Mathematical model for a freeform optical system

Point source to point target with two reflectors

Master internship
Industrial and Applied Mathematics

Hannah van Susteren
0993670

Supervisors:
J.H.M. ten Thije Boonkkamp
W.L. IJzerman

November 30, 2021
Eindhoven University of Technology
Department of Mathematics and Computer Science

Contents

1 Introduction

The importance of energy conservation is a topic which is greatly discussed. Energy conservation is important to battle climate change, as discussed in the United Nations Climate Change Conference, also known as COP26 [4]. The importance of energy conservation w.r.t. light has also been emphasize by Signify at COP26 [3]. Furthermore, energy conservation can have economical benefits as the energy prices are skyrocketing [2]. The U.S. Energy Information Administration estimates that 8% of the total residential and commercial electricity consumption is used for lightning [1]. Thus, the importance of efficient lighting is growing and as a result, more research on lightning is is carried out, including this report.

Optical systems can be studied using either forward methods or inverse methods. In forward methods, the target distribution is computed from the source distribution and the mapping of the optical system. A well-known forward method is Monte-Carlo ray-tracing [5]. However, this type of method has the disadvantage that it often relies on trial-and-error. That is, a target distribution is computed. If the result is not desired, the optical system is altered and the computation is repeated. This can be a long and tedious process if high precision results are needed.

Alternatively, there are inverse methods. These methods are based on a given source and target distribution. Subsequently, the mapping of the optical system is computed. This is a direct solution method. In the inverse method that we consider, geometrical optics is used. In geometrical optics, light propagation is approximated using rays which only travel in straight paths. This approximation is based on the fact that the wavelength is minuscule compared to the distances the light travels. Furthermore, we consider conservation of energy. That is, all light emitted by the source should arrive at the target. Using the conservation of energy and the formulation of an optical system, a partial differential equation can be derived for the optical surface, which can be either a lens or a reflector. This partial differential equation is known as the generalized Monge-Ampere equation.

In the derivation of the Monge-Ampere equation, optimal transport can be used. In optimal transport problems, a certain quantity has to be moved from one location to a different location. The goal is to find the mapping that moves the quantity while minimizing the effort or costs it takes to move the quantity. Using theory from optical transport, conditions can be found for finding a mapping. Furthermore, optimal transport can be used to formulate a cost function. This function relates optical surfaces and or Hamiltonian functions and is used in the derivation of the Monge-Ampère equation.

In order to solve the generalized Monge-Ampère equation, a numerical method is needed. The method that we use is the generalized least squares method. Alongside the Monge-Ampère equation, we impose a transport boundary condition. Subsequently, this system of equations can be solved using an iterative three-step minimization process. The process includes the minimization of three functionals that are used to impose the Monge-Ampere equation, impose the transport boundary condition and lastly compute the mapping.

In section 2 of this report, we formulate the mathematical model of our optical system. We start with a description of the geometry and introduce the variables that are needed. Subsequently, we derive the optical path length and the cost function. Next we introduce a transformation of coordinates from Cartesian coordinates to stereographic coordinates. We briefly discuss the mapping. Afterwards, we introduce energy conservation that is used to derive the Monge-Ampère equation. Lastly, we explain how optimal transport can be used to formulate problems in illumination optics. In section 3, we explain the numerical method. We give a concise explanation of all the steps of the generalized least squares method. These steps include functionals that impose the Monge-Ampère equation and the transport boundary conditions. Finally, we explain how to compute the mapping and the optical surfaces. In section 4, we study an example for the point to point optical system. In this example, we derive formulas for the optical surfaces and derive the mapping. In section 5, we discuss several aspects of the method that can be criticized.

2 Mathematical model

In this section, we will formulate the mathematical model of our optical system. The source intensity and target intensity are given. The goal is to design optical surfaces such that the twice reflected rays leaving from the source arrive at the target with the desired target intensity. We start with an introduction to the problem and the corresponding variables. Next we derive Hamilton's characteristic function and compute the cost function. Subsequently, we introduce stereographic coordinates that are preferred for point sources and targets. Furthermore, we formulate the optical mapping. Using this mapping and the stereographic coordinates, we formulate the conservation of energy and the corresponding transport boundary conditions. Finally, we introduce theory of optimal transport and see how this relates to illumination optics.

2.1 The optical system

We consider an optical system with a point source and a point target. The point source emits rays which are twice reflected before arriving at the point target. The optical system is depicted in Figure 1.

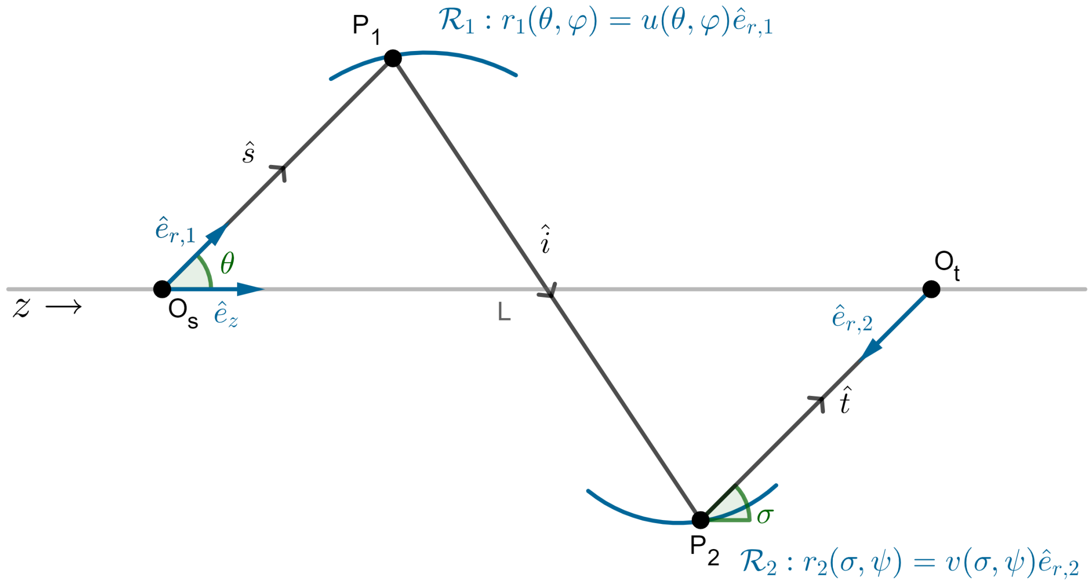

Figure 1: Illustration of the point-to-point optical system with two reflectors

The point source is located at O_s in the source plane $z = z_s = 0$. The rays leave the source with unit direction vector $\hat{\boldsymbol{s}} = (s_1, s_2, s_3)^\mathsf{T}$. The target is located at O_t in the target plane $z = z_t = L$, a distance L from the source. The unit direction vector $\hat{\boldsymbol{e}}_z = (0, 0, 1)$ denotes the direction from O_s to O_t. Rays arrive at the target with unit direction vector $\hat{\boldsymbol{t}} = (t_1, t_2, t_3)^\mathsf{T}$. Furthermore, we denote $\boldsymbol{q}_s = (0, 0)^\mathsf{T}$ and $\boldsymbol{p}_s = (s_1, s_2)^\mathsf{T}$ as the position and directions vectors of the rays emitted, projected on the source plane. Similarly, $\boldsymbol{q}_t = (0, 0)^\mathsf{T}$ and $\boldsymbol{p}_t = (t_1, t_2)^\mathsf{T}$ are the position and direction vectors of the rays arriving at the target, projected on the target plane. We denote the source and target domain by $\mathcal{X}$ and $\mathcal{Y}$, respectively.

