. D. FROESE, AND A. M. OBERMAN, *Two numerical methods for the elliptic Monge-Ampère equation*, ESAIM Math. Model. Numer. Anal., 44 (2010), pp. 737–758. DOI: 10.1051/m2an/2010017.
- [5] P. BENÍTEZ, J. C. MIÑANO, J. BLEN, R. MOHEDANO, J. CHAVES, O. DROSS, M. HERNÁNDEZ, J. L. ALVAREZ, AND W. FALICOFF, *SMS design method in 3d geometry: examples and applications*, in Proceedings of SPIE, vol. 5185 of Nonimaging Optics: Maximum Efficiency Light Transfer VII, 2004, pp. 18–29. DOI: 10.1117/12.506857.
- [6] P. BENÍTEZ, J. C. MIÑANO, J. BLEN, R. MOHEDANO, J. CHAVES, O. DROSS, M. HERNÁNDEZ, AND W. FALICOFF, *Simultaneous multiple surface optical design method in three dimensions*, Opt. Engrg., 43 (2004), pp. 1489–1502. DOI: 10.1117/1.1752918.
- [7] M. V. BERRY, *Oriental magic mirrors and the Laplacian image*, European J. Phys., 27 (2006), pp. 109–118. DOI: 10.1088/0143-0807/27/1/012.
- [8] S. C. BRENNER, T. GUD, M. NEILAN, AND L.-Y. SUNG, *C^0 penalty methods for the fully nonlin-*

- ear Monge-Ampère equation, *Math. Comp.*, 80 (2011), pp. 1979–1995. DOI: 10.1090/S0025-5718-2011-02487-7.
- [9] S. C. BRENNER AND M. NEILAN, *Finite element approximations of the three dimensional Monge-Ampère equation*, *ESAIM Math. Model. Numer. Anal.*, 46 (2012), pp. 979–1001. DOI: 10.1051/m2an/2011067.
- [10] F. BRICKELL, L. MARDER, AND B. S. WESTCOTT, *The geometrical optics design of reflectors using complex coordinates*, *J. Phys. A*, 10 (1977), pp. 245–260. DOI: 10.1088/0305-4470/10/2/014.
- [11] L. A. CAFFARELLI, S. A. KOCHENGIN, AND V. I. OLIKER, *On the numerical solution of the problem of reflector design with given far-filed scattering data*, in *Monge Ampère Equation: Applications to Geometry and Optimization*. Proceedings of the NSF-CBMS conference, Deerfield Beach, FL, USA, July 9–13, 1997, L. A. Caffarelli and M. Milman, eds., vol. 226 of *Contemporary Mathematics*, American Mathematical Society, Providence, RI, 1999, pp. 13–32. DOI: 10.1090/conm/226.
- [12] C. CASON, T. FROELICH, N. KOPP, R. PARKER, ET AL., *POV-Ray*. <http://www.povray.org>, 1991.
- [13] W. J. CASSARLY, *Iterative reflector design using a cumulative flux compensation approach*, in *International Optical Design Conference, OSA Technical Digest, Optical Society of America, Design of Illumination Systems: Optimization and Tolerancing Approaches (ITHA)*, 2010. DOI: 10.1364/IODC.2010.ITHA2.
- [14] J. CHAVES, *Introduction to Nonimaging Optics*, vol. 134 of *Optical Science and Engineering*, CRC Press, Boca Raton, FL, 2008. DOI: 10.1201/9781420054323.
- [15] W. DAHMEN, *B-splines in analysis, algebra and applications*, *Trav. Math.*, 10 (1998), pp. 15–76.
- [16] T. A. DAVIS AND S. I. DUFF, *An unsymmetric-pattern multifrontal method for sparse LU factorization*, *SIAM J. Math. Anal. Appl.*, 18 (1997), pp. 140–158. DOI: 10.1137/S0895479894246905.
- [17] C. DE BOOR, *A Practical Guide to Splines*, vol. 27 of *Applied Mathematical Sciences*, Springer, Heidelberg, 1978.
- [18] E. J. DEAN AND R. GLOWINSKI, *Numerical solution of the two-dimensional elliptic Monge-Ampère equation with Dirichlet boundary conditions: An augmented Lagrangian approach*, *C. R. Math. Acad. Sci. Paris*, 336 (2003), pp. 779–784. DOI: 10.1016/S1631-073X(03)00149-3.