The reflectors are denoted in spherical coordinates by parametrizations $\mathcal{R}_1 : \boldsymbol{r}_1(\theta,\varphi) = u(\theta,\varphi)\hat{\boldsymbol{e}}_{r,1}$ and $\mathcal{R}_2 : \boldsymbol{r}_2(\sigma,\psi) = v(\sigma,\psi)\hat{\boldsymbol{e}}_{r,2}$, where $\hat{\boldsymbol{e}}_{r,1}, \hat{\boldsymbol{e}}_{r,2}$ are the radial unit vectors from source and target, respectively. Furthermore, θ and σ are the polar angles, i.e., $\theta, \sigma \in [0,\pi]$. Furthermore, φ and ψ are the azimuthal angles, i.e., $\varphi, \psi \in [0, 2\pi)$. We choose the origin of the target coordinate system to be a point that approximates the optical surface v. We write $u(\theta,\varphi) = u(\hat{\boldsymbol{s}})$ and $v(\sigma,\psi) = v(\hat{\boldsymbol{t}})$.
In spherical coordinates, the unit direction vectors $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{t}}$ are

$$\hat{\boldsymbol{s}} = \begin{pmatrix} \sin(\theta)\cos(\varphi), \\ \sin(\theta)\sin(\varphi), \\ \cos(\theta) \end{pmatrix}, \qquad \hat{\boldsymbol{t}} = \begin{pmatrix} \sin(\sigma)\cos(\psi), \\ \sin(\sigma)\sin(\psi), \\ \cos(\sigma) \end{pmatrix}.$$

Next, we compute one of Hamilton's characteristic functions. Since we are interested in the source and target directions $\boldsymbol{p}_s, \boldsymbol{p}_t$, we consider the angular characteristic function T, which is defined as

$$T(z_s, z_t, \boldsymbol{p}_s, \boldsymbol{p}_t) = V(z_s, z_t, \boldsymbol{q}_s, \boldsymbol{q}_t) + \boldsymbol{q}_s \cdot \boldsymbol{p}_s - \boldsymbol{q}_t \cdot \boldsymbol{p}_t,$$

where V is the point characteristic. The point characteristic, also known as the optical path length, denotes the distance travelled by the light ray going from the source to the target. We have already seen that $\boldsymbol{q}_s = \boldsymbol{q}_t = (0,0)^\mathsf{T}$. Thus, the angular characteristic T is equal to the point characteristic V. Furthermore, it holds that

$$\frac{\partial T}{\partial \boldsymbol{p}_s} = \frac{\partial T}{\partial \boldsymbol{p}_t} = 0,$$

as shown in Section 2.7.4 of [6]. As a result, the angular characteristic T is independent of $\boldsymbol{p}_s$ and $\boldsymbol{p}_t$. Thus, the angular characteristic is constant. The optical path length can easily be found to be

$$V(z_s, z_t, \boldsymbol{q}_s, \boldsymbol{q}_t) = u(\hat{\boldsymbol{s}}) + v(\hat{\boldsymbol{t}}) + d(P_1, P_2),$$

where $d(P_1, P_2)$ is the distance between points P_1 and P_2 and P_1, P_2 are the points where the rays hit the first and second reflector, respectively. The distance between the reflectors can be computed using that the summation of vectors over a cycle is zero. That is,

$$\overrightarrow{O_sP_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2O_t} + \overrightarrow{O_tO_s} = 0.$$

Substituting all vectors we obtain

$$\begin{aligned} \overrightarrow{P_1P_2} &= -\overrightarrow{O_sP_1} - \overrightarrow{P_2} + \overrightarrow{O_tO_s}, \\ &= -u(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}} - v(\hat{\boldsymbol{t}})\hat{\boldsymbol{t}} + L\hat{\boldsymbol{e}}_z. \end{aligned}$$

Thus the distance between P_1 and P_2 equals

$$d(P_1, P_2) = |-u(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}} - v(\hat{\boldsymbol{t}})\hat{\boldsymbol{t}} + L\hat{\boldsymbol{e}}_z|,$$

where $|\cdot|$ denotes the Euclidean distance. Substituting the distance between the reflectors, we obtain

$$V(z_s, z_t, \boldsymbol{q}_s, \boldsymbol{q}_t) = u(\hat{\boldsymbol{s}}) + v(\hat{\boldsymbol{t}}) + |L\hat{\boldsymbol{e}}_z - u(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}} - v(\hat{\boldsymbol{t}})\hat{\boldsymbol{t}}|. \tag{1}$$

2.2 Derivation of cost function

In this section, we will derive the cost function. This is a function that relates the two optical surfaces by separating the variables related to the optical surfaces. It is typically written as

$$\tilde{u}(\hat{\boldsymbol{s}}) + \tilde{v}(\hat{\boldsymbol{t}}) = c(\hat{\boldsymbol{s}}, \hat{\boldsymbol{t}}),$$

where $\tilde{u}$ and $\tilde{v}$ relate to the first and second reflector, respectively. We can use (1) to obtain an expression for the cost function in terms of $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{t}}$. This is done by scaling and transforming the expressions for the reflectors, $u(\hat{\boldsymbol{s}})$ and $v(\hat{\boldsymbol{t}})$. First, bringing $u(\hat{\boldsymbol{s}}), v(\hat{\boldsymbol{t}})$ to the left-handside and squaring on both sides gives

$$V^2 - L^2 - 2u(\hat{\boldsymbol{s}})(V - L\hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}}) - 2v(\hat{\boldsymbol{t}})(V - L\hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}}) + 2u(\hat{\boldsymbol{s}})v(\hat{\boldsymbol{t}})(1 - \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{t}}) = 0. \tag{2}$$

To simplify computations, we scale the equation by introducing

$$\beta = \frac{V}{L}, \qquad \bar{u} = \frac{u(\hat{\boldsymbol{s}})}{L}, \qquad \bar{v} = \frac{v(\hat{\boldsymbol{t}})}{L}, \tag{3}$$

and obtain

$$\beta^2 - 1 - 2\bar{u}(\hat{\boldsymbol{s}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}}) - 2\bar{v}(\hat{\boldsymbol{t}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}}) + 2\bar{u}(\hat{\boldsymbol{s}})\bar{v}(\hat{\boldsymbol{t}})(1 - \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{t}}) = 0, \tag{4}$$

where we assume that $L \neq 0$, $\beta > 1$. In order to separate variables $\bar{u}$ and $\bar{v}$, we introduce functions $k_1(\hat{\boldsymbol{s}})$ and $k_2(\hat{\boldsymbol{t}})$, defined as

$$k_1(\hat{\boldsymbol{s}}) = \frac{1}{2\bar{u}(\hat{\boldsymbol{s}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}})}, \qquad k_2(\hat{\boldsymbol{t}}) = \frac{1}{2\bar{v}(\hat{\boldsymbol{t}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}})}. \tag{5}$$

Multiplying (4) with $k_1 k_2$, we obtain

$$\begin{aligned}
&(\beta^2 - 1)k_1 k_2 - k_1 - k_2 + \frac{1 - \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{t}}}{2(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}})} = 0,\\
&\left(k_1(\hat{\boldsymbol{s}}) - \frac{1}{\beta^2 - 1}\right)\left(k_2(\hat{\boldsymbol{t}}) - \frac{1}{\beta^2 - 1}\right) + \frac{1 - \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{t}}}{2(\beta^2 - 1)(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}})} - \frac{1}{(\beta^2 - 1)^2} = 0.
\end{aligned} \tag{6}$$

Finally, we introduce functions $\tilde{u}(\hat{\boldsymbol{s}})$ and $\tilde{v}(\hat{\boldsymbol{t}})$:

$$\tilde{u}(\hat{\boldsymbol{s}}) = \log\left(k_1(\hat{\boldsymbol{s}}) - \frac{1}{\beta^2 - 1}\right), \qquad \tilde{v}(\hat{\boldsymbol{t}}) = \log\left(k_2(\hat{\boldsymbol{t}}) - \frac{1}{\beta^2 - 1}\right), \tag{7}$$

where we assume the arguments of the logarithms to be positive. Substituting these functions gives an optimal transport formulation $\tilde{u}(\hat{\boldsymbol{s}}) + \tilde{v}(\hat{\boldsymbol{t}}) = c(\hat{\boldsymbol{s}}, \hat{\boldsymbol{t}})$ where

$$\tilde{u}(\hat{\boldsymbol{s}}) + \tilde{v}(\hat{\boldsymbol{t}}) = 2\log\left(\frac{1}{\beta^2 - 1}\right) + \log\left(1 - \frac{(\beta^2 - 1)(1 - \hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{t}})}{2(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{s}})(\beta - \hat{\boldsymbol{e}}_z \cdot \hat{\boldsymbol{t}})}\right). \tag{8}$$

2.3 Stereographic coordinates

We consider a point source and point target. Therefore, the source domain and target domain are in spherical coordinates. It is convenient to transform this to stereographic coordinates. In this transformation of coordinates, the unit direction vectors are projected on the equator plane

$z = 0$. The vectors $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{t}}$ are defined on the unit sphere S^2. As a result, the source and target domain become bounded and circular for cone-shaped beams. The stereographic coordinates $\boldsymbol{x}(\hat{\boldsymbol{s}})$ and $\boldsymbol{y}(\hat{\boldsymbol{t}})$ are defined as

$$\boldsymbol{x}(\hat{\boldsymbol{s}}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{1 \pm s_3} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \frac{1}{1 \pm \cos(\theta)} \begin{pmatrix} \sin(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) \end{pmatrix}, \tag{9a}$$

$$\boldsymbol{y}(\hat{\boldsymbol{t}}) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{1 \pm t_3} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \frac{1}{1 \pm \cos(\sigma)} \begin{pmatrix} \sin(\sigma)\cos(\psi) \\ \sin(\sigma)\sin(\psi) \end{pmatrix}. \tag{9b}$$

The corresponding inverse projections are

$$\hat{\boldsymbol{s}}(\boldsymbol{x}) = \frac{1}{1 + |\boldsymbol{x}|^2} \begin{pmatrix} 2x_1 \\ 2x_2 \\ \pm(1 - |\boldsymbol{x}|^2) \end{pmatrix}, \tag{10a}$$

$$\hat{\boldsymbol{t}}(\boldsymbol{y}) = \frac{1}{1 + |\boldsymbol{y}|^2} \begin{pmatrix} 2y_1 \\ 2y_2 \\ \pm(1 - |\boldsymbol{y}|^2) \end{pmatrix}. \tag{10b}$$

The $\pm$-signs in the stereographic projections (9a) and (9b) indicate the two options for the stereographic projection. The projection can be chosen from the south pole $(0, 0, -1)$ onto the plane $z = 0$, leading to a plus sign. Alternatively, the projection can be chosen from the north pole $(0, 0, 1)$ onto the plane $z = 0$, leading to a minus sign. When using a projection from the south pole, the stereographic projection is not defined for $s_3 = -1$ and $t_3 = -1$. For a projection from the north pole, the projection is not defined for $s_3 = 1$ and $t_3 = 1$. The projections can be chosen separately for the source and the target. For the light system depicted in 1, the best choice would be a projection from the south for $\hat{\boldsymbol{s}}$ as well as for $\hat{\boldsymbol{t}}$. This will result in a bounded domain on the plane $z = 0$ for $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{t}}$.

2.4 The mapping

We have previously mentioned that we consider an inverse method for the optical system. This means that given source intensity f and target intensity g, we want to find a mapping $m : \mathcal{X} \to \mathcal{Y}$ that maps the rays leaving the source such that they arrive at the target after being twice reflected. For stereographic source coordinates $\boldsymbol{x}$ and stereographic target coordinates $\boldsymbol{y}$, we denote the mapping as

$$\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{y}, \qquad \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y} \in \mathcal{Y},$$

where $\mathcal{X}, \mathcal{Y}$ are now the source and target domain in stereographic coordinates, respectively.

2.5 Energy conservation

An important conditions for an optical system is energy conservation. This means that all light emitted from the source should arrive at the target. The intensity of the point source is given by the distribution $f(\theta, \varphi)$ in lumen per steradians [lm/sr]. The intensity of the target is given by the distribution $g(\sigma, \psi)$, again given in [lm/sr]. Energy conservation states that

$$\int_A f(\theta, \varphi)\mathrm{d}S(\theta, \varphi) = \int_{\hat{\boldsymbol{t}}(A)} g(\sigma, \psi)\mathrm{d}S(\sigma, \psi),$$

for arbitrary $A \subset S^2$ and image set $\hat{\boldsymbol{t}}(A) \subset S^2$. As we use stereographic coordinates in the optical system, we have to apply the transformation for the energy balance. This gives

$$\int_{\boldsymbol{x}(A)} \tilde{f}(\boldsymbol{x}) \left| \frac{\partial \hat{\boldsymbol{s}}}{\partial x_1} \times \frac{\partial \hat{\boldsymbol{s}}}{\partial x_2} \right| \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{y}(\hat{\boldsymbol{t}}(A))} \tilde{g}(\boldsymbol{y}) \left| \frac{\partial \hat{\boldsymbol{t}}}{\partial y_1} \times \frac{\partial \hat{\boldsymbol{t}}}{\partial y_2} \right| \mathrm{d}\boldsymbol{y},$$

where $\tilde{f}$ and $\tilde{g}$ are the intensities in stereographic coordinates. Computing the Jacobians resulting from the coordinate transformation gives

$$\int_{\boldsymbol{x}(A)} \tilde{f}(\boldsymbol{x}) \frac{4}{(1+|\boldsymbol{x}|^2)^2} \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{y}(\hat{\boldsymbol{t}}(A))} \tilde{g}(\boldsymbol{y}) \frac{4}{(1+|\boldsymbol{y}|^2)^2} \mathrm{d}\boldsymbol{y}.$$

Thereafter, we substitute the mapping $\boldsymbol{y} = \boldsymbol{m}(\boldsymbol{x})$ to obtain

$$\int_{\boldsymbol{x}(A)} \tilde{f}(\boldsymbol{x}) \frac{4}{(1+|\boldsymbol{x}|^2)^2} \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{x}(A)} \tilde{g}(\boldsymbol{m}(\boldsymbol{x})) \frac{4}{(1+|\boldsymbol{m}(\boldsymbol{x})|^2)^2} |\det\big(\mathrm{D}\boldsymbol{m}(\boldsymbol{x})\big)| \mathrm{d}\boldsymbol{x}.$$

As this holds for arbitrary A, we obtain

$$\det\big(\mathrm{D}\boldsymbol{m}(\boldsymbol{x})\big) = \frac{\tilde{f}(\boldsymbol{x})}{\tilde{g}(\boldsymbol{m}(\boldsymbol{x}))} \frac{(1+|\boldsymbol{m}(\boldsymbol{x})|^2)^2}{(1+|\boldsymbol{x}|^2)^2}, \tag{11}$$

where we assume a positive determinant. Equation (11) is known as the Jacobian equation.

2.5.1 Transport boundary condition

The generalized Monge-Ampère equation is accompanied by the transport boundary condition:

$$\boldsymbol{m}(\partial \mathcal{X}) = \partial \mathcal{Y}.$$

This condition states that the boundary of source domain $\mathcal{X}$ is mapped to the boundary of the target domain $\mathcal{Y}$. The condition is derived from the implicit boundary condition $\boldsymbol{m}(\mathcal{X}) = \mathcal{Y}$ together with the edge-ray principle, the proof of which can be found in Section 4.5 of [6].

2.6 Optimal transport

In this section, we will explain the optimal transport problem and explain how this relates to optical design problems. We will use optimal transport to derive the Monge-Ampère equation. This derivation will be concise but useful as it gives more insight in the underlying minimization problem. For a more detailed derivation, see Section 4.3 of [6].