- [19] E. J. DEAN AND R. GLOWINSKI, *Numerical solution of the two-dimensional elliptic Monge-Ampère equation with Dirichlet boundary conditions: a least-squares approach*, *C. R. Math. Acad. Sci. Paris*, 339 (2004), pp. 887–892. DOI: 10.1016/j.crma.2004.09.018.
- [20] E. J. DEAN AND R. GLOWINSKI, *An augmented Lagrangian approach to the numerical solution of the Dirichlet problem for the elliptic Monge-Ampère equation in two dimensions*, *Electron. Trans. Numer. Anal.*, 22 (2006), pp. 71–96. EuDML: eudml.org/doc/127457.
- [21] E. J. DEAN AND R. GLOWINSKI, *Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type*, *Comput. Methods Appl. Mech. Engrg.*, 195 (2006), pp. 1344–1386. DOI: 10.1016/j.cma.2005.05.023.
- [22] E. J. DEAN AND R. GLOWINSKI, *On the numerical solution of the elliptic Monge-Ampère equation in dimension two: a least-squares approach*, in *Partial differential equations*, R. Glowinski and P. Neittaanmäki, eds., vol. 16 of *Computational Methods in Applied Sciences*, Springer, Heidelberg, 2008, pp. 43–63.
- [23] J. E. DENNIS AND H. H. W. MEI, *Two new unconstrained optimization algorithms which use function and gradient values*, *J. Optim. Theory Appl.*, 28 (1979), pp. 453–482. DOI: 10.1007/BF00932218.
- [24] J. E. DENNIS AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall series in *Computational Mathematics*, Prentice-Hall, Englewood Cliffs, NJ, 1983. DOI: 10.1137/1.9781611971200.
- [25] H. W. ENGL AND A. NEUBAUER, *Reflector design as an inverse problem*, in *Proceedings of the Fifth European Conference on Mathematics in Industry*, M. Heiliö, ed., vol. 7 of *European Consortium for Mathematics in Industry*, B. G. Teubner, Stuttgart, 1991, pp. 13–24.
- [26] X. FENG, R. GLOWINSKI, AND M. NEILAN, *Recent developments in numerical methods for fully nonlinear second order partial differential equations*, *SIAM Rev.*, 55 (2013), pp. 205–267. DOI: 10.1137/110825960.
- [27] X. FENG AND M. NEILAN, *Mixed finite element methods for the fully nonlinear Monge-Ampère equation based on the vanishing moment method*, *SIAM J. Numer. Anal.*, 47 (2009), pp. 1226–1250. DOI: 10.1137/070710378.
- [28] X. FENG AND M. NEILAN, *Vanishing moment method and moment solutions for fully nonlinear second order partial differential equations*, *J. Sci. Comput.*, 38 (2009), pp. 74–98. DOI:

- 10.1007/s10915-008-9221-9.
- [29] X. FENG AND M. NEILAN, *Analysis of Galerkin methods for the fully nonlinear Monge-Ampère equation*, J. Sci. Comput., 47 (2011), pp. 303–327. DOI: 10.1007/s10915-010-9439-1.
- [30] M. FINCKH, H. DAMMERTZ, AND H. P. A. LENSCH, *Geometry construction from caustic images*, in Computer Vision – ECCV 2010, K. Daniilidis, P. Maragos, and N. Paragios, eds., vol. 6315 of Lecture Notes in Computer Science, Springer, Heidelberg, 2010, pp. 464–477. DOI: 10.1007/978-3-642-15555-0_34.
- [31] F. R. FOURNIER, W. J. CASSARLY, AND J. P. ROLLAND, *Optimization of single reflectors for extended sources*, in Proceedings of SPIE, no. 71030I in Illumination Optics, 2008. DOI: 10.1117/12.800377.
- [32] B. D. FROESE, *A numerical method for the elliptic Monge-Ampère equation with transport boundary conditions*, SIAM J. Sci. Comput., 34 (2012), pp. A1432–A1459. DOI: 10.1137/110822372.
- [33] B. D. FROESE AND A. M. OBERMAN, *Convergent finite difference solvers for viscosity solutions of the elliptic Monge-Ampère equation in dimensions two and higher*, SIAM J. Numer. Anal., 49 (2011), pp. 1692–1714. DOI: 10.1137/100803092.
- [34] B. D. FROESE AND A. M. OBERMAN, *Fast finite difference solvers for singular solutions of the elliptic Monge-Ampère equation*, J. Comput. Phys., 230 (2011), pp. 818–834. DOI: 10.1016/j.jcp.2010.10.020.