The goal in optimal transport is to find a mapping $\boldsymbol{m} : \mathcal{X} \to \mathcal{Y}$ that transforms a measure μ on $\mathcal{X}$ to a measure ν on $\mathcal{Y}$ while minimizing the cost of transport. We denote the push-forward of μ by $\boldsymbol{m}$ as $\boldsymbol{m}_{\#}(\mu)$. The first formulation of an optimal transport problem was given by Gaspard Monge in 1781 [7]. He formulated the problem as follows.

Formulation 2.1 (Monge)**.** *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a Borel-measurable function. Let μ and ν be probability measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then the aim is to find a mapping m: $\mathcal{X} \to \mathcal{Y}$ that attains the infimum*

$$\inf_{\boldsymbol{m} \in \mathcal{M}} I[\boldsymbol{m}] = \inf_{\boldsymbol{m} \in \mathcal{M}} \int_{\mathcal{X}} c(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x})) \mathrm{d}\mu(\boldsymbol{x}), \tag{12}$$

where $\mathcal{M}$ is the set of all measure-preserving mappings such that $\boldsymbol{m}_{\#}(\mu) = \nu$, that is, $\nu = \mu \circ \boldsymbol{m}^{-1}$.

Monge's formulation of the optimal transport problem can be ill-posed as an $\boldsymbol{m}$ such that $\boldsymbol{m}_{\#}(\mu) = \nu$ does not always exist [Section 4.3 of [6]]. In 1942, Kantorovich formulated the same problem.

Formulation 2.2 (Monge-Kantorovich). *Let $\Gamma(\mu, \nu)$ be the collection of probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ on $\mathcal{X}$ and ν on $\mathcal{Y}$. That is, for all $A \subset \mathcal{X}$: $\gamma(A \times \mathcal{Y}) = \mu(A)$ and for all $B \subset \mathcal{Y}$: $\gamma(\mathcal{X} \times B) = \nu(B)$. The aim is to find the infimum*

$$\inf_{\gamma \in \Gamma(\mu,\nu)} J[\gamma] = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\gamma(\boldsymbol{x}, \boldsymbol{y}). \tag{13}$$

This formulation is known as the Monge-Kantorovich formulation. It can be shown that under some conditions on the cost function, a pair (u, v) and a mapping $\boldsymbol{m}$ can be found that minimizes the transportation costs as stated in the Monge-Kantovorich formulation (13). This can be shown using c-convex/c-concave functions and c-convex/c-concave pairs. All further definitions and derivations start with the formulation of the cost function:

$$v(\boldsymbol{y}) - u(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{y}).$$

Note that this formulation differs a minus sign from previous notations. However, the two formulations can be interchanged.
We start with the definitions for c-transforms [Section 4.3 of [6]].

Definition 1. *The c-transform $v^* : \mathcal{X} \to \mathbb{R}$ of $v : \mathcal{Y} \to \mathbb{R}$ is defined as*

$$\forall \boldsymbol{x} \in \mathcal{X} : v^*(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{Y}} \left(-c(\boldsymbol{x}, \boldsymbol{y}) + v(\boldsymbol{y})\right). \tag{14}$$

Definition 2. *The c-transform $u^* : \mathcal{Y} \to \mathbb{R}$ of $u : \mathcal{X} \to \mathbb{R}$ is defined as*

$$\forall \boldsymbol{y} \in \mathcal{Y} : u^*(\boldsymbol{y}) = \inf_{\boldsymbol{x} \in \mathcal{X}} \left(c(\boldsymbol{x}, \boldsymbol{y}) + u(\boldsymbol{x})\right). \tag{15}$$

Next we define c-convex and c-concave functions.

Definition 3. *A function $u : \mathcal{X} \to \mathbb{R}$ is c-convex if there exists $v : \mathcal{Y} \to \mathbb{R}$ such that the c-transform of v equals u, i.e., $u = v^*$.*

Definition 4. *A function $u : \mathcal{X} \to \mathbb{R}$ is c-concave if the supremum and infimum in* (14) *and* (15) *are interchanged and there exists $v : \mathcal{Y} \to \mathbb{R}$ such that the c-transform of v equals u, i.e., $u = v^*$.*

Lastly, we have to define a c-convex/c-concave pair.

Definition 5. *For a conjugate pair (u, v) with $u = v^*$ and $v = u^*$, leading to $u = u^{**}$, we define a c-convex pair as*

$$\forall \boldsymbol{x} \in \mathcal{X} : u(\boldsymbol{x}) = \max_{\boldsymbol{y} \in \mathcal{Y}} \left(v(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})\right), \tag{16a}$$

$$\forall \boldsymbol{y} \in \mathcal{Y} : v(\boldsymbol{y}) = \min_{\boldsymbol{x} \in \mathcal{X}} \left(u(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y})\right), \tag{16b}$$

where u is c-convex and v is c-concave.

Definition 6. *For a conjugate pair* (u, v) *with* $u = v^*$ *and* $v = u^*$*, leading to* $u = u^{**}$*, we define a c-concave pair as*

$$\forall \boldsymbol{x} \in \mathcal{X} : u(\boldsymbol{x}) = \min_{\boldsymbol{y}\in\mathcal{Y}} \left(v(\boldsymbol{y}) - c(\boldsymbol{x}, \boldsymbol{y})\right), \tag{17a}$$

$$\forall \boldsymbol{y} \in \mathcal{Y} : v(\boldsymbol{y}) = \max_{\boldsymbol{x}\in\mathcal{X}} \left(u(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y})\right), \tag{17b}$$

where u *is c-concave and* v *is c-convex.*

Now that all definitions are given, we can formulate conditions for a minimizer of the optimal transport problem.

Theorem 1. *For a bounded and continuous cost function* c *with injective* $\nabla_x c(\boldsymbol{x}, \cdot)$*, there exists a conjugate pair* (u, v) *that minimizes the Monge-Kantorovich problem* (13) *and a corresponding unique map that minimizes the Monge-problem* (12)*. Furthermore, there exists a c-convex/c-concave* u *such that*

$$\nabla u(x) + \nabla_x c(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x})) = \boldsymbol{0}, \tag{18}$$

which is a stationary point of $u(\boldsymbol{x}) + c(\boldsymbol{x}, \boldsymbol{y})$ *in* (16b) *or* (17b)*. As a result, the mapping can be found implicitly using* (18)*.*

A proof of existence can be found in [8].

2.6.1 Derivation of the Monge-Ampère equation

In the previous section, we saw that there exists a mapping $\boldsymbol{m} : \mathcal{X} \to \mathcal{Y}$ that satisfies the Jacobian equation (11) and a conjugate pair (u, v) that satisfies $v(\boldsymbol{y}) - u(\boldsymbol{x}) = c(\boldsymbol{x}, \boldsymbol{y})$. Furthermore, the mapping $\boldsymbol{m}$ can be found implicitly using (18). However, it should hold that the mixed Hessian matrix, defined as

$$\boldsymbol{C} = \mathrm{D}_{\boldsymbol{xy}} c = \begin{pmatrix} \frac{\partial^2 c}{\partial x_1 \partial y_1} & \frac{\partial^2 c}{\partial x_1 \partial y_2} \\ \frac{\partial^2 c}{\partial x_2 \partial y_1} & \frac{\partial^2 c}{\partial x_2 \partial y_2} \end{pmatrix},$$

is invertible. In order to obtain the minimum or maximum defined in the c-convex/c-concave pairs, it should hold that

$$\boldsymbol{P}(\boldsymbol{x}) = -\mathrm{D}^2 u(\boldsymbol{x}) - \mathrm{D}_{\boldsymbol{xx}} c(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x})) \tag{19}$$

is symmetric negative definite (SND) or symmetric positive definite (SPD), respectively. Differentiating (18) w.r.t. $\boldsymbol{x}$, we obtain

$$\mathrm{D}^2 u(\boldsymbol{x}) + \mathrm{D}_{\boldsymbol{xx}} c(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x})) + \boldsymbol{C}(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x}))\mathrm{D}\boldsymbol{m}(x) = \boldsymbol{0}.$$

Substituting $\boldsymbol{P}$ from (19) gives

$$\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x}))\mathrm{D}\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{P}(\boldsymbol{x}). \tag{20}$$

Now taking the determinants and using the Jacobian equation (11) gives

$$\det\big(\mathrm{D}\boldsymbol{m}(\boldsymbol{x})\big) = \frac{\det\big(\boldsymbol{P}(\boldsymbol{x})\big)}{\det\big(\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x}))\big)} = \frac{f(\boldsymbol{x})}{g(\boldsymbol{m}(\boldsymbol{x}))} = F(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x})), \tag{21}$$

where we assume that f, g have incorporated the Jacobians due to coordinate transformation and denote the right-handside by F. Finally, we substitute $\boldsymbol{P}$ again to obtain the generalized Monge-Ampère equation

$$\det\Big(\mathrm{D}^2 u(\boldsymbol{x}) + \mathrm{D}_{\boldsymbol{x}\boldsymbol{x}} c(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x}))\Big) = \det\big(\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{m}(\boldsymbol{x}))\big) \frac{f(\boldsymbol{x})}{g(\boldsymbol{m}(\boldsymbol{x}))}.$$

3 Numerical method

In this section, we will explain the numerical method which is the generalized least squares algorithm. A full documentation of this algorithm is found in Section 6 of [6]. We will give a concise explanation of the algorithm.

In the first part of the algorithm, the mapping $\boldsymbol{m}$ is computed using a 3-step iterative minimization process. The mapping $\boldsymbol{m}$ can be computed from the generalized Monge-Ampère equation (20) subject to (11). Alongside the Monge-Ampère equation, there is the transport boundary condition which states

$$\boldsymbol{m}(\partial\mathcal{X}) = \partial\mathcal{Y}. \tag{22}$$

From (21) and (22), the mapping can be computed using a constrained minimization problem. This is done via an iterative process minimizing three functionals.
The first functional J_I, defined as

$$J_I[\boldsymbol{m}, \boldsymbol{P}] = \tfrac{1}{2} \iint_{\mathcal{X}} \|\boldsymbol{C}\,\mathrm{D}\boldsymbol{m} - \boldsymbol{P}\|^2 \mathrm{d}\boldsymbol{x},$$

where $\|\cdot\|$ denotes the Frobenius norm, enforces (20) under the constraint $\det(\boldsymbol{P}) = F\det(\boldsymbol{C})$.
The second functional J_B, defined as

$$J_B[\boldsymbol{m}, \boldsymbol{b}] = \tfrac{1}{2} \oint_{\partial\mathcal{X}} |\boldsymbol{m} - \boldsymbol{b}|^2 \mathrm{d}s,$$

where $\boldsymbol{b} : \partial\mathcal{X} \to \partial\mathcal{Y}$ and $|\cdot|$ denotes the L_2-norm, enforces the transport boundary condition (22). Lastly, functional J, defined as

$$J[\boldsymbol{m}, \boldsymbol{P}, \boldsymbol{b}] = \alpha J_I[\boldsymbol{m}, \boldsymbol{P}] + (1-\alpha) J_B[\boldsymbol{m}, \boldsymbol{b}],$$

is a weighted average of functionals J_I and J_B for $0 < \alpha < 1$. By choosing α, you can either emphasize the boundary or emphasize the interior.
The iterative process starts with initial guess $\boldsymbol{m}^0$ and cost matrix $\boldsymbol{C}(\cdot, \boldsymbol{m}^0)$. The iteration steps are

$$\begin{aligned}
\boldsymbol{b}^{n+1} &= \underset{\boldsymbol{b}\in\mathcal{B}}{\operatorname{argmin}}\ J_B[\boldsymbol{m}^n, \boldsymbol{b}], \\
\boldsymbol{P}^{n+1} &= \underset{\boldsymbol{p}\in\mathcal{P}}{\operatorname{argmin}}\ J_I[\boldsymbol{m}^n, \boldsymbol{P}], \\
\boldsymbol{m}^{n+1} &= \underset{\boldsymbol{m}\in\mathcal{M}}{\operatorname{argmin}}\ J[\boldsymbol{m}, \boldsymbol{P}^{n+1}, \boldsymbol{b}^{n+1}],
\end{aligned}$$

where the spaces are defined as

$$\begin{aligned}
&\mathcal{B} = \{\boldsymbol{b} \in C^1(\partial\mathcal{X})^2 \,|\, \boldsymbol{b}(x) \in \partial\mathcal{Y}\}, \\
&\mathcal{P} = \{\boldsymbol{P} \in C^1(\mathcal{X})^{2\times 2} \,|\, \boldsymbol{P} \text{ SND/SPD}, \det(\boldsymbol{P}) = \det\big(\mathrm{F}(\cdot, \boldsymbol{m})\big)\det\big(\boldsymbol{C}(\cdot, \boldsymbol{m})\big)\}, \\
&\mathcal{M} = C^2(\mathcal{X})^2.
\end{aligned}$$

As an initial guess $\boldsymbol{m}^0$, we map the smallest box enclosing the source domain $\mathcal{X}$, i.e., $[a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}]$ to the smallest box enclosing the target domain $\mathcal{Y}$, i.e., $[c_{\min}, c_{\max}] \times [d_{\min}, d_{\max}]$. Furthermore, the source grid is discretized using a rectangular grid of size $N_1 \times N_2$:

$$\begin{aligned}
x_{1,i} &= a_{\min} + (i-1)h_1, \\
x_{2,j} &= b_{\min} + (j-1)h_2,
\end{aligned}$$

where h_1 and h_2 are the constant grid sizes. We will now explain each step of the 3-step minimization process separately.

3.1 Computation of $\boldsymbol{b}$

In the computation of $\boldsymbol{b}$, we minimize the functional J_B as this imposes the transport boundary condition. The minimization can be performed pointwise since we do not integrate over derivative of $\boldsymbol{b}$, thus we compute

$$\min_{b_{ij}\in\mathcal{B}} \tfrac{1}{2}|m_{ij} - b_{ij}|^2,$$

where $m_{ij} = \boldsymbol{m}(x_{ij})$ and $b_{ij} = \boldsymbol{b}(x_{ij})$. The point $\boldsymbol{b}$ is the point on the boundary $\partial\mathcal{Y}$ that minimizes the distance between $\boldsymbol{m}(\partial\mathcal{X})$ and the target boundary $\partial\mathcal{Y}$. This is computed for all points $\boldsymbol{x}_{ij}$. The minimum is computed using skew projections on line segments that are parallel to fragments of the boundary $\partial\mathcal{Y}$. A full explanation of the skew projection can be found in Section 6.1.1 of [6].

3.2 Computation of P

In the computation of $\boldsymbol{P}$, we minimize the functional J_I, which imposes the generalized Monge-Ampère equation (21). This minimization can again be performed pointwise since we do not integrate over derivative of $\boldsymbol{P}$. We want to solve the following constrained minimization problem

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|\boldsymbol{Q}_s - \boldsymbol{P}\|^2, \\ &\text{subject to} && \det(\boldsymbol{P}) = \mathrm{F}\det\big(\boldsymbol{C}(\cdot,\boldsymbol{m})\big)). \end{aligned}$$

where $\boldsymbol{Q}_s = \frac{1}{2}(\boldsymbol{Q}+\boldsymbol{Q}^{\mathsf{T}})$ is the symmetric part of Q which is the central difference approximation of $\mathrm{D}\boldsymbol{m}$. An additional constraint is $\mathrm{tr}(\boldsymbol{P}) \leq 0$ or $\mathrm{tr}(\boldsymbol{P}) \geq 0$ for a c-convex or c-concave solution u, respectively. In order to solve the minimization problem, the Lagrangian is computed. Subsequently, the partial derivatives of the Lagrangian w.r.t elements of $\boldsymbol{P}$ and the Lagrangian multiplier are computed and set equal to zero. Different stationary points can be found for different values of the Lagrangian multiplier. A complete overview is given in Section 6.1.2 of [6].