- [35] D. GILBARG AND N. S. TRUDINGER, *Elliptic partial differential equations of second order*, vol. 224 of Grundlehren der Mathematischen Wissenschaften, Springer, Heidelberg, 2nd ed., 1983.
- [36] C. E. GUTIÉRREZ, *The Monge–Ampère Equation*, vol. 44 of Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, Basel, 2001.
- [37] C. E. GUTIÉRREZ, *Fully Nonlinear PDEs in Real and Complex Geometry and Optics*, vol. 2087 of Lecture Notes in Mathematics, Springer, Heidelberg, 2014, ch. Refraction Problems in Geometric Optics, pp. 95–150.
- [38] K. HÖLLIG, U. REIF, AND J. WIPPER, *Weighted extended B-spline approximation of Dirichlet problems*, SIAM J. Numer. Anal., 39 (2001), pp. 442–462. DOI: 10.1137/S0036142900373208.
- [39] A. KARAKHANYAN AND X.-J. WANG, *On the reflector shape design*, J. Differential Geom., 84 (2010), pp. 561–610. <http://projecteuclid.org/euclid.jdg/1279114301>.
- [40] S. A. KOCHENGIN AND V. I. OLIKER, *Determination of reflector surfaces from near-field scattering data*, Inverse Problems, 13 (1997), pp. 363–373. DOI: 10.1088/0266-5611/13/2/011.
- [41] S. A. KOCHENGIN AND V. I. OLIKER, *Determination of reflector surfaces from near-field scattering data. II: Numerical solution*, Numer. Math., 79 (1998), pp. 553–568. DOI: 10.1007/s002110050351.
- [42] S. A. KOCHENGIN AND V. I. OLIKER, *Computational algorithms for constructing reflectors*, Comput. Vis. Sci., 6 (2003), pp. 15–21. DOI: 10.1007/s00791-003-0103-2.
- [43] M. KURZ, D. OBERSCHMIDT, N. SIEDOW, R. FESSLER, AND J. JEGOROV, *Mit schnellem Algorithmus zur perfekten Freiformoptik*, Mikroproduktion, 3 (2009), pp. 10–12.
- [44] P.-L. LIONS, N. S. TRUDINGER, AND J. I. E. URBAS, *The Neumann problem for equations of Monge-Ampère type*, Comm. Pure Appl. Math., 39 (1986), pp. 539–563. DOI: 10.1002/cpa.3160390405.
- [45] K. MADSEN, H. B. NIELSEN, AND O. TINGLEFF, *Methods for non-linear least squares problems*, technical report, Department of Mathematical Modelling, Danmarks Tekniske Universitet, Copenhagen, 2004. http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3215/pdf/imm3215.pdf.
- [46] S.-Y. MAK AND D.-Y. YIP, *Secrets of the chinese magic mirror replica*, Phys. Ed., 36 (2001), pp. 102–107. DOI: 10.1088/0031-9120/36/2/302.
- [47] A. MAS, I. MARTÍN, AND G. PATOW, *Fast inverse reflector design (FIRD)*, Computer Graphics Forum, 28 (2009), pp. 2046–2056. DOI: 10.1111/j.1467-8659.2009.01430.x.
- [48] J. C. MIÑANO, P. BENÍTEZ, W. LIN, J. INFANTE, F. MUÑOZ, AND A. SANTAMARÍA, *An application of the SMS method for imaging designs*, Optics Express, 17 (2009), pp. 24036–24044. DOI: 10.1364/OE.17.024036.
- [49] J. C. MIÑANO AND J. C. GONZÁLEZ, *New method of design of nonimaging concentrators*, Appl. Opt., 31 (1992), pp. 3051–3060. DOI: 10.1364/AO.31.003051.
- [50] G. MONGE, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l’Académie Royale des Sciences de Paris, (1781), pp. 666–704. <http://gallica.bnf.fr/ark:/12148/bpt6k35800/f796.image>.
- [51] F. MUÑOZ, P. BENÍTEZ, O. DROSS, J. C. MIÑANO, AND W. A. PARKYN, *Simultaneous multiple surface design of compact air-gap collimators for light-emitting diodes*, Opt. Engrg., 43

- (2004), pp. 1522–1530. DOI: 10.1117/1.1753588.