3.3 Computation of $\boldsymbol{m}$

For the computation of $\boldsymbol{m}$, we minimize the functional J. This minimization can not be done pointwise since we integrate over a derivative of $\boldsymbol{m}$. Therefore, we compute the first variation w.r.t. $\boldsymbol{m}$ and set it equal to zero. The first variation is

$$\delta J[\boldsymbol{m},\boldsymbol{P},\boldsymbol{b}](\boldsymbol{\eta}) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(J[\boldsymbol{m}+\varepsilon\boldsymbol{\eta},\boldsymbol{P},\boldsymbol{b}] - J[\boldsymbol{m},\boldsymbol{P},\boldsymbol{b}]\right).$$

This can be rewritten to be

$$\delta J[\boldsymbol{m},\boldsymbol{P},\boldsymbol{b}](\boldsymbol{\eta}) = \alpha\int_{\mathcal{X}}(\boldsymbol{C}\,\mathrm{D}\boldsymbol{m} - \boldsymbol{P}) : \boldsymbol{C}\,\mathrm{D}\boldsymbol{\eta}\,\mathrm{d}\boldsymbol{x} + (1-\alpha)\oint_{\partial\mathcal{X}}(\boldsymbol{m}-\boldsymbol{b})\cdot\boldsymbol{\eta}\,\mathrm{d}s, \tag{23}$$

where $:$ denotes the Frobenius inner product defined as $A : B = \sum_{i,j}\bar{A_{ij}}B_{ij}$. Using the fundamental lemma of calculus of variations, we can formulate a boundary value problem from

(23):

$$\nabla \cdot (\boldsymbol{C}^{\mathsf{T}}\boldsymbol{C}\,\mathrm{D}\boldsymbol{m}) = \nabla \cdot (\boldsymbol{C}^{\mathsf{T}}\boldsymbol{P}), \qquad \boldsymbol{x} \in \mathcal{X}, \tag{24a}$$

$$(1-\alpha)\boldsymbol{m} + \alpha(\boldsymbol{C}^{\mathsf{T}}\boldsymbol{C}\,\mathrm{D}\boldsymbol{m})\hat{\boldsymbol{n}} = (1-\alpha)\boldsymbol{b} + \alpha\boldsymbol{C}^{\mathsf{T}}\boldsymbol{P}\hat{\boldsymbol{n}}, \qquad \boldsymbol{x} \in \partial\mathcal{X}, \tag{24b}$$

where $\hat{\boldsymbol{n}}$ is the outward unit normal on the source boundary $\partial\mathcal{X}$. Boundary value problem (24) is discretized using the finite volume method which leads to a linear system that is solved using the `mldivide` operator in `Matlab`.

3.4 Computation of u

After convergence of the minimization problem which computes $\boldsymbol{m}$, we can compute the optical surface u. The surface u can be computed from (18). However, due to numerical errors, $\nabla_x c$ is not conservative anymore. As a result, a direct integration method does not give a unique solution. Thus we compute the least-squares solution. This is done by minimizing the functional

$$I[u] = \tfrac{1}{2}\int_{\mathcal{X}} |\nabla u + \nabla_x c(\cdot, \boldsymbol{m})|^2 \mathrm{d}\boldsymbol{x}.$$

Similarly as before, this is solved by setting the first variation to zero:

$$\begin{aligned}\delta I[u](v) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\big(I[u+\varepsilon v] - I[u]\big)\\ &= \int_{\mathcal{X}} \big(\nabla u + \nabla_x c(\cdot, \boldsymbol{m})\big) \cdot \nabla v \,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Using Gauss's theorem and the fundamental lemma of calculus of variations, we obtain the following BVP:

$$\Delta u = -\nabla \cdot \nabla_x c(\cdot, \boldsymbol{m}), \qquad \boldsymbol{x} \in \mathcal{X}, \tag{25a}$$

$$\nabla u \cdot \hat{\boldsymbol{n}} = -\nabla_x c(\cdot, \boldsymbol{m}) \cdot \hat{\boldsymbol{n}}, \qquad \boldsymbol{x} \in \partial\mathcal{X}, \tag{25b}$$

where $\hat{\boldsymbol{n}}$ is the outward unit normal of $\partial\mathcal{X}$. This BVP still does not give a unique solution. We can compute a unique solution by choosing the value of u in one point. The computation of this value is explained in [9]. This value is added as a constraint to the BVP. As a result, we can uniquely solve BVP (25).

4 Example

In this section, we will consider a relatively simple example. The optical system is shown in Figure 2.

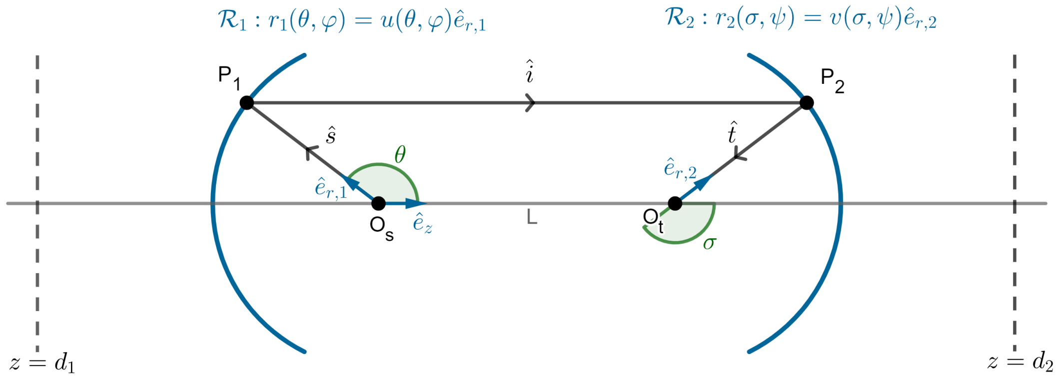

Figure 2: Illustration of the point-to-point example with two reflectors

We consider the source to be at the origin, i.e., $O_s = (0, 0, 0)$. The target is located at a distance L away along the z-axis, i.e., $O_t = (0, 0, L)$. We consider two identically shaped paraboloids as reflectors. The first reflector has focus O_s and directrix $z = d_1$. The second reflector has focus O_t and directrix $z = d_2$. Using these reflectors, all rays leaving the point source will be reflected into a collimated beam. This beam is again reflected such that all rays arrive at the point target. For the source and target, we consider a stereographic projection from the north pole.

4.1 Optical surfaces

A paraboloid with focus (a, b, c) and directrix $z = d$ in Cartesian coordinates is given by

$$(x - a)^2 + (y - b)^2 + (z - c)^2 = (z - d)^2. \tag{26}$$

4.1.1 First reflector

For the first reflector $\mathcal{R}_1$, we have focus $O_s = (0, 0, 0)$ and directrix $z = d_1$. The location of the directrix can be computed using a ray leaving the source in direction $(0, 0, -1)$. The optical path length will then be

$$V = 4|\tfrac{1}{2}d_1| + L. \tag{27}$$

Dividing (27) by L and taking the directrix on the negative z-axis, we get $d_1 = -\frac{1}{2}L(\beta - 1)$ Thus the formula in Cartesian coordinates reads

$$x^2 + y^2 = -2zd_1 + d_1^2. \tag{28}$$

Using the parametrization $\mathcal{R}_1 : \boldsymbol{r}_1(\theta, \varphi) = u(\theta, \varphi)\hat{\boldsymbol{e}}_{r,1}$ and the inverse projection (10a) from the north pole, we obtain Cartesian coordinates in terms of the stereographic source coordinates

x_1, x_2:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = u\,\hat{\boldsymbol{s}}(\boldsymbol{x}) = \frac{u}{1+|\boldsymbol{x}|^2}\begin{pmatrix} 2x_1 \\ 2x_2 \\ -(1-|\boldsymbol{x}|^2) \end{pmatrix}.$$

Substituting these Cartesian coordinates into (28) gives

$$\frac{4|\boldsymbol{x}|^2u^2}{(1+|\boldsymbol{x}|^2)^2} = \frac{2ud_1(1-|\boldsymbol{x}|^2)}{1+|\boldsymbol{x}|^2} + d_1^2. \tag{29}$$

Solving for the location of the optical surface $u = u(\boldsymbol{x})$ gives

$$u(\boldsymbol{x}) = -\tfrac{1}{2}d_1(1+|\boldsymbol{x}|^2). \tag{30}$$

Thus we have obtained a formula for the first optical surface $u(\boldsymbol{x})$ in terms of stereographic coordinates $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$.

In the derivation of the cost function, we have applied a transformation given by (3),(5) and (7). This transformation also has to be applied to (30). Thus we first obtain $\bar{u}$:

$$\bar{u}(\boldsymbol{x}) = \tfrac{1}{4}(\beta-1)(1+|\boldsymbol{x}|^2).$$

Next we compute k_1:

$$k_1(\boldsymbol{x}) = \frac{2}{(\beta-1)(1+|\boldsymbol{x}|^2)\left(\beta+\frac{1-|\boldsymbol{x}|^2}{1+|\boldsymbol{x}|^2}\right)}.$$

Lastly, we substitute k_1 into (7) to obtain $\tilde{u}$:

$$\tilde{u}(\boldsymbol{x}) = \log\left(\frac{2}{(\beta-1)(1+|\boldsymbol{x}|^2)\left(\beta+\frac{1-|\boldsymbol{x}|^2}{1+|\boldsymbol{x}|^2}\right)} - \frac{1}{\beta^2-1}\right)$$

$$\tilde{u}(\boldsymbol{x}) = \log\left(\frac{\beta+1-(\beta-1)|\boldsymbol{x}|^2}{(\beta^2-1)(\beta+1+(\beta-1)|\boldsymbol{x}|^2)}\right). \tag{31}$$

4.1.2 Second reflector

The second reflector has focus $O_t = (0,0,L)$ and directrix $z = d_2$. As we consider two identically shaped reflectors, we can easily see that $d_2 = L + |d_1| = \frac{1}{2}L(\beta+1)$. Using (26), the reflector in Cartesian coordinates is

$$x^2 + y^2 = 2z(L-d_2) + d_2^2 - L^2. \tag{32}$$

Similar as for the first reflector, we use parametrization $\mathcal{R}_2 : \boldsymbol{r}_2(\phi,\psi) = v(\phi,\psi)\hat{\boldsymbol{e}}_{r,2}$ and inverse projection (10b) from the south pole to obtain Cartesian coordinates written in stereographic coordinates. However, we want a projection w.r.t. the target $O_t = (0,0,L)^\mathsf{T}$. Thus we have inverse projection

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -v\,\hat{\boldsymbol{t}}(\boldsymbol{y}) + \begin{pmatrix} 0 \\ 0 \\ L \end{pmatrix} = \frac{-v}{1+|\boldsymbol{y}|^2}\begin{pmatrix} 2y_1 \\ 2y_2 \\ -(1-|\boldsymbol{y}|^2) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ L \end{pmatrix}.$$

Note that $\hat{\boldsymbol{t}}$ points towards the target, instead of away from the target, leading to a minus sign. Substituting this into (32) gives the following equation for the location of the second optical surface $v = v(\boldsymbol{y})$:

$$\begin{aligned}\frac{4|\boldsymbol{y}|^2 v^2}{(1+|\boldsymbol{y}|^2)^2} &= 2(v\frac{1-|\boldsymbol{y}|^2}{1+|\boldsymbol{y}|^2} + L)(L-d_2) + d_2^2 - L^2, \\ \frac{4|\boldsymbol{y}|^2 v^2}{(1+|\boldsymbol{y}|^2)^2} &= 2v\frac{1-|\boldsymbol{y}|^2}{1+|\boldsymbol{y}|^2}(L-d_2) + (L-d_2)^2 \end{aligned} \tag{33}$$

Solving for v gives

$$v(\boldsymbol{y}) = -\tfrac{1}{2}(L-d_2)(1+|\boldsymbol{y}|^2).$$

We can now use that $L - d_2 = d_1$ and obtain

$$v(\boldsymbol{y}) = -\tfrac{1}{2}d_1(1+|\boldsymbol{y}|^2). \tag{34}$$

Similar as for u, we have to apply the transformations (3), (5), (7). Firstly,

$$\bar{v}(\boldsymbol{y}) = \tfrac{1}{4}(\beta-1)(1+|\boldsymbol{y}|^2).$$

Subsequently,

$$k_2(\boldsymbol{y}) = \frac{2}{(\beta-1)(\beta-1+(\beta+1)|\boldsymbol{y}|^2)}.$$

Lastly, we obtain

$$\begin{aligned}\tilde{v}(\boldsymbol{y}) &= \log\left(\frac{2}{(\beta-1)(1+|\boldsymbol{y}|^2)\left(\beta+\frac{1-|\boldsymbol{y}|^2}{1+|\boldsymbol{y}|^2}\right)} - \frac{1}{\beta^2-1}\right) \\ \tilde{v}(\boldsymbol{y}) &= \log\left(\frac{\beta+1-(\beta-1)|\boldsymbol{y}|^2}{(\beta^2-1)(\beta+1+(\beta-1)|\boldsymbol{y}|^2)}\right). \end{aligned} \tag{35}$$

Now that we have computed both reflectors, we can make a 3D plot of our optical system. This can be seen in Figure 3.

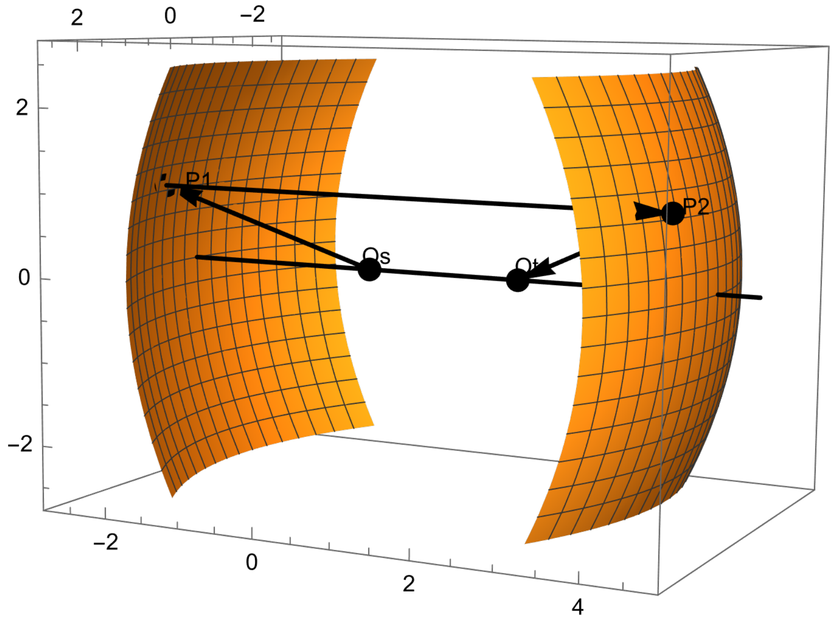

Figure 3: 3D illustration of the example, $L = 2$, $\beta = 6$

4.2 The mapping

For this example, we can show see that the mapping in stereographic coordinates is $\boldsymbol{m}(\boldsymbol{x}) = -\boldsymbol{x}$. We start with the fact that vectors over a cycle are zero:

$$\overrightarrow{O_sP_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2O_t} + \overrightarrow{O_tO_s} = 0,$$
$$u\hat{\boldsymbol{s}} + v\hat{\boldsymbol{t}} = (L - d(P_1, P_2))\hat{e}_z.$$