- [52] A. NEUBAUER, *Design of 3d-reflectors for near field and far field problems*, in Large Scale Optimization with Applications. Part I: Optimization in Inverse Problems and Design, L. T. Biegler, T. F. Coleman, A. R. Conn, and F. N. Santosa, eds., vol. 92 of The IMA Volumes in Mathematics and its Applications, Springer, Heidelberg, 1997, pp. 101–118. DOI: 10.1007/978-1-4612-1962-0_6.
- [53] H. B. NIELSEN, *Damping parameter in Marquardt’s method*, technical report imm-rep-1999-05, Department of Mathematical Modelling, Danmarks Tekniske Universitet, Copenhagen, 1999. <http://www2.imm.dtu.dk/pubdb/p.php?648>.
- [54] A. M. OBERMAN, *Computing the convex envelope using a nonlinear partial differential equation*, Math. Models Methods Appl. Sci., 18 (2008), pp. 759–780. DOI: 10.1142/S0218202508002851.
- [55] A. M. OBERMAN, *Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the Hessian*, Discrete Contin. Dyn. Syst. Ser. B, 10 (2008), pp. 221–238. DOI: 10.3934/dcdsb.2008.10.221.
- [56] V. I. OLIKER, *A rigorous method for synthesis of offset shaped reflector antennas*, Computing Letters, 2 (2006), pp. 29–49. DOI: 10.1163/157404006777491981.
- [57] G. PATOW AND X. PUEYO, *A survey of inverse surface design from light transport behavior specification.*, Computer Graphics Forum, 24 (2005), pp. 773–789. DOI: 10.1111/j.1467-8659.2005.00901.x.
- [58] L. PIEGL AND W. TILLER, *The NURBS Book*, Springer, Heidelberg, 2. ed., 1997.
- [59] H. RIES AND J. A. MUSCHAWECK, *Tailored freeform optical surfaces*, J. Opt. Soc. Amer. A, 19 (2002), pp. 590–595. DOI: 10.1364/JOSAA.19.000590.
- [60] N. SAVAGE, *Optical design software*, Nature Photonics, 1 (2007), pp. 598–599. DOI: 10.1038/nphoton.2007.190.
- [61] J. S. SCHRUBEN, *Formulation of a reflector-design problem for a lighting fixture*, J. Opt. Soc. Amer. A, 62 (1972), pp. 1498–1501. DOI: 10.1364/JOSA.62.001498.
- [62] G. STRANG, *Linear Algebra and Its Applications*, Academic Press, New York, 2nd ed., 1980.
- [63] N. S. TRUDINGER AND X.-J. WANG, *On the second boundary value problem for Monge-Ampère type equations and optimal transportation*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 8 (2009), pp. 143–174. DOI: 10.2422/2036-2145.2009.1.07.
- [64] J. URBAS, *Oblique boundary value problems for equations of Monge-Ampère type*, Calc. Var. Partial Differential Equations, 7 (1998), pp. 19–39. DOI: 10.1007/s005260050097.
- [65] USC-SIPI, *USC-SIPI image database*. <http://sipi.usc.edu/database/>.
- [66] C. VILLANI, *Topics in Optimal Transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.
- [67] X.-J. WANG, *On the design of a reflector antenna*, Inverse Problems, 12 (1996), pp. 351–375. DOI: 10.1088/0266-5611/12/3/013.
- [68] X.-J. WANG, *On the design of a reflector antenna II*, Calc. Var. Partial Differential Equations, 20 (2004), pp. 329–341. DOI: 10.1007/s00526-003-0239-4.
- [69] T. WEYRICH, P. PEERS, W. MATUSIK, AND S. RUSINKIEWICZ, *Fabricating microgeometry for custom surface reflectance*, ACM Trans. Graphics, 28 (2009), pp. 32:1–6. DOI: 10.1145/1531326.1531338.
- [70] R. WINSTON, J. C. MIÑANO, AND P. BENÍTEZ, *Nonimaging Optics*, Academic Press, New York, 2005.
- [71] X. ZHU, J. NI, AND Q. CHEN, *An optical design and simulation of LED low-beam headlamps*, J. Phys. Conf. Ser., 276 (2011), p. 012201. DOI: 10.1088/1742-6596/276/1/012201.
- [72] S. ZWICK, R. FESSLER, J. JEGOROV, AND G. NOTNI, *Resolution limitations for tailored picture-generating freeform surfaces*, Optics Express, 20 (2012), pp. 3642–3653. DOI: 10.1364/OE.20.003642.