For the first two components, we obtain

$$u\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + v\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \mathbf{0}.$$

Substituting the stereographic projections, we obtain

$$u(1 - s_3)\boldsymbol{x} + v(1 - t_3)\boldsymbol{y} = \mathbf{0}, \tag{36}$$
$$\frac{2u\boldsymbol{x}}{1 + |\boldsymbol{x}|^2} + \frac{2v\boldsymbol{y}}{1 + |\boldsymbol{y}|^2} = \mathbf{0}.$$

Taking the inner product with $\boldsymbol{x}$ as well as $\boldsymbol{y}$ results in two equations:

$$\frac{2|\boldsymbol{x}|^2}{1 + |\boldsymbol{x}|^2}u + \frac{2\boldsymbol{x} \cdot \boldsymbol{y}}{1 + |\boldsymbol{y}|^2}v = \mathbf{0},$$
$$\frac{2\boldsymbol{x} \cdot \boldsymbol{y}}{1 + |\boldsymbol{x}|^2}u + \frac{2|\boldsymbol{y}|^2}{1 + |\boldsymbol{y}|^2}v = \mathbf{0}.$$

As this is linear system and we do not look for the trivial solution, we set the determinant of the linear operator equal to zero. This gives

$$(\boldsymbol{x} \cdot \boldsymbol{y})^2 = |\boldsymbol{x}|^2|\boldsymbol{y}|^2.$$

Thus we obtain that $\boldsymbol{x} = \pm\boldsymbol{y}$. However, $\boldsymbol{x} = \boldsymbol{y}$ contradicts (36). Thus we find that the mapping is $\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{y} = -\boldsymbol{x}$.

4.3 Cost function

If we use steregraphic projections from the north pole, the cost function (8) becomes

$$
\begin{aligned}
c(\boldsymbol{x},\boldsymbol{y}) &= 2\log\left(\frac{1}{\beta^2-1}\right) + \log\left(\frac{(\beta+1)^2 + 2(\beta^2-1)\boldsymbol{x}\cdot\boldsymbol{y} + (\beta-1)^2|\boldsymbol{x}|^2|\boldsymbol{y}|^2}{\left(\beta+1+(\beta-1)|\boldsymbol{x}|^2\right)\left(\beta+1+(\beta-1)|\boldsymbol{y}|^2\right)}\right),\\
&= \log\left(\frac{(\beta+1)^2 + 2(\beta^2-1)\boldsymbol{x}\cdot\boldsymbol{y} + (\beta-1)^2|\boldsymbol{x}|^2|\boldsymbol{y}|^2}{(\beta^2-1)^2\left(\beta+1+(\beta-1)|\boldsymbol{x}|^2\right)\left(\beta+1+(\beta-1)|\boldsymbol{y}|^2\right)}\right)
\end{aligned}
$$

Using (31) and (35), we can compute $\tilde{u} + \tilde{v}$:

$$
\begin{aligned}
\tilde{u}(\boldsymbol{x}) + \tilde{v}(\boldsymbol{y}) &= \log\left(\frac{(\beta+1-(\beta-1)|\boldsymbol{y}|^2)(\beta+1-(\beta-1)|\boldsymbol{x}|^2)}{(\beta^2-1)^2(\beta+1+(\beta-1)|\boldsymbol{y}|^2)(\beta+1+(\beta-1)|\boldsymbol{x}|^2)}\right),\\
&= \log\left(\frac{(\beta+1)^2 - (\beta^2-1)(|\boldsymbol{x}|^2+|\boldsymbol{y}|^2) + (\beta-1)^2|\boldsymbol{x}|^2|\boldsymbol{y}|^2}{(\beta^2-1)^2(\beta+1+(\beta-1)|\boldsymbol{x}|^2)(\beta+1+(\beta-1)|\boldsymbol{y}|^2)}\right).
\end{aligned}
$$

If we substitute the mapping $\boldsymbol{y} = -\boldsymbol{x}$, we find that

$$
\begin{aligned}
c(\boldsymbol{x},-\boldsymbol{x}) &= \log\left(\frac{(\beta+1)^2 - 2(\beta^2-1)|\boldsymbol{x}|^2 + (\beta-1)^2|\boldsymbol{x}|^2|\boldsymbol{x}|^2}{(\beta^2-1)^2\left(\beta+1+(\beta-1)|\boldsymbol{x}|^2\right)^2}\right),\\
\tilde{u}(\boldsymbol{x}) + \tilde{v}(-\boldsymbol{x}) &= \log\left(\frac{(\beta+1)^2 - 2(\beta^2-1)|\boldsymbol{x}|^2 + (\beta-1)^2|\boldsymbol{x}|^2|\boldsymbol{x}|^2}{(\beta^2-1)^2(\beta+1+(\beta-1)|\boldsymbol{x}|^2)^2}\right).
\end{aligned}
$$

Thus for our example, we have found that it indeed holds that $\tilde{u}(\boldsymbol{x}) + \tilde{v}(\boldsymbol{y}) = c(\boldsymbol{x},\boldsymbol{y})$.

5 Discussion

In Section 2.2, we have derived the cost function. There are two points of discussion in this derivation. Firstly, in the beginning of the derivation, we have squared equation (2). Squaring an equation that needs to be solved can introduce a second solution while in the first place, only one solution is desired. This phenomena was also discussed in Section 3.7 of [6]. Unfortunately, we were unable to explain the meaning of this second solution.

Secondly, in the transformation of u to $\tilde{u}$ and similar for v, we have taken the logarithm. This was done in (7). Similar computations were done in Section 3.4 of [10]. It is only allowed to take the logarithm over a positive argument. However, we were unable to verify whether the argument is always positive. This means that we might compute solutions that are not defined.

Lastly, we note that in the computation of the reflectors (29) and (33), we solved quadratic equations. Thus, two solutions could be found for both u as well as for v. For both cases, we found one paraboloid that had the desired geometry and a second solution which was not defined on our domain. Therefore, we discarded the second solution.

References

[1] Energy Information Administration (2021) *How much electricity is used for lighting in the United States?*. Retrievable from `https://www.eia.gov/tools/faqs/faq.php?id=99\&t=3`

[2] European Commission (2021) *Energy prices and costs in Europe*. Retrievable from `https://ec.europa.eu/energy/data-analysis/energy-prices-and-costs_en\#energy-prices-and-costs-dashboards`

[3] Climate Home News (2021) *Signify calls on leaders to accelerate the transition to smart LED lightning at Cop26*. Retrievable from `https://www.climatechangenews.com/2021/11/05/signify-calls-leaders-accelerate-transition-smart-led-lighting-cop26/`

[4] UN climate change conference (2021) *Energy*. Retrievable from `https://ukcop26.org/energy/`

[5] A.S. Glassner (1991). *An introduction to ray tracing.* London : Academic Press

[6] L. B. Romijn (2021). *Generation Jacobian equations in Freeform Optical Design.* Eindhoven University of Technology Library

[7] G. Monge. (1781) *Mémoire sur la théorie des déblais et des remblais* Histoire de l'Académie Royale des Sciences de Paris

[8] F. Cavalletti and M. Huesmann (2015) *Existence and uniqueness of optimal transport maps* In Annales de l'Institut Henri Poincare (C) Non Linear Analysis, volume 32 Elsevier, 2015.

[9] A.H. van Roosmalen, M.J.H. Anthonissen, W.L. IJzerman, J.H.M. ten Thije Boonkkamp *Design of a freeform two-reflector system to collimate and shape a point source distribution.* Optics Expressv29 n16 (2021): 25605-25625

[10] N.K. Yadav (2018). *Monge-Ampère problems with non-quadratic cost function : application to freeform optics.* Eindhoven University of Technology Library